The Effect of Incentive Framing and Descriptive Norms on Internal Whistleblowing*

CLARA XIAOLING CHEN, University of Illinois at Urbana–Champaign†

JENNIFER E. NICHOL, Northeastern University

FLORA H. ZHOU, Georgia State University

ABSTRACT

Firms are under increasing pressure to implement effective internal whistleblowing systems. While firms can provide incentives to encourage internal whistleblowing, it remains controversial how such incentives should be structured. We examine whether the effectiveness of incentives encouraging internal whistleblowing is a joint function of the framing of such incentives (reward or penalty) and the strength of descriptive norms supporting internal whistleblowing. We predict and find in a lab experiment that penalties lead to a greater increase in internal whistleblowing (compared to rewards) when descriptive norms supporting whistleblowing are stronger. Our study contributes to the previous accounting literature on dishonesty and the role of management control systems design in promoting honesty and ethical norms in organizations. We also contribute to an emerging accounting literature on the links between controls, norms, and individual behavior by distinguishing between descriptive norms and injunctive norms and by highlighting the interplay between these two types of norms. Our results have important implications for organizations considering adopting incentives to encourage internal whistleblowing.

L’incidence de la nature des incitatifs et des normes descriptives sur la dénonciation interne

RÉSUMÉ

Les entreprises sont exposées à des pressions croissantes les poussant à instaurer des systèmes de dénonciation interne efficaces. Bien qu’elles puissent proposer des incitatifs visant à encourager la dénonciation interne, la façon de structurer ces incitatifs demeure un sujet de controverse. Les auteures se demandent si l’efficacité des incitatifs encourageant la dénonciation interne est fonction à la fois de la nature de ces incitatifs (récompense ou pénalité) et de la solidité des normes descriptives sur lesquelles repose la dénonciation interne. En procédant à une expérience de laboratoire, les auteures confirment leur

* Accepted by Alan Webb. We thank Alan Webb, two anonymous reviewers, Robert Bloomfield, Tim Bauer, Jan Bouwens, Anthony Bucaro, Richard Crowley, Michael Donohoe, Brooke Elliott, Cassandra Estep, Shawn Gordon, Sean Hillison, Scott Jackson, Khim Kelly, Marsha Keune, Elena Klevsky, Lisa Koonce, Victor Maas, Greg McPhee, Tatiana Sandino, William Snyder, Ivo Tafkov, Bill Tayler, Brian White, Michael Williamson, Gretchen Winter, and Steve Wu, workshop participants at Florida State University, Georgia State University, Harvard Business School, Kent State University, Michigan State University, University of Denver, University of Illinois at Urbana–Champaign, and University of South Carolina, and participants at the American Accounting Association Annual Meeting, the Management Accounting Section Midyear Meeting, the European Accounting Association Annual Conference, and the Conference Board Global Council on Business Conduct Conference for their helpful comments and suggestions. We thank the Campus Research Board and the Center for Professional Responsibility in Business and Society at the University of Illinois at Urbana–Champaign for generous financial support.

† Corresponding author.
prévision selon laquelle les pénalités (comparativement aux récompenses) mènent à une croissance plus grande de la dénonciation interne lorsque les normes descriptives sur lesquelles repose la dénonciation sont plus solides. L’étude contribue aux écrits comptables précédents sur la malhonnêteté et le rôle de la conception des systèmes de contrôle de gestion dans la promotion de l’honnêteté et des normes éthiques au sein des organisations. Les auteures contribuent également aux travaux de recherche récents en comptabilité portant sur les liens entre les contrôles, les normes et le comportement individuel, en établissant la distinction entre normes descriptives et normes injonctives et en mettant en évidence les interactions de ces deux catégories de normes. Ces résultats ont d’importantes conséquences pour les organisations qui envisagent l’adoption de mesures incitatives visant à encourager la dénonciation interne.

1. Introduction

Whistleblowing is “the disclosure by organizational members of illegal, immoral, or illegitimate organizational acts or omissions to parties who can take action to correct the wrongdoing” (Miceli and Near 1992, xv). External whistleblowing involves reporting misconduct to an outside agency such as the SEC, while internal whistleblowing refers to the disclosure of misconduct to other parties within an organization (Bowen, Call, and Rajgopal 2010). Firms are under increasing pressure to implement effective internal whistleblowing systems. New regulations such as the Dodd-Frank Act offer financial incentives for external whistleblowing that may lure employees away from reporting internally.1 Such external incentives create pressure for firms to enhance their internal whistleblowing policies in order to avoid the potential SEC investigation, fines, and negative publicity associated with external whistleblowing. Internal whistleblowing is less costly for an organization because it provides an opportunity for the organization to remediate the misconduct before it snowballs into a serious issue, thereby reducing an organization’s risk of public reputation damage.

Under the pressure of such regulatory changes, some practitioners have proposed using financial incentives to encourage internal whistleblowing (Pearlman and Mufson 2012). However, it is unclear how the incentives encouraging internal whistleblowing should be structured. Specifically, it is unclear whether a reward or penalty framing of such incentives is more effective. In practice, firms that implement incentives often rely on penalties for not blowing the whistle by including a requirement to report misconduct in the employees’ code of conduct. For example, Berkshire Hathaway’s Code of Conduct specifies that a failure to report a violation of the code of conduct may incur “penalties up to and including removal from office or dismissal, ... demotion or re-assignment of the individual involved and suspension with or without pay or benefits.” Other companies have similar statements in their codes of conduct with varying degrees of penalty specificity. In this study, we examine whether the effectiveness of financial incentives encouraging internal whistleblowing is a function of the framing of such incentives (reward or penalty) as well as the strength of the descriptive norms supporting internal whistleblowing. We develop our theoretical arguments by building on both the psychology and behavioral economics literatures on norms, which distinguish between descriptive norms and injunctive norms (e.g., Bicchieri 2006; Cialdini, Reno, and Kallgren 1990; Thogersen 2008).

Descriptive norms represent individuals’ perceptions of what other people commonly do. In practice, firms vary considerably in their descriptive norms regarding whistleblowing. For example, 41 percent of frauds in the healthcare industry are reported by

---

1. The Dodd-Frank Act, which came into effect in 2010, established a whistleblower program enabling the SEC to pay an award (10 percent to 30 percent of the amount recovered) to whistleblowers who provide original information about a violation.
employees, whereas only 14 percent of frauds in the other industries are reported by employees (Dyck, Morse, and Zingales 2010). In contrast, injunctive norms specify individuals’ perceptions of what ought to be done, that is, what actions would be approved or disapproved through social sanctions (Cialdini et al. 1990).

Bicchieri (2006) develops a model suggesting that descriptive and injunctive norms interact synergistically with each other to influence individual behavior (Thogersen 2008). This interaction affects individuals’ preferences to conform to the target behavior such that only when both descriptive norms and injunctive norms are strong will an individual think that other people expect him or her to conform to the target behavior suggested by injunctive norms. In addition, psychology research suggests that penalties are more likely than rewards to communicate injunctive norms in the moral domain (Mulder 2008). In a whistleblowing context, a penalty for not blowing the whistle suggests that whistleblowing is one’s obligation, which communicates a strong injunctive norm. A reward for whistleblowing, on the other hand, implies that whistleblowing is not necessarily a part of one’s normal responsibilities, which does not communicate a strong injunctive norm. Building on these two streams of literature (Bicchieri 2006 and Mulder 2008), we predict an ordinal interaction between incentive framing and descriptive norms such that penalties (which are perceived as strong injunctive norms) will result in a greater increase in internal whistleblowing (compared to rewards) when descriptive norms supporting whistleblowing are stronger.

We test our prediction using 147 undergraduate students enrolled in a large university. We use a 2 × 2 between-subjects design that manipulates the framing of incentives encouraging internal whistleblowing (Reward for Whistleblowing versus Penalty for Not Whistleblowing) and the strength of the descriptive norms supporting internal whistleblowing (Strong Descriptive Norms versus Weak Descriptive Norms). We operationalize incentive framing with financial incentive framing in our experiment because this design choice allows us to hold constant the magnitude of economic incentives and isolate the effects of incentive framing on injunctive norms. Participants in our experiment chose whether or not to blow the whistle on observed overstatements of self-reported performance on a task. As predicted, we find that penalties result in a greater increase in whistleblowing (compared to rewards) when descriptive norms supporting whistleblowing are stronger.

Our study makes the following contributions: First, we contribute to the accounting literature on dishonesty and the role of management control systems design in promoting honesty and ethical norms in organizations (e.g., Evans, Hannan, Krishnan, and Moser 2001; Hannan, Rankin, and Towry 2006; Maas and Van Rinsum 2013; Rankin, Schwartz, and Young 2008; Towry 2003; Webb 2002; Zhang 2008). Prior accounting literature on dishonesty has focused primarily on how dishonesty changes in response to contextual variables (e.g., Evans et al. 2001; Hannan et al. 2006; Maas and Van Rinsum 2013; Rankin et al. 2008), but little is known about the factors affecting how people respond to dishonesty (e.g., whistleblowing). The limited evidence on whistleblowing in the accounting literature has examined the effects of team identity (Towry 2003), fairness (Zhang 2008; Seifert, Sweeney, Joireman, and Thornton 2010), and the ability to collude (Zhang 2008). For example, both Towry (2003) and Zhang (2008) examine whistleblowing decisions when rewards are offered for whistleblowing. Our study complements these studies by examining whistleblowing in response to penalties as well as rewards and by revealing that incentive framing (penalty versus reward) differentially communicates injunctive norms regarding whistleblowing. In particular, we provide evidence consistent with the discussion in Luft (1994) that penalty contracts communicate different noneconomic messages to employees than do bonus contracts. In doing so, our study also adds to the long stream of accounting literature that examines the effects of contract framing on organizational outcomes (e.g., Christ, Sedatole, and Towry 2012; Frederickson and Waller 2005; Hannan, Hoffman, and Moser 2005; Luft 1994).
Second, we contribute to a burgeoning accounting literature on the links between controls, norms, and individual behaviors (e.g., Cardinaels and Yin 2015; Chen and Sandino 2012; Davidson and Stevens 2013; Hannan, Towry, and Zhang 2013; Tayler and Bloomfield 2011). By distinguishing between descriptive norms and injunctive norms and by highlighting the interplay between these two types of norms, our study enhances our understanding of how informal controls based on social norms interact with formal control systems.

Our results have important implications for organizations considering adopting incentives to encourage internal whistleblowing. Our results suggest that incentives framed as penalties will be more effective at encouraging internal whistleblowing (compared to rewards) when strong descriptive norms supporting whistleblowing exist in an organization and no less effective than rewards when strong descriptive norms supporting whistleblowing do not exist in an organization. Thus, firms may be better off implementing penalties for whistleblowing and focusing on increasing the strength of the firm’s descriptive norms for whistleblowing. Although descriptive norms vary across firms, firms can consider ways to strengthen descriptive norms that are supportive of whistleblowing such as periodically reporting to employees that the corporation has investigated whistleblower complaints (Lipman 2012). Furthermore, by definition, descriptive norms are individuals’ perceptions of what people do and, as such, may not be accurate (Lapinski and Rimal 2005; Prentice and Miller 1993; Wenzel 2005). To prevent employees from forming perceptions that descriptive norms for whistleblowing are weak, firms can disseminate actual statistics that suggest more employees report misconduct than people may think. For example, the 2011 National Business Ethics Survey (NBES) shows that 65 percent of employees who observed misconduct chose to report misconduct in the year 2011 (ERC 2012). Statistics like this would help firms foster descriptive norms that are more supportive of whistleblowing.

The remainder of this paper is organized as follows. Section 2 develops the hypothesis. Section 3 describes the experimental design and procedures, and section 4 presents the results. Section 5 concludes.

2. Background and hypothesis development

**Reward versus penalty**

Based on the existing accounting literature, the main effects of rewards versus penalties on internal whistleblowing are ambiguous. Prospect theory suggests that a penalty-based incentive may be more likely to increase internal whistleblowing than a reward-based incentive because individuals tend to weigh losses more heavily than commensurate gains (Kahneman and Tversky 1979; Taylor 1991). Thus, incentives that penalize employees for not reporting observed misconduct should represent a more compelling consequence than financial incentives that commensurately reward employees for reporting observed misconduct. However, although prior research has demonstrated that penalty-framed incentives induce greater effort than reward-framed incentives (e.g., Hannan et al. 2005; Church, Libby, and Zhang 2008), in a moral dilemma such as whistleblowing, it is not necessarily effort that must be motivated in order to obtain the desired response. An individual may perceive the threat of a penalty as unfair, particularly if an individual perceives the required action as threatening one’s freedom (Brehm 1966; Brehm and Brehm 1981). Rewards, on the other hand, may induce greater trust and perceived fairness than penalties (e.g., Christ et al. 2012; Hannan et al. 2005; Luft 1994). When an incentive for whistleblowing is structured as a reward, it may be perceived as a practical measure that reimburses whistleblowers for the risks they take in reporting observed misconduct. If fairness is an important concern in a whistleblowing setting, rewards may prove more effective than penalties at motivating whistleblowing.
The above discussion suggests that we cannot make a clear prediction about the main effects of rewards and penalties on internal whistleblowing. Drawing on the psychology and behavioral economics literatures on social norms, we argue that whether rewards or penalties are more effective in encouraging internal whistleblowing depends critically on the injunctive norms conveyed by the incentives as well as an important contextual factor, that is, the strength of the descriptive norms supporting internal whistleblowing. Psychology research suggests that, in the moral domain, punishments are more likely than rewards to express injunctive norms (Mulder 2008). Two experiments conducted by Mulder (2008) demonstrate that, in a social dilemma in the moral domain, punishments communicate that an act is required (i.e., people are punished when they don’t do what they are supposed to do) while rewards communicate that an act is optional. Following this argument, a penalty suggests that whistleblowing is one’s obligation whereas a reward implies that whistleblowing is not necessarily a part of one’s normal responsibilities.

Below, we discuss our hypothesis development based on descriptive norms and injunctive norms.

**Hypothesis development**

Social norms represent the rules and standards that are understood by members of a group and guide behavior without the force of laws (Sherif 1936). Prior research has shown that social norms can significantly influence personal choices, especially in morally ambiguous situations (Cialdini et al. 1990; Cialdini, Kallgren, and Reno 1991; Lapinski and Rimal 2005; Robinson and O’Leary-Kelly 1998; Tayler and Bloomfield 2011). The psychology literature distinguishes between descriptive norms and injunctive norms (e.g., Cialdini et al. 1990).2 This literature suggests that people have strong preferences to behave in a manner similar to those around them (e.g., Asch 1951; Sherif 1936; Turner 1991), so descriptive norms play an important role in an individual’s behavior because they provide information about the prevailing norms of conduct. In addition, an individual’s behavior is also influenced by perceived approval or disapproval by other people, so injunctive norms also play a role in determining an individual’s behavior (e.g., Ajzen 1988, 1991; Schwartz 1977).

While early research in psychology did not consider the impact of descriptive norms and injunctive norms together, more recent research in psychology and behavioral economics highlights the importance of considering the interactions between descriptive and injunctive norms (Bicchieri 2006; Smith and Louis 2008; Smith et al. 2012; Thogersen 2008). Bicchieri (2006) develops a model of social norms showing that, once a social norm is activated, an individual will follow the norm only if the individual believes that: (i) a large proportion of the relevant population conforms to the social norm (i.e., a descriptive norm), and (ii) a large proportion of the relevant population expects one to conform to the social norm (i.e., an injunctive norm).3 The model implies that if either the descriptive norm or the injunctive norm is weak, variation in the strength of the other norm will have

---

2. Tayler and Bloomfield (2011, 754) define personal norms as “self-based standards or expectations for behavior that flow from our internalized values.” These personal norms are enforced through the anticipation of self-enhancement or self-deprecation (Cialdini et al. 1991; Schwartz 1973, 1977). Schwartz (1973, 1977) differentiates personal norms from injunctive social norms by locating the standards and the sanctions for action for personal norms inside the self and the standards and sanctions for action for injunctive norms in social approval or disapproval. Thus, personal norms and injunctive norms are two related yet distinct constructs. Consistent with this, experimental evidence in Cialdini et al. (1991) suggests that injunctive norms have a separate and additional influence on individual behavior beyond personal norms.

3. Bicchieri (2006) uses the terms “empirical expectations” and “normative expectations” in place of “descriptive norms” and “injunctive norms.” We use “descriptive norms” and “injunctive norms” to be consistent with Thogersen (2008) and other psychology papers that build upon Bicchieri (2006).
little impact on an individual’s behavior. The intuition behind this model is as follows: when the injunctive norm is weak, there is no implicit rule for a target behavior. Under such circumstances, even if other people exhibit the target behavior, their behavior does not necessarily indicate that an individual is expected to follow what other people do, thereby limiting the effect of the descriptive norm on an individual’s behavior. Similarly, when there is an implicit rule communicated by an injunctive norm, but other people’s behavior often violates the implicit rule (i.e., the descriptive norm is weak), an individual would interpret other people’s behavior as signaling a lack of legitimacy of the rule. As a result, an individual would not think that other people expect him or her to comply with the implicit rule. Thus, weak descriptive norms would reduce the impact of the injunctive norm on an individual’s behavior. Only when both descriptive norms and injunctive norms are strong will an individual prefer to conform to the implicit rule. In summary, this model predicts that descriptive and injunctive norms interact synergistically to induce norm-conforming behavior such that an individual will conform to a social norm only when both descriptive and injunctive norms are strong.

Building on Bicchieri’s (2006) model, we predict that a penalty contract (which communicates a strong injunctive norm) will have a greater impact on whistleblowing behavior when descriptive norms supporting whistleblowing are stronger. On the other hand, a reward-based incentive does not communicate an injunctive norm or at best communicates a weak injunctive norm because a reward suggests that whistleblowing is not necessarily expected. Because the injunctive norm is likely to be weak to nonexistent in this setting, we do not expect a relationship between descriptive norms and whistleblowing.

Taken together, we predict an ordinal interaction between incentive framing and descriptive norms such that penalties will result in a greater increase in internal whistleblowing (compared to rewards) when descriptive norms supporting whistleblowing are stronger. In other words, we predict that a penalty contract, combined with strong descriptive norms supporting whistleblowing, will lead to the greatest likelihood of whistleblowing.

We do not predict a main effect of incentive framing because when descriptive norms supporting whistleblowing are weak, even strong injunctive norms communicated by penalties are likely to be ineffective according to Bicchieri’s model. Similarly, we do not predict a main effect of descriptive norms because rewards would communicate a weak injunctive norm at best, making descriptive norms ineffective at inducing internal whistleblowing. Thus, we posit the following hypothesis:

**Hypothesis.** Penalties lead to a greater increase in internal whistleblowing (compared to rewards) when descriptive norms supporting whistleblowing are stronger.

We note that our prediction on the interaction effect of descriptive norms and incentive framing on internal whistleblowing is not without tension. The majority of the experimental evidence in prior literature is consistent with Bicchieri’s theory (e.g., Schultz, Khazian, and Zaleski 2008; Smith and Louis 2008; Smith et al. 2012). For example, Bicchieri’s theory has received empirical support in settings such as political attitude (the second experiment in Smith and Louis 2008) and conservation (Schultz et al. 2008 and Smith

---

4. One exception is the first experiment in Smith and Louis (2008, 656), where the authors find that nonsupportive descriptive norms and supportive injunctive norms result in the highest level of targeted behavior. This result is likely driven by the choice of a very salient political issue in the experimental task. If the majority of peers approved of a particular action on a salient and important political issue but the peers were not taking this action yet, the participant may have felt an obligation to engage in the action (i.e., “if nobody else is doing it, but we all think it’s a good idea, then I’ll have to do it”). In our experimental setting, however, whistleblowing is a morally ambiguous action, so we do not expect this result to generalize to our setting.
et al. 2012). However, there are important differences between our study and prior studies that find the predicted interactive effect between descriptive and injunctive norms. First, prior studies have used injunctive norms that are developed from the participants’ peers (e.g., Gockeritz et al. 2010; Smith et al. 2012; Thogersen 2008), whereas the source of the injunctive norms in our study is an organization’s top management who have designed the incentive system. Prior studies speculate that such a difference in the source of injunctive norms could lead to different effects on individual behavior (Lee et al. 2007; Rimal and Real 2003; Smith and Louis 2008). Second, we manipulate the presence versus absence of injunctive norms indirectly via incentive schemes, whereas prior studies manipulate the strength of injunctive norms directly (e.g., Smith et al. 2012). To the extent that the incentive framing used in our study manipulates injunctive norms indirectly, we believe this is a weaker manipulation of the injunctive norm compared to prior studies and thus biases against finding results. These differences between our study and prior studies suggest that the predicted ordinal interaction of descriptive norms and incentive schemes on whistleblowing is not a forgone conclusion.

3. Method

We recruited 147 undergraduate student volunteers enrolled in a large university to participate in one of eleven 60-minute experimental sessions. We randomly assigned participants to 49 three-person groups and all groups interacted primarily via computer. In each three-person group, we randomly assigned two participants to be regular participants and the third participant to be the “misconduct” participant, giving us 98 (49 × 2) regular participants from which to draw our observations of interest. We will explain this design choice in more detail below.

Experimental procedures

We conducted our experiment using z-Tree software (Fischbacher 2007). To ensure anonymity, we assigned each participant a code name. To induce some level of group identification, the code names were written on colored notes indicating the participants’ group membership (e.g., “Red Team,” “Blue Team”) (Towry 2003). At the beginning of each session, participants were assigned to partitioned computer stations close to the other members of their group and were asked to introduce themselves to their group members by their code names. Afterwards, each participant stood up one at a time and stated his/her code name aloud to the room to increase participants’ awareness that any actions they might take under their code name during the course of the experiment would not be anonymous to the room at large. Beyond these acts, participants did not communicate with each other during the experiment. All subsequent interactions among the group members occurred via computer.

5. Prior studies using the survey methodology find mixed results for the interactive effect between descriptive norms and injunctive norms (Lee, Geisner, Lewis, Neighbors, and Larimer 2007; Rimal and Real 2003, 2005). We note that these survey studies suffer from several limitations: First, they rely on self-reported measures of targeted behavior, which are likely to be biased in domains where social desirability concern is salient (e.g., conservation, drinking) (Hamilton 1985). Second, due to the lack of independence between descriptive norms and injunctive norms in natural settings, survey studies do not allow these two types of norms to be orthogonal to each other, leading to potential confounding effects. Third, the cross-sectional design of survey studies does not allow causal inferences to be drawn. It is also likely that respondents rationalize their behavior by referring to descriptive norms (Rimal and Real 2003).

6. We do not think this difference would have an effect on the shape of the interaction. Under the theoretical framework of Bicchieri (2006), descriptive norms and injunctive norms are conceptualized as continuums. As one type of norm moves up on the continuum, the impact of the other type of norm on the targeted behavior will become stronger.
We chose to disclose the code names of the whistleblowers in our experiment to create a social disincentive for whistleblowing. In reality, the social disincentive of whistleblowing tends to be very high (Dyck et al. 2010). Whistleblowers often face risks such as losing their jobs and their potential to be employed in the future and being socially ostracized or harassed. We operationalized the social disincentive for whistleblowing in our experimental setting by informing participants that their whistleblowing decisions would be exposed to everyone in the room at the end of the experiment. This design choice reflects the way whistleblowing often plays out in organizations. Although in theory organizations are supposed to provide anonymity to whistleblowers, anonymity has thus far been the exception rather than the rule for the following reasons: (i) many whistleblowers choose not to remain anonymous (Dyck et al. 2010; ERC 2010); (ii) many organizations have inadequate systems to protect the identity of whistleblowers (Curtis 2006; Greenberg 2011); and (iii) even when organizations make every effort to protect the anonymity of whistleblowers, the identity of the whistleblower is usually revealed in the investigation process inadvertently. In summary, the empirical evidence suggests that most whistleblowing activity is not anonymous in practice and even if an organization promises to keep the identity of a whistleblower confidential, concerns related to potential exposure are likely to be high. Therefore, the threat of exposure to the whistleblower in our research design should increase the generalizability of our results.

We test our hypothesis using a reporting game. All participants (1) performed a task, (2) reported task performance to the experimenter, (3) viewed any overstating by one of their group members, and (4) chose whether to blow the whistle on observed overstating. Participants repeated steps (1) to (3) across three periods. We explain our design choices for each of these steps in more detail below.

For step (1), participants had up to four minutes to answer 10 quantitative problem solving, quantitative data sufficiency, verbal sentence correction, and verbal critical reasoning questions similar to questions on graduate tests such as the LSAT and GMAT. After the four minutes expired, the computer scored participant responses and each participant saw his or her actual score.

For step (2), regular participants reported a score to the experimenter as their official score (Official Score). Misconduct participants were offered a financial reward of $2 for overstating their scores by a fixed two points each period. Regular and misconduct participants were only informed of their own payment schemes. Misconduct participants had the option of not overstating their scores and thus forgoing the reward. Participants were informed that at least one of their group members would be shown any overstatements in which they had engaged and would have an opportunity to report them. All participants were informed that their Official Scores would be displayed publicly to the room at the end of the experiment unless a group

---

7. Only 3 percent of reports about internal misconduct are reported through company hotlines (ERC 2010). The majority of whistleblowers choose to report wrongdoing directly to someone they know within the organization instead of to a hotline (ERC 2010). When employees report wrongdoing directly to someone within the organization, anonymity is not an option. Even when hotlines are used and anonymity is an option, whistleblowers often do not remain anonymous. Dyck et al. (2010) report that 63 percent of the employee whistleblowers in their sample chose not to remain anonymous.

8. Many organizations implement only the most rudimentary systems in order to appear to protect potential whistleblowers (Greenberg 2011). Curtis (2006) suggests that many managers make little effort to safeguard the privacy of the whistleblower. When asked what best described the level of confidentiality protection for whistleblowers in their organization, none of the respondents selected the responses indicating the strongest levels of confidentiality. Instead, 26 percent said “confidential to the extent possible,” 16 percent chose “casual effort,” and 58 percent said “no steps taken” to ensure confidentiality.

9. Misconduct participants were not given the option to overstate by anything other than this fixed amount.
member blew the whistle on them for overstating, in which case the participants’ actual scores (rather than their Official Scores) would be displayed publicly to the room.  

For step (3), regular participants were only shown the misconduct participant’s overstatement amount, which was kept constant at two points each period, conditioning on the misconduct participant choosing to overstate. We made the design choices in steps (2) and (3) in order to hold the overstatement observed by regular participants constant across all conditions. If some conditions induced more overstating than other conditions, whistleblowing behavior could be a function of the magnitude of the overstatement rather than the incentives for whistleblowing. Our experimental design rules out this potential concern. Out of 49 misconduct participants, only two chose not to overstate. When a misconduct participant chose to not overstate, there was no overstatement to show to his/her peers and thus no opportunity for whistleblowing, so we excluded those two groups (with two observations from regular participants in each group) from our analyses. To avoid any deception, misconduct participants were also exposed to their group members’ (regular participants) overstatements at the end of the final period and given the opportunity to report their group members. This exposure and whistleblowing opportunity ensured that regular participants still experienced the threat of being reported as stated, but their variable overstating would not change the overstatement amount viewed by other regular participants.

Finally, for step (4), regular participants chose whether to blow the whistle on the overstatements they observed.

Our operationalization of misconduct within the experimental task is representative of a common type of misconduct employees observe and report on internally in practice. Out of the 35 types of misconduct documented in the NBES 2011 report (ERC 2012), three types are closely related to the setting we use in our experiment: falsifying expense reports, falsifying and/or manipulating financial reporting information, and falsifying time reports or hours worked. These three types of misconduct have a reporting rate between 61 percent and 66 percent, indicating that our setting generalizes to the type of misconduct employees tend to encounter and report.

Participants had a trial run of the reporting game and then repeated the same reporting game for three periods, after which participants filled out a postexperimental questionnaire and received their payment for participating in the experiment.

We used data for Period 1 for our analysis, but we gathered data for Periods 2 and 3 to explore an idea that examines the evolution of norms as a result of financial incentives. We focus on Period 1 for the data analysis because in Periods 2 and 3, participants observed repeated overstating by group members despite their whistleblowing, which theory suggests would decrease the perceived legitimacy of injunctive norms and reduce the impact of injunctive norms on the participants.

10. We intentionally made the test very challenging so that most participants received low scores on the test. This created a social cost of embarrassment for participants due to knowing their low test scores would be publicized to their peers. Social embarrassment is important in our experiment because it provided the motivation for a participant to overstate. We made this design choice so that regular participants would find overstatements by the misconduct participants believable.

11. We did not show the misconduct participant’s Official Score or actual score to the regular participants in order to reduce social comparisons between the regular participants’ own scores and their peers’ scores. Such social comparisons could bias whistleblowing tendencies and are not the focus of our study. For example, if the regular participant had a score similar to the misconduct participant’s score, it is likely to increase the competition between the regular participant and the misconduct participant and cause the regular participant to be more willing to blow the whistle (Festinger 1954; Garcia, Tor and Schiff 2013).

12. Out of the two misconduct participants who chose not to overstate, one comes from the Penalty/Strong Descriptive Norms treatment and the other comes from the Reward/Weak Descriptive Norms treatment.
Appendix 1 presents the experimental instructions provided to both the regular participants and the misconduct participants.

**Experimental design**

We randomly assigned each three-member group to one of four experimental conditions created by crossing incentive framing and the strength of descriptive norms supporting whistleblowing. We discuss our manipulations of these variables below.

**Incentive framing**

Our first manipulated factor is the framing of the incentive provided to participants to encourage whistleblowing. In the reward conditions, the regular participants received a base pay of $16 and a reward of $1 each time they chose to report a group member’s overstatement in each of the four periods (including the trial run). In the penalty conditions, the regular participants received a base pay of $20 and a penalty of $1 each time they chose to *not* report a group member’s overstatement in each of the four periods. Thus, $20 could be made in each condition across four periods if a participant always blew the whistle in each period. We set the potential pay across conditions equal based solely on economic predictions. We chose to keep the potential pay, as opposed to expected pay, constant based on the following two considerations: first, in our experiment, it is difficult, if not impossible, to predict ex ante the average frequency of whistleblowing across the four conditions in order to set the expected pay equal across our treatments. Second, setting the potential pay equal across the four treatments based solely on economic predictions allows us to focus on the effects of behavioral factors on whistleblowing. Appendix 2 presents the pay schemes used for both the regular participants and the misconduct participants.

**Descriptive norms**

Our second manipulated factor is the strength of descriptive norms for internal whistleblowing. We manipulate the strength of descriptive norms for internal whistleblowing by presenting information to participants about people’s behavior in a previous experiment similar to the experiment in which they were taking part. Specifically, participants in the strong descriptive norms supporting whistleblowing condition (Strong Descriptive Norms) were told the following: “In a previous experiment similar to this one, most people OVERWHELMINGLY chose to report their group members.”

Participants in the weak descriptive norms supporting whistleblowing condition (Weak Descriptive Norms) were told the following: “In a previous experiment similar to this one, most people OVERWHELMINGLY chose to *not* report their group members.”

To avoid deception, we took extra steps to ensure that the claims regarding the strength of descriptive norms that we made to our participants in the manipulation were honest. Specifically, we first ran two pilot studies and then used the pilot results to support our claims in the strength of descriptive norms manipulation.13

---

13. In Pilot 1, we did not manipulate descriptive norms but we followed the same basic procedures of the experiment. We were able to achieve variations in whistleblowing behavior across treatments in Pilot 1. In Pilot 2, we followed the same procedures but also manipulated descriptive norms similar to the manipulation in our official experiment, relying on the results found in Pilot 1 to inform participants about whistleblowing descriptive norms in a prior experiment. We also included a “no financial incentives” condition in Pilot 2 and obtained a majority of participants who blew the whistle in the Strong Descriptive Norms treatment and a majority of participants who did not blow the whistle in the Weak Descriptive Norms treatment. Therefore, in our official experiment, we were able to truthfully tell participants that in an experiment similar to the one they were in, most participants did or did not report misconduct by drawing on different observations from Pilot 2.
Dependent variable—internal whistleblowing

Our dependent variable is the participant’s internal whistleblowing decision, which is coded as one if a participant reports the overstatement by his or her group member in the period directly following the trial run, and zero otherwise.

4. Results

For data analysis purposes, we use the regular participants only. After eliminating two groups (four regular participants—two from the Penalty/Strong Descriptive Norms treatment and two from the Reward/Weak Descriptive Norms treatment) whose misconduct members chose not to overstate their performance, we had 94 usable observations.14 As mentioned before, we use Period 1 for the data analysis.

Since whistleblowing is defined as “the disclosure by organizational members of illegal, immoral, or illegitimate organizational acts or omissions to parties who can take action to correct the wrongdoing” (Miceli and Near 1992, xv), we check whether our participants consider overstating as immoral or unacceptable in our experiment. We find that the majority of our participants indicate that overstating self-reported performance is wrong and that the majority chose not to overstate their own performance. While all participants did very poorly in the test (the highest score was 50 percent) and thus faced an incentive to overstate their performance due to potential social embarrassment, only 33 out of the 94 regular participants chose to overstate. This suggests that the majority of our participants did not view the act of overstating their scores as acceptable. Furthermore, in the postexperimental questionnaire, we asked participants to rate the extent to which they think overstating self-reported performance is wrong on an 11-point scale, with 0 being “Strongly Disagree” and 10 being “Strongly Agree” that it is wrong. The average response is 8.16, indicating that the majority of our participants think it is wrong to overstate self-reported performance.

To examine whether descriptive norms varied as the manipulations intended, we asked participants to estimate the percentage of their peers likely to report their group members’ overstating with the following request: “Please estimate the percentage of participants in this experiment that you believe reported their group members for overstating.” Since the response to this question could be influenced by manipulations of both descriptive norms and incentive framing, we conduct a full factorial ANOVA including both incentive framing and descriptive norms on the estimated percentage. Consistent with our manipulation, the results show that the main effect of descriptive norms is significant ($p = 0.008$) and that the main effect of incentive framing ($p = 0.492$) and the interaction effect ($p = 0.507$) are not significant. Specifically, consistent with our manipulation, participants in the Strong Descriptive Norms condition think a higher percentage of their peers were likely to report their group members’ overstating than those in the Weak Descriptive Norms condition (70 percent versus 53 percent, $p = 0.005$).

We now present analyses investigating the interaction of incentive framing (Penalty or Reward) and the strength of descriptive norms (Strong Descriptive Norms or Weak Descriptive Norms) on internal whistleblowing.

Test of Hypothesis

Because our dependent variable is a binary variable, we use a nominal logistic model, which is appropriate for the analysis of binomially distributed categorical outcomes.

---

14. Note that misconduct participants’ instructions were constant across all treatments, so any decision to abstain from overstating was not related to our manipulations. We did not manipulate incentives or descriptive norms for misconduct participants.
TABLE 1
The effects of incentive framing and descriptive norms on whistleblowing using a nominal logistic model

<table>
<thead>
<tr>
<th>Panel A: Frequency (percentage) of whistleblowing*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
</tr>
<tr>
<td>Reward for whistleblowing</td>
</tr>
<tr>
<td>Penalty for not blowing the whistle</td>
</tr>
</tbody>
</table>

| Panel B: Effect likelihood ratio test from a logistic model of whistleblowing† |
| Factor                                        | df | Chi-square | p-value‡ |
| Incentive framing§                            | 1  | 7.09       | 0.008    |
| Descriptive norms¶                           | 1  | 8.66       | 0.003    |
| Incentive framing×Descriptive norms          | 1  | 2.80       | 0.095    |

<table>
<thead>
<tr>
<th>Panel C: Planned contrast and follow-up simple effect tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
</tr>
<tr>
<td>Overall test: Penalties lead to a greater increase in whistleblowing (compared to rewards) when descriptive norms supporting whistleblowing are stronger.**</td>
</tr>
<tr>
<td>Follow-up simple effect tests:</td>
</tr>
<tr>
<td>Effect of descriptive norms given penalty-based incentives</td>
</tr>
<tr>
<td>Effect of descriptive norms given reward-based incentives</td>
</tr>
<tr>
<td>Effect of incentive framing when descriptive norms supporting whistleblowing are stronger</td>
</tr>
<tr>
<td>Effect of incentive framing given descriptive norms supporting whistleblowing are weaker</td>
</tr>
</tbody>
</table>

**Notes:** * Whistleblowing is the participant’s whistleblowing decision, which is coded as 1 if a participant reports overstatement by his or her group member in the period directly following the trial run, and 0 otherwise. Percentage of Whistleblowing is the number of regular participants who report overstatement by group members divided by the total number of regular participants in each treatment.
† The chi-square test for the whole model shows that the p-value of the logistic model is 0.0041, indicating that the model has a good fit overall.
‡ Chi-square tests are one-sided. Therefore, all the reported p-values in this table are original p-values.
§ We manipulated Incentive Framing between subjects at two levels. Participants were either promised a reward for reporting observed misconduct or threatened with a penalty for not reporting observed misconduct.
¶ We manipulated Descriptive Norms between subjects at two levels of strength. Participants were told that in a previous experiment, others overwhelmingly chose to report their peers (Strong Descriptive Norms) or to not report their peers (Weak Descriptive Norms).
** We use contrast weights of +3 for the Penalty/Strong Descriptive Norms condition, −1 for the Reward/Strong Descriptive Norms condition, −1 for the Penalty/Weak Descriptive Norms condition, and −1 for the Reward/Weak Descriptive Norms condition based on our theoretical predictions.
Specifically, we test our hypothesis by calculating the likelihood-ratios using a chi-square test. Table 1 summarize these results. We use a planned contrast test and pairwise simple contrasts to provide an efficient and powerful test of the ordinal interaction hypothesis (Buckless and Ravenscroft 1990; Bobko 1986; Strube and Bobko 1989). Given the theoretical prediction, we use contrast weights of +3 for the Penalty/Strong Descriptive Norms condition, −1 for the Reward/Strong Descriptive Norms condition, −1 for the Penalty/Weak Descriptive Norms, and −1 for the Reward/Weak Descriptive Norms conditions to test the ordinal interaction. Panel C of Table 1 provides the results of our contrast test. The hypothesis test using the above contrast weights is statistically significant ($\chi^2 = 11.77, p < 0.001$), supporting our hypothesis.

Follow-up simple effects shown in Table 1, panel C confirm that internal whistleblowing is significantly higher in the Penalty/Strong Descriptive Norms condition than in either the Penalty/Weak Descriptive Norms condition ($p = 0.005$) or the Reward/Strong Descriptive Norms condition ($p = 0.008$). Table 1, panel C also shows no significant difference in internal whistleblowing between the Penalty/Weak Descriptive Norms condition and the Reward/Weak Descriptive Norms condition ($p = 0.389$) or between the Reward/Strong Descriptive Norms condition and the Reward/Weak Descriptive Norms condition ($p = 0.265$). These simple effects suggest that strong descriptive norms increase whistleblowing given penalty-based incentives but do not have a significant influence on whistleblowing given reward-based incentives. In addition, penalties lead to higher whistleblowing than rewards when descriptive norms supporting whistleblowing are strong, but not when descriptive norms supporting whistleblowing are weak. These results are consistent with Bicchieri’s model, indicating that, when either injunctive norms or descriptive norms are weak, the other type of norms does not have a significant impact on whistleblowing behavior.

To rule out the possibility that the above result is driven by the variations in the degree to which regular participants overstated their own scores, we examine regular participants’ overstating behavior across conditions and conduct the planned contrast analysis after controlling for regular participants’ overstatements in Period 1. We find that regular participants’ overstating behavior does not vary significantly across conditions. All of our results remain qualitatively similar after controlling for regular participants’ overstatements.

In summary, we find strong support for an ordinal interaction effect between incentive framing and descriptive norms such that penalties lead to a greater increase in internal whistleblowing as compared to rewards when descriptive norms supporting whistleblowing are stronger.

Supplemental analyses of postexperimental questions

To shed light on the process underlying individuals’ whistleblowing decisions, we conduct additional analyses of the postexperimental questions we collected. Specifically, we first conduct a factor analysis of the full set of postexperimental questions to identify the
underlying constructs. We then conduct additional analyses to examine how our treatments influence these constructs and how these constructs influence participants’ whistleblowing decisions.

Table 2, panel A presents the descriptive statistics of participants’ responses to the postexperimental questionnaire items by condition. Table 2, panel B presents the results of the factor analysis of the postexperimental questionnaire items. The factor analysis extracts five factors, accounting for 70 percent of the total variance in the data. We eliminate questions with factor loadings below 0.7 (Comrey and Lee 1992). Items 5 and 6 load on Factor 1, which we label “image concerns,” indicating participants’ concerns about how they appear to their group members. Items 1 and 2 load on Factor 2, which captures participants’ perceived clarity of expectations to blow the whistle. Item 12 loads on Factor 3, reflecting participants’ perceived group identity. Items 8 and 10 load on Factor 4. We name Factor 4 “personal norms” because it captures the participants’ standards for behavior based on their internalized values. Item 11 loads on Factor 5, reflecting participants’ perceived fairness of payment. We use the average of participants’ responses to the questions that load on each factor for our analyses.

We conduct supplemental analyses on these factors. We first examine whether Incentive Framing and Descriptive Norms influence these factors using the analysis of variance model (ANOVA). If the factors are influenced by the treatments, we further examine whether these factors influence whistleblowing after controlling for Incentive Framing and Descriptive Norms.

First, we find that the “clarity of expectations” construct (Factor 2) is jointly influenced by incentive framing and descriptive norms. Specifically, we find that the clarity of expectations to blow the whistle is higher in the Penalty/Strong Descriptive Norms condition than in the other three conditions (7.72 for Penalty/Strong Descriptive Norms versus 5.46 for Penalty/Weak Descriptive Norms, 5.83 for Reward/Strong Descriptive Norms, and 4.73 Reward/Weak Descriptive Norms, \( p < 0.001 \), two-tailed), suggesting that an individual will have clearer expectations to follow the norms when both injunctive norms and descriptive norms are strong. In addition, we also find that the “clarity of expectations” construct (Factor 2) also influences whistleblowing.

Second, we find that participants are more concerned about how they appear to their group members (Factor 1) in the Reward/Strong Descriptive Norms condition compared to the other three conditions (7.08 for Reward/Strong Descriptive Norms versus 5.71 Penalty/Weak Descriptive Norms, 5.78 Penalty/Strong Descriptive Norms, and 5.38 Reward/Weak Descriptive Norms, \( p = 0.007 \), two-tailed). Participants in the Reward/Strong Descriptive Norms condition are more likely to think they would look greedy to group members if they reported their group members. This suggests that using rewards to encourage whistleblowing when the descriptive norms supporting whistleblowing are already strong adds a social stigma to whistleblowing. However, this factor does not influence whistleblowing after controlling for the effects of our treatments, suggesting that image concerns do not drive our results.

Third, the analysis on pay fairness (Factor 5) indicates that participants appear to find their compensation fair in general but more fair when there are strong descriptive norms supporting whistleblowing (mean = 8.48 and 7.20 for Strong Descriptive Norms and Weak Descriptive Norms, respectively; \( p\)-value = 0.009, two-tailed). This may be due to participants requiring a higher compensation in return for being asked to report a peer when social costs associated with whistleblowing are perceived to be higher. However, we do not find a main effect of incentive framing or an interaction effect between descriptive norms and incentive framing on the perceived fairness of compensation. Perceived fairness of compensation also does not influence whistleblowing after controlling for the effects of our treatments. These results suggest that perceived fairness of compensation is unlikely to drive our results.
TABLE 2
Descriptive statistics and factor analysis of postexperimental questionnaire responses

<table>
<thead>
<tr>
<th>Questionnaire items</th>
<th>Reward for whistleblowing</th>
<th>Penalty for not blowing the whistle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strong Descriptive Norms</td>
<td>Weak Descriptive Norms</td>
</tr>
<tr>
<td>(1) I was expected to report my group members for overstating their scores</td>
<td>6.69 (2.49)</td>
<td>8.33 (2.22)</td>
</tr>
<tr>
<td>(2) I knew exactly whether I was expected to report my group members</td>
<td>4.96 (2.47)</td>
<td>7.11 (1.94)</td>
</tr>
<tr>
<td>(3) I am sure I did the right thing in my decision(s) to report or not report my group members</td>
<td>6.69 (2.99)</td>
<td>8.00 (1.97)</td>
</tr>
<tr>
<td>(4) It was difficult to decide whether to report my group members</td>
<td>4.73 (3.17)</td>
<td>4.05 (3.39)</td>
</tr>
<tr>
<td>(5) I would look bad to others if I reported my group members</td>
<td>6.00 (4.54)</td>
<td>3.89 (3.19)</td>
</tr>
<tr>
<td>(6) Other people are likely to think any decision I made to report my group members was motivated by greed</td>
<td>8.15 (6.23)</td>
<td>7.67 (6.04)</td>
</tr>
<tr>
<td>(7) Other people are likely to think any decision I made to report my group members was motivated by fear of the consequences</td>
<td>4.62 (3.13)</td>
<td>6.00 (2.79)</td>
</tr>
<tr>
<td>(8) Other people are likely to think any decision I made to report my group members was motivated by ethics</td>
<td>6.00 (2.73)</td>
<td>5.94 (2.26)</td>
</tr>
<tr>
<td>(9) I was concerned about how my decisions would affect others</td>
<td>5.91 (2.54)</td>
<td>6.39 (2.89)</td>
</tr>
<tr>
<td>(10) Overstating self-reported performance is wrong</td>
<td>8.31 (6.80)</td>
<td>8.56 (7.79)</td>
</tr>
<tr>
<td>(11) The way I expect to be paid is fair</td>
<td>8.58 (7.38)</td>
<td>8.33 (7.00)</td>
</tr>
<tr>
<td>(12) As a whole, I liked the members in my group and felt attached to them</td>
<td>4.27 (4.19)</td>
<td>5.72 (4.71)</td>
</tr>
</tbody>
</table>

Panel B: Factor analysis

<table>
<thead>
<tr>
<th>Items</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.092</td>
<td>0.880</td>
<td>-0.008</td>
<td>-0.007</td>
<td>0.040</td>
</tr>
<tr>
<td>2</td>
<td>-0.068</td>
<td>0.885</td>
<td>0.137</td>
<td>0.123</td>
<td>-0.059</td>
</tr>
<tr>
<td>3</td>
<td>-0.247</td>
<td>0.450</td>
<td>-0.087</td>
<td>0.466</td>
<td>0.268</td>
</tr>
<tr>
<td>4</td>
<td>0.625</td>
<td>-0.054</td>
<td>0.503</td>
<td>0.054</td>
<td>-0.259</td>
</tr>
<tr>
<td>5</td>
<td>0.734</td>
<td>-0.092</td>
<td>0.128</td>
<td>-0.070</td>
<td>0.158</td>
</tr>
<tr>
<td>6</td>
<td>0.755</td>
<td>0.106</td>
<td>-0.193</td>
<td>-0.006</td>
<td>0.112</td>
</tr>
</tbody>
</table>

(The table is continued on the next page.)
Our treatments do not have a significant effect on group identity (Factor 3) or personal norms (Factor 4). These two factors also do not influence whistleblowing in our study.

Item 7 does not load significantly on any of the factors, so we analyze this item separately. We find that participants in the Penalty condition are significantly more likely to agree with the statement in Item 7: “Other people are likely to think any decision I made to report my group members was motivated by fear of the consequences” (mean = 5.86 and 4.67 for Penalty and Reward, respectively; p-value = 0.048, two-tailed). This is consistent with loss aversion playing a role in the whistleblowing decision. However, we do not find a main effect of descriptive norms or an interaction effect between incentive framing and descriptive norms on the participants’ response to this statement, which suggests that loss aversion is unlikely to drive our results.

Overall, these additional analyses help rule out potential alternative explanations for our results and give us greater confidence that our results are driven by the interaction between incentive framing and descriptive norms.

5. Conclusion

We report the results of an experiment where participants in three-person groups (1) performed a task, (2) reported task performance to the experimenter, (3) viewed any overstating by one of their group members, and (4) chose whether to blow the whistle on observed overstatements. We examine whether the effectiveness of incentives encouraging internal whistleblowing is a function of the framing of such incentives (a reward for whistleblowing or penalty for not whistleblowing) and the strength of the descriptive norms supporting whistleblowing (strong or weak). We find an ordinal interaction between incentive framing and descriptive norms such that penalties lead to a greater increase in internal whistleblowing (compared to rewards) when descriptive norms...
supporting whistleblowing are strong than when descriptive norms supporting whistleblowing are weak.

Our results have important implications for organizations that use or plan to use incentives to encourage internal whistleblowing in order to reduce potential negative consequences associated with the external whistleblowing encouraged under the Dodd-Frank Act. The main takeaway from our study is consistent with a recent framework that views management control as a system where the value of one management control practice depends on the use of another management control practice (Grabner and Moers 2013). Specifically, our results suggest that the effect of a formal control (i.e., an incentive encouraging internal whistleblowing) depends on the informal controls (i.e., descriptive norms supporting whistleblowing) that are already in place in an organization. Ceteris paribus, a penalty will be more effective than a reward in an organization when there are strong descriptive norms encouraging whistleblowing and no less effective than a reward in an organization with weak descriptive norms encouraging whistleblowing. Thus, organizations may be better off implementing penalties for whistleblowing while working toward increasing the strength of the firm’s descriptive norms for whistleblowing.

Limitations of our paper provide opportunities for future research. First, our experimental study abstracts away from many factors in the real world, including the variations in the implementation cost or difficulty for penalties versus rewards. In reality, the ability of an organization to enforce a penalty contract or make the threat of a penalty credible can vary across organizations. Future research can examine whether perceived penalty enforceability reduces a penalty’s injunctive norm influence. That said, our findings on the theoretical implications of penalties versus rewards should be generalizable as long as organizations have a strong internal control and sanction system to make the penalty a credible threat. Along the same lines, in both the reward and penalty conditions, the enforceability of the contract is perfect in our experiment. We made this design choice to isolate and increase the chance of detecting the effects of incentive framing. In reality, implementing an incentive contract for whistleblowing, regardless of it being a reward or a penalty contract, usually involves some level of uncertainty from the potential whistleblower’s perspective. That said, we would expect to observe the same directional results if we introduced incentive uncertainty into the research design.

Second, the real costs and benefits of internal whistleblowing can be much greater in magnitude in the real world. It is possible that changing the magnitude of the costs and benefits would moderate our results. Future research can use other research methodologies to examine whether the results documented in our laboratory experiment generalize to a real-world setting.

Third, we examine only a single period in our data analysis, whereas in the real world, employees interact with each other repeatedly. Such interactions may change the descriptive norms perceived by the employees. Also, the framing of the incentives per se may change perceived descriptive norms in the long term.

Finally, the source of the injunctive norms in our study is the top management of the organization who have designed the incentive system. In the design of management control systems, injunctive norms and descriptive norms can arise from different referent groups. Prior psychology research provides mixed evidence on how the source of the norms, or the reference group identity, influences the impact of social norms (e.g., Rimal and Real 2005; Schultz et al. 2008). Exploring whether and how referent group identity moderates the effects of descriptive norms and injunctive norms on individual behavior in organizations would be an interesting area for future research.
Appendix 1

Experimental instrument

Panel A: Experimental instrument for the regular participants

[Screen 1]
This experiment consists of four periods. Each period begins with a task where you are asked to answer a set of questions. You will receive a fixed amount of <Absent/Penalty Treatments = $5; Reward Treatments = $4> each period for this task. Once you have finished answering the questions, the computer will tell you which questions you answered correctly. You will then be asked to submit your Official Score.

Each period’s Official Score will be publicly displayed to everyone in the room at the end of the experiment and will identify you by your team color and code name.

NO CALCULATORS, PAPER, PENCILS, ETC. ARE ALLOWED WHEN ANSWERING QUESTIONS. Please answer all questions as well as you can “in your head”.

[Screen 2]
If the Official Score you submit in any period is higher than your actual score in that period, your overstatement will be shown to one of your group members, who will have an opportunity to report you. If your peer chooses to report you, your actual score (rather than your Official Score) will be displayed at the end of the experiment.

Similarly, if one of your peers submits an Official Score higher than his/her actual score, the overstatement will be shown to you and you will have an opportunity to report it. If you report it, the score shown for your peer at the end of the experiment will be lower than he/she submitted.

In addition to everyone’s question task scores being publicly displayed at the end of the experiment as described, anyone who has chosen to report a peer will be publicly listed with his/her Team color, code name, and the percentage of times he/she reported a peer shown for all to see.

THE EXPERIMENTER WILL ANNOUNCE OUT LOUD AT THE END OF THE EXPERIMENT HOW MANY TIMES YOU CHOSE TO REPORT A PEER!

[Screen 3–13: Question Task]

[Screen 14]
You answered the following number of questions correctly in this period: <score displayed>
The above number represents your actual score for this period. Please enter your Official Score. This score can be more than your actual score. However, if you report an Official Score that is higher than your actual score, your overstatement will be shown to one of your group members who will have an opportunity to report you.

[Screen 15 (shown if Misconduct Participant overstated score)]
One of your group members has overstated his/her score by the following amount: 2

[Strong Descriptive Norms:] Do you wish to report your group member? In a previous experiment similar to this one, most people OVERWHELMINGLY chose to report their group members.

[Weak Descriptive Norms:] Do you wish to report your group member? In a previous experiment similar to this one, most people OVERWHELMINGLY chose to NOT report their group members.

[Financial Incentive Reward:] Your pay will be increased by $1.00 if you choose to report your group member this period.

[Financial Incentive Penalty:] Your pay will be decreased by $1.00 if you choose to not report your group member this period.

Please type “Yes” if you wish to report him/her and “No” if you do not wish to report (capitalize first letter only): <Enter Response>

Remember, your decision will be publicly displayed and announced after the final period.
Panel B: Experimental instrument for the misconduct participants

The following presents the instructions for the misconduct participants, whose responses were only used to provide misconduct for the experiment:

This experiment consists of four periods. Each period begins with a game where you are asked to answer a set of questions. Once you have finished playing the game, the computer will tell you which questions you answered correctly. You will then be asked to submit a report of your score for payment. You will receive a fixed amount of $2 each period for playing the game, regardless of your actual score.

In addition, you will have an opportunity to earn a reward by over-reporting by a specific amount each period. If you agree, you will receive $2 each time you over-report by the requested amount. Furthermore, if you over-report every period by the amount requested, you will receive an additional reward of $4 once all the periods are complete.

Your over-reporting will be shown to your group members, who will have an opportunity to report you, but your identity will be kept secret. Whether your peers choose to report you or not, it will not affect your pay.

Once all periods are complete, anyone who has chosen to report a peer will be publicly listed with his/her Team color, code name, and the number of times he/she reported a peer shown on the computer screen for all to see. However, it will NOT say who was reported and over-reporting will NOT be publicly displayed.

Appendix 2

Payment schemes for reward, penalty, and misconduct participants

<table>
<thead>
<tr>
<th>Participant Type</th>
<th>Base Pay</th>
<th>Penalty Calculation</th>
<th>Total Pay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penalty participant pay</td>
<td>Base = $20</td>
<td>Penalty = (4 periods × $1.00) = $0 if report, −$4 if not report</td>
<td>$20 if report, $16 if not report</td>
</tr>
<tr>
<td>Reward participant pay</td>
<td>Base = $16</td>
<td>Reward = (4 periods × $1.00) = $4 if report, $0 if not report</td>
<td>$20 if report, $16 if not report</td>
</tr>
<tr>
<td>Misconduct participant</td>
<td>Base = (4 periods × $2) = $8 Over-report = (4 periods × $2) = $8 if over-report, $0 if not</td>
<td>Over-report bonus for all 4 = $4 if always over-report, $0 if not</td>
<td>$20 if always over-report, $8 if never over-report</td>
</tr>
</tbody>
</table>

References


