Whistleblowers, Amnesty, and Managerial Fraud: 
An Experimental Investigation*

January, 2016

Stephen J. Cotten  
Department of Economics and Finance  
University of Houston - Clear Lake  
Houston, TX 77058  
cotten@uhcl.edu

Rudy Santore**  
Department of Economics  
University of Tennessee  
Knoxville, TN 37996  
rsantore@utk.edu

JEL Codes: C72, C92, K42  

Keywords: Whistleblower, Leniency, Fraud, Experiment

ABSTRACT  
We examine the effects of increasing transparency and accountability within managerial teams, as well as granting amnesty to whistleblowers, on the incentive to commit fraud. We develop a theoretical model to investigate these issues and then use it as the foundation for a controlled laboratory experiment. Theory predicts that, although whistleblowing should not occur in equilibrium, the mere threat of whistleblowing can reduce fraud if managers in the team are sufficiently heterogeneous. In contrast, amnesty provisions are predicted to have no effect on behavior. Results from the laboratory experiment support many of the model’s predictions and also shed light on behavior not predicted by theory. For example, contrary to theory, amnesty provisions have the desirable effect of reducing unreported fraud.

*We would like to thank William Neilson and Christian Vossler for helpful comments. We would also like to thank the University of Tennessee for funding assistance.  
** Corresponding author.
1. INTRODUCTION

Throughout the 1980s and 1990s corporations embraced the recommendations of agency theory by utilizing equity compensation to incentivize managers (Murphy, 1999). However, in the aftermath of multiple high-profile corporate fraud scandals, such as Enron and Worldcom, both researchers and policy-makers were forced to acknowledge that accounting fraud may be an unexpected consequence of these compensation schemes. Lawmakers responded by passing the Sarbanes-Oxley Act (SOX) of 2002, which attempts to reduce fraud by strengthening corporate governance and reporting requirements. Two important provisions of SOX are the protection it provides to whistleblowing employees and the requirement that management establish and affirm the adequacy of their financial reports.¹ ²

In this paper, we consider the implications of increasing transparency and accountability, as well as granting amnesty to whistleblowers, on the incentives of managers in a team to commit fraud. We first develop a theoretical model that captures the essence of such policies and then test the theory in a laboratory setting. In our model, a team of two managers who each hold a share of equity in a firm sends a report regarding the firm’s financial health to the market. These reports may misrepresent the firm’s health (hereafter referred to as ‘fraud’) allowing one or both managers to manipulate the market price. If fraud is not reported by one of the managers, there is a possibility that it will be detected by regulators and both managers will be held accountable.

¹ As discussed by Watnick (2007), “[i]n attempting to reform American business practices, Congress impressed into service corporate officers, directors, and other corporate employees, enlisting them as ‘foot soldiers’ in the fight against corporate fraud. Congress did so by requiring those who witness corporate fraud to report what they know about it and by offering commiserate protection from retaliation under the ‘whistleblower protection’ provisions contained within Sarbanes-Oxley.”
² 18 U.S.C. 1350, “Failure of corporate officers to certify financial reports,” requires management to ensure, with threat of criminal penalties for noncompliance, that “information contained in the periodic report fairly presents, in all material respects, the financial condition and results of operations of the issuer.”
To examine the implications of requiring that managers “sign off” on financial statements, we allow for two different information environments. In the first environment, fraud is symmetric information so that managers observe whether fraud has taken place and are capable of blowing the whistle. In the second environment, fraud is asymmetric information so neither manager can blow the whistle. We also examine the effect of allowing whistleblowers to receive amnesty for reporting fraud, which of course is only possible in a symmetric information environment. Finally, we examine the interactions of the above policy features when sanctions impose different financial costs (e.g., reductions in lifetime earnings) and personal costs (e.g., public embarrassment) on each of the managers.

We demonstrate that the total fraud committed when fraud is symmetric information differs from the asymmetric case only if the disutility from being sanctioned differs across managers. This result is due to the fact that one manager’s fraud yields financial benefits to both managers as each holds an equity share in the firm and can sell that equity at an inflated price. When sanctions have the same impact on both managers they each prefer to commit the same amount of fraud, so there is no incentive to blow the whistle. Only if managers are sufficiently heterogeneous does the threat of blowing the whistle reduce the equilibrium amount of fraud.

The model also predicts that, other things equal, amnesty provisions for whistle blowers will have no effect on behavior. This follows from the equilibrium prediction that it is never optimal to choose a level of fraud that would induce the other manager to blow the whistle. These theoretical predictions, however, rest on the assumption that managers are rational and able to coordinate perfectly. In practice, coordination may be difficult, especially when managers are heterogeneous. Whether or not these predictions, and the assumptions on which they rest, describe actual behavior is ultimately an empirical question.
To address the empirical relevance of the model’s predictions we report the results of a laboratory experiment with treatments designed to implement the various policy scenarios. Testing these predictions with field data would be quite challenging, if not impossible, due to the practical limitations of field data. The laboratory provides several advantages in terms of both observability and control. First, in the lab we can observe all fraud that is committed, not merely the detected or reported fraud. Second, we can control both the probability of detection and the magnitude of sanctions, which will differ across firms and their management teams. Third, it is easy to investigate counterfactual environments, such as giving full amnesty to whistleblowers. Finally, a clean randomized experimental design allows a relatively simple and straightforward econometric analysis.

When there is no amnesty, our findings are consistent with the main predictions of the model regarding the impact of symmetric information and whistle blowing. For treatments in which managers are identical (that is, they face identical sanctions), the ability to observe the other’s fraud and blow the whistle does not have a statistically significant effect on total fraud. Whistle blowing does occur, but at a very low rate (5.6%). At the same time, for the treatments in which the managers face heterogeneous penalties, total fraud is significantly less when whistle blowing is possible, which is also consistent with theory. However, coordination is clearly difficult for subjects with at least one subject blowing the whistle in approximately one out of every four rounds.

The prediction that amnesty will have no impact on behavior is not supported by the data. Indeed, we find that providing amnesty to whistleblowers yields a significant reduction in the fraud that goes unreported. In addition, subjects blow the whistle significantly more often when there is amnesty, which helps to explain the decrease in unreported fraud.
There is a relatively small theoretical literature on whistleblowing within the context of corporate malfeasance and, to our knowledge, there is no experimental work on the topic. Heyes and Kapur (2009) develop a model of whistleblowing behavior in which agents may report environmental violations, although their results can be applied to other contexts. Unlike our model, in which whistleblowers act to avoid being punished for the managerial team’s actions, their model focuses on various motivations for whistleblowers to act to prevent harm to others. These authors find that the optimal enforcement policy is dictated in large part by the motivation of the potential whistleblower.

More recent theoretical work by Friebel and Guriev (2012) demonstrates that the mere potential for whistleblowing may yield inefficiency as an unintended consequence. In their model, upper management can manipulate earnings and mislead investors regarding the financial health of the firm. However, to prevent lower division managers from blowing the whistle, upper management must give lower management additional compensation. Friebel and Guriev show that the optimal contract reduces the division manager’s incentive to provide effort thus causing inefficiency.

Despite some similarities, our paper differs from Friebel and Guriev (2011) in a few important aspects. First, our managers hold identical positions within the managerial team and each is able to actively participate in the fraud, while Friebel and Guriev have two levels of manager and only top management can manipulate earnings. Second, unlike Friebel and Guriev, we do not allow the team to manipulate the compensation packages in order to prevent whistle-

---

3 Dyck et al. (2010) conduct an empirical examination of reported corporate fraud cases between 1996 and 2004 and find that a wide variety of actors act as whistleblowers.

4 There is also a body of work on whistle blowing in the context of anti-trust enforcement. Some examples are Aubert et al. (2006), Hinloopen and Soetevent (2008), and Beaton-Wells and Tran (2015).
blowing. Third, whistleblowing in our model is motivated by the potential for team punishment when fraud is detected, which can occur even if no manager blows the whistle. In contrast, Friebel and Guriev assume that a manipulation of earnings is detected only if a division manager blows the whistle.

2. The Model

Two risk-neutral managers are employed by a firm. Each manager receives identical compensation packages in the form of salary and a share of equity, \( \alpha \leq \frac{1}{2} \).\(^5\) We assume the compensation packages meet each manager’s participation constraint and for notational convenience normalize the salary to zero without loss of generality.

The true value of the firm, \( V^T \), is known to the managers but not to owners and investors. The managers issue a report on the firm’s value to the market, \( V^R = V^T + F \), where \( F \) is a non-negative fraudulent inflation of the true value.\(^6\) Each manager contributes to this report of the firm’s value and is able to fraudulently inflate the reported value of the firm. Manager \( i \)’s choice of fraud is denoted by \( f_i \), with total fraud being the sum of the two, \( F = f_i + f_j \). Thus, each manager is individually capable of inflating the perceived value of the firm without assistance from the other manager.

We wish to analyze two different information environments, each corresponding to a different strategic setting. In the \textit{symmetric information} scenario, each manager observes the

\(^5\) Although equity compensation is a common method of motivating managerial effort, it has also been shown to induce managerial fraud (see, for example, Goldman & Slezak, 2007). We do not model the managers’ incentives to provide effort because the results would be similar to single agent models, with higher levels of equity compensation inducing greater team effort, and detract from our focus on whistle-blowing.

\(^6\) We treat investors as naïve in this model. In a more general model where investors rationally anticipate some level of fraud, \( F^e \), managers would still have an incentive to commit fraud. This is because the amount investors are willing to pay for the firm would then be \( V^R - F^e = V^T + F - F^e \). Thus, there remains a marginal incentive to fraud even if it is anticipated.
other manager’s fraud prior to the report being issued to the public. In this scenario, a manager can contact authorities (blow the whistle) prior to the fraudulent information being released to the public. In the asymmetric information scenario, we assume that any fraud committed by one manager is not observed by the other manager. Since fraud is private knowledge it is impossible for one manager to blow the whistle on the other. We begin with a discussion of the symmetric information setting and then move on to the asymmetric information scenario.

**Symmetric Information - Whistleblowing**

In addition to actively engaging in illegal behavior, when a manager learns of the fraud committed by the other manager he or she may passively engage in fraud by not reporting it to authorities before the report is issued to the market. We assume that if a manager does ‘blow the whistle’ on the fraud, the ensuing investigation detects and prevents all fraudulent activity so that the managers do not profit from the fraud. In other words, if a manager blows the whistle investors observe $V^R = V^T$ and are willing to purchase shares of the firm at that valuation. If neither manager blows the whistle, the report is issued and the managers sell their equity stakes at the inflated market value.\(^7\) However, if fraud goes unreported there is a positive probability, $\rho$, that authorities will detect the fraud and the managers will be punished.\(^8\)

For symmetric information environments, both managers are presumed to have a fiduciary responsibility to owners and can be sanctioned for not reporting the fraud. If neither manager blows the whistle and fraud is later detected, the manager’s sanctions are based on the average

---

\(^7\) Bar-Gill and Bebchuk (2003) discuss liquidity motivations for the manager to sell before the true value is learned. These authors argue that incentives to misreport exist even if managers cannot sell their shares in the short term.

\(^8\) One can interpret $\rho$ as the probability that authorities have sufficient evidence to build a case against the managerial team, and $\rho$ is treated as exogeneous for simplicity. Making $\rho$ a function of fraud does not qualitatively change the results, as only the expected sanction matters.
level of fraud, $\frac{F}{2}$. Thus, a manager who does not commit fraud, but also does not blow the whistle, receives the same sanction as a manager who was responsible for the fraud. This assumption is consistent with the view that managers are responsible for reporting any malfeasance. On the other hand, if one or both managers blow the whistle, manager $i$’s punishment is a function of his or her individual choice of (attempted) fraud, $f_i$. Therefore, a whistleblower who did not commit fraud receives no punishment.

Let $X()$ denote the convex sanction function that determines the punishment to an individual manager if the fraud is detected or reported, where $X' > 0$, and $X'' > 0$. This sanction is a function of the fraud for which the manager is held accountable, which need not equal the fraud actually committed by the manager as discussed above. All managers receive disutility from being sanctioned, and a manager’s type determines his or her disutility. We assume that there are two possible types of manager and let $t_i = L, H$ denote manager $i$’s type. A type $L$ manager (referred to as a low) incurs disutility $\eta_L X()$ from being sanctioned while a type $H$ manager (referred to as a high) incurs disutility $\eta_H X()$ from being sanctioned, where $\eta_H$ and $\eta_L$ are constants and $\eta_H > \eta_L$. Manager types are common knowledge.

We allow for the possibility that, in exchange for revealing fraud, managers receive a reduced penalty. While the Sarbanes-Oxley Act provides protections from retaliation, reducing the cost to potential whistleblowers, it does not explicitly guarantee immunity or even leniency. Without partial or complete amnesty, managers who commit fraud may incur penalties for their

---

9 One could punish each manager based on the entire amount of fraud with no qualitative change in results.

10 For example, younger managers may suffer greater future pay losses than managers nearing retirement. Mittendorf (2008) uses a similar construction in analyzing shirking. Schmidt (2005) provides a non-technical discussion of how encouraging whistleblowing affects different types of managers; Cooter and Porat (2001) examine nonlegal sanctions that will typically differ across agents. Differences in risk-aversion would also yield similar results but would be analytically less tractable.
actions, despite the fact that they reported the wrongdoing. At the same time, blowing the 
whistle allows managers to protect themselves from being punished for the actions of others if 
they find the total fraud and expected sanction to be too great. To model this partial amnesty (or 
leniency), we let $\theta$ represent the proportional reduction in penalty for the whistleblower, where 
$0 \leq \theta \leq 1$. We assume that the amnesty is granted to both managers if they both blow the 
whistle. Thus, the utility cost for manager $i$ of type $t_i$ who blows the whistle is $(1 - \theta)\eta_i X(f_i)$, 
and the utility cost for manager $i$ of type $t_i$ when the only the other manager blows the whistle is 
$\eta_i X(f_i)$.

To summarize, the timing of events is as follows:

Stage 1: Each manager simultaneously and individually chooses a level of fraud.

Stage 2: Each manager observes the fraud chosen by the other manager in Stage 1 and decides 
whether or not to blow the whistle.

Stage 3(a): If one or both managers blow the whistle, each is held responsibility for his or her 
individual fraud levels. If there is an amnesty provision, any manager who blew the whistle 
receives a reduced sanction.

Stage 3(b): If neither manager blows the whistle, with exogenous probability, $\rho > 0$, the 
difference in reported and true value is determined to be fraudulent and each manager is held 
accountable for $F/2$.

It follows from the above discussion that the expected utility for manager $i$ of type $t$ 
ultimately depends on $i$’s choice of fraud, the fraud choice of other manager, and whether she or 
the other manager blows the whistle.
We focus on subgame perfect equilibria for our sequential game. Most of the equilibrium outcomes are fairly intuitive so the following discussion stresses intuition rather than formal proofs.

The first step in understanding the equilibrium outcomes is to recognize that each manager type has a preferred level of total fraud. Conditional on the other manager not blowing the whistle, this preferred level of fraud equalizes the marginal benefit of additional fraud, \( \alpha \), and the marginal disutility of additional fraud, \( \rho \eta t \Delta X' \left( \frac{f_i + f_j}{2} \right) \). We let \( P_t \) denoted the preferred level of fraud for a manager of type \( t \). It is easy to verify that \( P_L > P_H \); as one would expect, the type who incurs the greater disutility from being sanctioned prefers less total fraud.

It can be rational to blow the whistle if too much fraud has been committed (i.e. fraud is at a level at which the expected sanctions exceed the expected gains), so the threat of whistleblowing is credible. At the same time, it is never sequentially rational to choose a level of fraud that leads to whistleblowing when there is less than full amnesty (\( \theta < 1 \)). Consistent with our view of managers who are rational and commit fraud strategically, we ignore equilibria that are based on managers playing weakly dominated strategies. With full amnesty, for example, there exist equilibria in which both managers blow the whistle even if only one manager commits fraud, regardless of the magnitude. However, for the manager who did not commit fraud blowing the whistle is weakly dominated as long as the fraud is not too great. This is because this manager

\[
EU_i = \begin{cases} 
\alpha V_T - (1 - \theta)\eta t_i X(f_i), & \text{if manager } i \text{ blows the whistle} \\
\alpha V_T - \eta t_i X(f_i), & \text{if only manager } j \text{ blows the whistle} \\
\alpha (V_T + f_i + f_j) - \rho \eta t_i X' \left( \frac{f_i + f_j}{2} \right), & \text{if no manager blows the whistle}
\end{cases}
\]
would be no worse off if the offending manager blows the whistle (on himself) and strictly better off *ex ante* if the other manager does not blow the whistle.

The nature of the managers’ behavior ultimately depends on whether or not the managers are the same type. If both managers are of type $t$, then the equilibrium is for the managers to coordinate on their mutually preferred level of fraud so that $f_i^* + f_j^* = P_t$. While the individual fraud levels are not unique, a natural focal point would be a symmetric outcome: $f_i^* = f_j^* = P_t/2$.

For the case where the managers are different types, the high chooses zero fraud, $f_H^* = 0$, because he or she anticipates that the other manager will choose more fraud than she prefers. Since the high type manager chooses zero fraud in equilibrium, blowing the whistle entails no sanction for the high so this ‘threat’ is credible. Given that the high type chooses zero fraud in equilibrium, it is optimal for a high to blow the whistle if the expected sanction, $\rho_\eta_H X \left( \frac{f_L}{2} \right)$, exceeds the manager’s gain from the higher equity valuation, $\alpha f_L$. Therefore, the low must choose a level of fraud that is below the high type’s whistleblowing threshold, which we denote by $W_H(0)$.\(^{11}\) It follows that the low-type manager $j$ chooses the lesser of $P_L$ and the maximum fraud that would not induce the high to blow the whistle.\(^{12}\)

**Result 1 (Whistleblowing):** Suppose that the managers are the same type $t=L,H$, then the equilibrium fraud equals the level preferred by that type, $f_i^* + f_j^* = P_t$, and neither manager blows the whistle.

\(^{11}\) As such $W_H(0)$ is the whistleblowing threshold when the high commits zero fraud, which is the equilibrium prediction. More generally, a high type manager will blow the whistle if the utility from doing so, $-(1-\theta)\eta_H X \left( \frac{f_H}{2} \right)$, is greater than manager’s gain from the higher equity valuation less the expected sanction, $\alpha(f_L + f_H) - \rho_\eta_H X \left( \frac{f_L + f_H}{2} \right)$.

\(^{12}\) It can be shown that If $\eta_H$ and $\eta_L$ are sufficiently close in magnitude, then $P_L < W_H$. 

---

10
**Result 2** *(Whistleblowing)*: Suppose that the two managers are different types, then the high-type manager chooses zero fraud $f_H^* = 0$, the low-type manager chooses $f_L^* = \min\{P_L, W_H\}$, and neither manager blows the whistle.

Note that the level or existence of an amnesty provision does not alter the equilibrium predictions. If managers are the same type, then they are predicted to coordinate on their preferred levels of fraud. And if they are different types, then the high-type manager is predicted to choose zero fraud, so his or her sanction from blowing the whistle is the same (zero) with or without amnesty.

**Asymmetric Information – No Whistleblowing**

In contrast to the above, we now consider the possibility that fraud is asymmetric information, with neither partner observing the other’s actions. One could also interpret the model in this section as describing behavior when there are significant personal reputation and career cost of being involved in a scandal, regardless of one’s personal actions. Although not formally modelled, it is easy to show that if these personal reputation costs are large relative to the expected sanction, blowing the whistle would not be optimal. Regardless of the interpretation, the only change in the model is that now both managers know that whistleblowing is not possible. If fraud is detected each manager is held responsible for one half of the total fraud. Thus, the payoff to a manager is simply a function of the total fraud:

$$EU_i = \alpha(V^T + F) - \rho\eta_i X \left(\frac{F}{2}\right).$$

(2)

where $= f_i + f_j$. Clearly the optimal levels of fraud for the two manager types are different and are the same as in the symmetric information version of the model, namely, $P_L$ and $P_H$. 

The equilibrium predictions are straightforward. If both managers are the same type \( t (= L, H) \), then they are predicted to coordinate on their most preferred level of fraud: \( f_i^* + f_j^* = P^t \). However, if they are different types, then the high type should choose zero fraud, \( f_H^* = 0 \), while the low type chooses his or her preferred level, \( f_L^* = P_L \).

**Result 3 (No Whistleblowing):** Suppose that the managers are the same type \( t = L, H \), then the equilibrium fraud equals the level preferred by that type, \( f_i^* + f_j^* = P_t \).

**Result 4 (No Whistleblowing):** Suppose that the two managers are different types, then the high-type manager chooses zero fraud \( f_H^* = 0 \) and the low-type manager chooses \( f_L^* = P_L \).

3. Experimental Design

We conduct a series of laboratory experiments to test the predictions of our model and explore the effects of behavior that is inconsistent with the rational agents postulated by our theory. With field data, we can typically observe only the fraud that is detected, while in a controlled laboratory setting, we are able to observe all fraud and control for other characteristics that are exogenous in real world settings.

The experiments are designed to test the effect of changes in three main treatment variables on the fraud chosen by members of a managerial team: whistleblowing (information symmetry), amnesty, and heterogeneous managerial preferences. Hereafter, the term “amnesty” should be understood to mean “full amnesty” \( (\theta = 1) \) corresponding to the experimental design that we utilized. In addition, we use the following abbreviated notation: Two Low-Type Managers (LL), One High One Low Type Manager (HL), No Whistleblowing (NW), Whistleblowing (W),
Amnesty (A), and No Amnesty (NA). Because amnesty applies only when there symmetric information and the possibility of whistle-blowing, our full-factorial design involves six treatments. Two of the treatments (LL-NW and HL-NW) do not allow whistleblowing and differ only in the manager composition. The remaining four (LL-W-NA, HL-W-NA, LL-W-A, HL-NW-NA) allow whistleblowing and vary both amnesty and manager type composition.

A total of 180 subjects recruited from the general population at the University of Tennessee participated in one of 12 experimental sessions conducted at the Experimental Economics Laboratory. There were two sessions (replications) of each treatment, with 16 subjects in the first and 14 subjects in the second (N=30 per treatment). Approximately 82% of subjects had previously participated in an economics experiment. Average earnings were roughly $20.00, and sessions lasted for approximately one hour on average.

In all six treatments, subjects played the role of a manager in a group with one other manager. Subjects were given a project with a true value of $100, but were given the opportunity to fraudulently inflate that value by $0, $20, $40, $60, $80, or $100 units. Both subjects in a group simultaneously chose how much fraud to commit, the sum of which was added to the actual value to form the reported value. Subjects stood to earn 50% of the reported value, penalties for fraud notwithstanding. There was then a 25% chance\(^\text{13}\) that any fraudulent inflation would be detected and penalties would be levied on both managers based on the total amount of fraud committed.

In the symmetric information treatments, an intermediate stage was inserted where managers were given the option of preventing the fraud (whistleblowing) after the total had been revealed.

\(^{13}\) Five rounds (25% of the twenty) were randomly chosen in advance for the fraud to be detected, to hold the timing of detection rounds constant across all treatments.
but before any detection would occur. If either manager chose to blow the whistle, the fraud was cancelled out but each manager was punished based on their individual choice of fraud.\textsuperscript{14} In the amnesty treatments, a whistleblowing manager was exempt from penalty. Finally, in treatments with different manager types, subjects were randomly assigned a type, \textit{high} or \textit{low}, and remained that type for the entirety of the experiment while being randomly matched with a player of the different type each round.

Sanctions for each level of fraud were chosen discretely to give clean integer predictions while following the assumptions and structure of the theory. Subjects were given tables with payoffs for all possible outcomes in the game. Instructions for all treatments, as well as these tables (from which all parameters can be determined), are located in an appendix.

Decisions were made via computer. The experiments were programmed and conducted with the software \textit{z}-Tree (Fischbacher, 2007). The software collected all decisions and made all relevant earnings calculations. Written instructions were provided to each participant and displayed on-screen. The experiment moderator read instructions aloud, one screen at a time, and any questions were answered prior to proceeding to the next instruction screen. During the instructions, participants were asked to answer four questions on the computer to assess their understanding of how earnings are calculated, and were unable to proceed until the questions were correctly answered. Those who were unable to perform the calculations on their own were able to ask an experiment moderator for help, who would then re-explain procedures and field questions. Following the instructions, there was one unpaid practice round, where questions were encouraged and addressed. Upon conclusion of the experiment, a short questionnaire was

\textsuperscript{14}To avoid context effects, the problem was not framed directly as fraud. Managers were asked to deviate from an actual value to form a reported value, and were penalized based on the deviation.
administered to obtain demographic information and qualitative assessments of the experiment design and instruction clarity.

The experiment was parameterized such that $p^L = 80$ (risk-neutral low-type managers maximize their expected utility when there are 80 units of total fraud) and $W^H(0) = 40$ (a risk-neutral high-type manager who has committed 0 units of fraud will tolerate a maximum of 40 units of total fraud before blowing the whistle). For the symmetric information setting, some of the fraud that is committed may be reported. Thus, we consider both attempted fraud, which is the fraud each manager chose in the first stage of the game, and unreported fraud, which is fraud that was not reported. Of course, for the asymmetric information case where whistleblowing is not possible, attempted and unreported fraud are the same.

4. Results

The following analysis proceeds in two stages, with each utilizing a 2x2 design. In the first stage, we compare fraud in the asymmetric and symmetric information environments under the assumption that whistleblowers receive no reduction in sanctions (no amnesty). In the second stage, we focus on the symmetric information environment and investigate the effect of amnesty on both fraud and whistle-blowing.

Design 1: Varying Information Environments

Our first set of hypotheses is that allowing a manager to observe the other’s fraud and blow the whistle, other things equal, should reduce total fraud if and only if the managers are different types.

**Hypothesis 1(A):** Total fraud in $HL-W-NA$ will be less than in $HL-NW$. 
Hypothesis 1(B): Total fraud in LL-W-NA will be the same as in LL-NW.

In Table 1 we report the theoretical risk-neutral point fraud predictions and observed attempted and unreported fraud levels for the four treatments in which there is asymmetric information (no whistle-blowing) or symmetric information (whistle-blowing) with no amnesty. Unreported and attempted fraud are necessarily the same for the no whistleblowing treatments.

Table 1: Total Fraud Predictions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>NW</th>
<th>W-NA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Prediction</td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td></td>
<td>$F^* = 80$</td>
<td>$F^* = 80$</td>
</tr>
<tr>
<td></td>
<td>Attempted Fraud</td>
<td>89.40 (2.51)**</td>
<td>91.60 (3.28)***</td>
</tr>
<tr>
<td></td>
<td>Unreported Fraud</td>
<td>89.40 (2.51)***</td>
<td>84.67 (2.65)</td>
</tr>
<tr>
<td>HL</td>
<td></td>
<td>$F^* = 80$</td>
<td>$F^* = 40$</td>
</tr>
<tr>
<td></td>
<td>Attempted Fraud</td>
<td>79.73 (4.70)</td>
<td>51.80 (3.99)***</td>
</tr>
<tr>
<td></td>
<td>Unreported Fraud</td>
<td>79.73 (4.70)</td>
<td>40.93 (3.30)</td>
</tr>
</tbody>
</table>

Standard errors are in parentheses. Observed values are group means over all periods. N=15 for each treatment cell. $H_0$: Predicted = Observed. *, **, and *** represent a significant difference using a t-test at the 10%, 5%, and 1% levels respectively. Results are qualitatively similar with non-parametric tests.

Hypothesis 1(A) is supported by comparing the HL-W-NA treatment to the HL-NW treatment. Examining both attempted and unreported fraud, fraud is significantly lower when there is both a high and low manager and whistleblowing is allowed ($p=0.00$ for both). Hypothesis 1(B) is supported by comparing the LL-W-NA treatment to the LL-NW treatment. For neither attempted fraud ($p=0.58$) nor unreported fraud ($p=0.28$) is there a statistically significant difference.

The theoretic predictions of observed behavior do a reasonable job of anticipating subject behavior. The observed attempted fraud in three of the four treatments is greater than the predicted values, and these differences are statistically significant, but the largest departure is 30% of the predicted value. The only exception is treatment HL-NW which was almost exactly in line with the prediction of 80. However, the observed unreported fraud in the NA treatments
is closer to the predicted values and the differences are not statistically significant with the exception of $LL$-$NW$. In sum, the theoretical predictions are well supported by our experimental results. While the attempted fraud was observed to be a bit higher than predicted, much of this excess fraud was reported by one or both of the managers.

*Design 2: Varying Whistleblower Amnesty*

We now focus exclusively on the symmetric information (whistleblowing) treatments and compare behavior as we vary the punishment regime. In one regime managers face full punishment for their role in the fraud even if they blow the whistle, and in the other whistleblowers receive complete amnesty. Recall that whistleblowing is not predicted to occur in equilibrium, because it is irrational to commit a level of fraud high enough to cause other managers to report it to authorities. However, avoiding whistleblowing requires, in part, coordinating on fraud levels correctly and knowing the preferences of other managers. In practice such coordination is not likely to be perfect and some whistleblowing would be expected. Table 2 provides the predicted and observed fraud levels with and without amnesty.

*Table 2: Total Fraud Predictions*

<table>
<thead>
<tr>
<th></th>
<th>Prediction</th>
<th>$W$-$NA$</th>
<th>$W$-$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F^* = 80$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$LL$</td>
<td>$91.60$ $(3.28)^{***}$</td>
<td>$85.40$ $(1.79)^{***}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$84.67$ $(2.65)$</td>
<td>$51.67$ $(4.28)^{***}$</td>
<td></td>
</tr>
<tr>
<td>$HL$</td>
<td>$51.80$ $(3.99)^{***}$</td>
<td>$94.33$ $(4.64)^{***}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$40.93$ $(3.30)$</td>
<td>$19.47$ $(4.45)^{***}$</td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are in parentheses. Observed values are group means over all periods. N=15 for each treatment cell. $H_0$: Predicted = Observed. *, **, and *** represent a significant difference using a t-test at the 10%, 5%, and 1% levels respectively. Results are qualitatively similar with nonparametric tests.
The observed attempted fraud in all four treatments is greater than the predicted values, and these differences are statistically significant (with \( p \)-values ranging from 0.00 to 0.02). The observed unreported fraud levels in the No Amnesty treatments are closer to the predicted values and the differences are not statistically significant. Interestingly, in the amnesty treatments, \( LL-W-A \) and \( HL-W-A \), the unreported fraud levels are below the predicted values (\( p\)-value = 0.00 for both). The differences between the attempted and unreported fraud values are much larger in the amnesty treatments because of a much greater rate of whistleblowing – a behavioral effect not predicted by theory. We return to this observation when we consider the rates of whistleblowing in each of the four treatments.

In addition to the point predictions for total fraud, Table 2 also illustrates qualitative treatment effects with corresponding testable hypotheses. One set of hypotheses is that, relative to when both managers are lows, total fraud will be less when one of the managers is a high.

**Hypothesis 2(A):** Total fraud in \( HL-W-NA \) will be less than in \( LL-W-NA \).

**Hypothesis 2(B):** Total fraud in \( HL-W-A \) will be less than in \( LL-W-A \).

Examining attempted fraud, we find mixed support for Hypothesis 2. Introducing heterogeneous managers decreases attempted fraud without amnesty (91.60 vs. 51.80, \( p = 0.00 \)) but increases attempted fraud (85.40 vs. 94.33, \( p = 0.06 \)) with amnesty. This latter result may likely be explained by the fact that some players, particularly high-type managers, committed fraud anticipating that they would blow the whistle. The results are stronger when we look at unreported fraud with both treatment effects as expected (84.67 vs. 40.93, \( p=0.00 \) and 51.67 vs. 19.47, \( p = 0.00 \)).
Another set of hypotheses derived from theory is that, other things equal, amnesty will not affect the total amount of fraud:

**Hypothesis 3(A):** Total fraud in $LL-W-NA$ will be the same as in $LL-W-A$.

**Hypothesis 3(B):** Total fraud in $HL-W-NA$ will be the same as in $HL-W-A$.

These hypotheses are easily rejected for unreported fraud ($p=0.00$). Hypothesis 3(B) is also rejected for attempted fraud ($p=0.00$). We are unable to reject Hypothesis 3(A) for attempted fraud ($p=0.18$), although it did fall with amnesty as in the other three comparisons.

According to theory, whistleblowing will not occur in equilibrium. Table 3 shows the frequency with which whistleblowing was observed at the group level; specifically, the frequency that either one or both managers in a given group reported the fraud. With the exception of $LL-W-NA$, fraud was reported rather frequently in each of the other three treatments. While the theory clearly fails with regard to the non-observance of whistle-blowing, it is instructive to investigate whether it fails in a systematic manner.

### Table 3. Frequency Of At Least One Whistleblower In A Group

<table>
<thead>
<tr>
<th></th>
<th>W-NA</th>
<th>W-A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whistleblowing Rate (total)</td>
<td>5.6% (1.4%)</td>
<td>36.3% (4.2%)</td>
</tr>
<tr>
<td>Whistleblowing Rate (lows)</td>
<td>2.8% (0.7%)</td>
<td>22.5% (3.0%)</td>
</tr>
<tr>
<td><strong>HL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whistleblowing Rate (total)</td>
<td>27.0% (2.9%)</td>
<td>73.0% (3.4%)</td>
</tr>
<tr>
<td>Whistleblowing Rate (lows)</td>
<td>13.3% (1.8%)</td>
<td>55.7% (3.4%)</td>
</tr>
<tr>
<td>Whistleblowing Rate (highs)</td>
<td>14.3% (3.2%)</td>
<td>52.7% (4.3%)</td>
</tr>
</tbody>
</table>

The total rate represents the rate at which a group had at least one whistleblower. The breakdown for lows and highs represents the rate at which individuals of each type blew the whistle. Standard errors are in parentheses.
The prediction that whistleblowing will not occur is based on the implicit assumption that managers are able to coordinate perfectly. However, in practice, coordination among managers is likely to be more difficult when they are different types (that is, if they face different sanctions). For this reason, we expect to see more whistle blowing for these treatments, which leads to our next hypothesis.

**Hypothesis 4(A):** Whistleblowing will be more frequent in HL-W-NA than in LL-W-NA.

**Hypothesis 4(B):** Whistleblowing will be more frequent in HL-W-A than in LL-W-A.

We can see in Table 3 that whistleblowing is indeed more frequent when the managers are different types ($p=0.00$ for both). Interestingly, high and low managers blow the whistle at roughly the same rate in the $HL$ treatments.

It is also evident from Table 3 that whistleblowing is more frequent, other things equal, when there is amnesty and these differences are statistically significant. Specifically, we can state two additional findings that are not predicted by theory:

**Finding 1:** Whistleblowing is more frequent in LL-W-A than in LL-W-NA ($p=0.00$).

**Finding 2:** Whistleblowing is more frequent in HL-W-A than in HL-W-NA ($p=0.00$).

Although we do not formally extend the model to allow for non-standard preferences, prospect theory (Kahneman and Tversky, 1979) provides one explanation for why subjects are more likely to blow the whistle when there is amnesty. Prospect theory proposes that agents are risk seeking when the safe option is in the loss range, but risk averse when the safe option is at the reference point (no loss or gain). When there is no amnesty, a manager who has committed
fraud essentially must choose between a certain loss (the sanction if she blows the whistle) or a gamble that the fraud will not be detected. Whereas with amnesty, blowing the whistle is a safe option that entails no loss. Prospect theory would thus predict whistle blowing to be more attractive under amnesty.

**Table 4**: Comparison of attempted fraud for highs and lows.

<table>
<thead>
<tr>
<th></th>
<th>HL-NW</th>
<th>HL-W-NA</th>
<th>HL-W-A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lows Average (Std Error)</strong></td>
<td>51.27 (1.71)</td>
<td>28 (1.94)</td>
<td>48.33 (3.04)</td>
</tr>
<tr>
<td><strong>Highs Average (Std Error)</strong></td>
<td>28.47 (4.87)</td>
<td>23.8 (2.49)</td>
<td>46 (3.29)</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>-22.8***</td>
<td>-4.2*</td>
<td>-2.33</td>
</tr>
</tbody>
</table>

Clustered standard errors are in parentheses. Observed values are means of individual whistle-blowing, clustered based on subject ID. N=30 for each treatment cell. H_0: Difference = 0. *, **, and *** represent a significant difference using a t-test at the 10%, 5%, and 1% levels respectively.

As a final exercise, we now explore whether the manager-types behaved in accordance with our theoretical predictions. Table 4 provides the average attempted fraud choices for highs and lows in each of the *HL* treatments.\(^{15}\) Theory predicts that in every *HL* treatment, the high manager should produce zero fraud, but that was not observed; high-type managers chose non-trivial levels of fraud in all three treatments. Failure to find support for this point prediction, however, is not surprising. High-type managers face a decision for which it is only possible to err on one ‘side;’ that is, by choosing positive fraud. Nevertheless, we can test the qualitative prediction that the lows should commit more fraud than the highs in each *HL* treatment.

**Hypothesis 5**: In all *HL* treatments, the lows commit more fraud than the highs.

\(^{15}\) Our focus on individual behavior motivates the use of attempted fraud, but the qualitative results are unchanged if we use unreported fraud.
The relative behavior of highs and lows is qualitatively consistent with the theory in two of the three treatments. Table 4 shows that consistent with Hypotheses 5, the lows committed more fraud than the highs in both $HL-NW (p < 0.01)$ and $HL-W-NA (p < 0.10)$. However, for treatment $HL-W-A$ there was no difference between types, contrary to Hypothesis 5.

5. Conclusion

Large-scale corporate frauds often require the tacit, if not explicit, cooperation of multiple agents within a firm. It is somewhat ironic that equity compensation, intended to align the interests of managers and owners, may not only encourage intentional misrepresentation of the firm’s true financial status but also discourage managers from aggressively monitoring and/or reporting such behavior. However, policies that increase transparency and accountability, such as those that require management to affirm the accuracy of financial statements, can counteract these incentives.

In theory, as long as fraud is symmetric information among managers the mere threat of having the whistle blown can curtail fraud, at least when manager are heterogeneous in their willingness to tolerate such behavior. Somewhat interestingly, granting amnesty to whistleblowers is not predicted alter the frequency of whistleblowing or the equilibrium level of fraud. These predictions, premised on perfectly coordination, are only partially supported by our experimental results.

On the one hand, consistent with these predictions, we find a significant drop in unreported fraud for treatments where a one manager can observe the other manager’s behavior and the managers receive different punishments. On the other hand, contrary to theory, we find that
granting amnesty to whistle blowers dramatically reduces unreported fraud, primarily because managers are more likely to report it.

Our experimental results thus lend empirical support for policies that increase transparency and accountability among members of a managerial team. At the same time, they make a stronger case for leniency provisions by identifying positive behavioral effects in scenarios where standard theory suggests there should be none. Together this suggests that modifying reporting requirements can be an effective complement to audits in combating corporate fraud.
References


