Leaks and Internal Communication*

Snehal Banerjee† Peicong (Keri) Hu‡ Taejin Kim§

April 2025

Abstract

We study how whistleblower policies affect a firm's internal communication, public information and expected value. An informed manager, who can divert cash for private benefit, manipulates internal information available to the employee to cover up evidence of his misconduct. Given her information, the employee chooses whether to blow the whistle and an action which affects internal alignment. Larger whistleblower rewards for the employee and harsher penalties for the manager reduce misconduct but also worsen internal communication. Public information quality worsens with harsher penalties but improves with larger rewards. Finally, when misalignment is sufficiently costly, such policies decrease firm value.

JEL: G10, G12, G14, G32
Keywords: leaks, whistleblower, disclosure, internal communication, efficiency

*We thank Jeremy Bertomeu, Bradyn Breon-Drish, Jesse Davis, Ron Dye, Elia Ferracuti, Michael Fishman, Naveen Gondhi, Xu Jiang, Doron Levit, Sivan Frenkel, Wouter Dessein, Lin Nan, Chao Tang, Jing Li, Yan Xiong, Robert Goex, Alfred Wagenhofer, Evgeny Petrov, and participants at the SFS Cavalcade (2018), the China International Conference in Finance (2018), the European Finance Association Annual Meeting (2018), the Econometric Society European Summer Meeting (2018), the Cambridge Corporate Finance Theory Symposium (2018), Accounting and Economics Society Asia-Pacific Webinar (2024), the Inaugural AES Annual Conference (2024), the Stanford Institute for Theoretical Economics Meeting (2024), the SFS Cavalcade Asia-Pacific (2024), DAR&DART Accounting Theory Seminar (2025), and seminar participants at Seoul National University, KAIST, Yonsei University, and Korea University for helpful comments and discussions.

†Email: snehalb@umich.edu. Ross School of Business, University of Michigan.
‡Email: kerihu@hku.hk. Faculty of Business and Economics, The University of Hong Kong.
§Email: taejink@korea.ac.kr. Korea University Business School, Korea University.

1 Introduction

Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman.

— Louis Brandeis, Other People's Money – and How Bankers Use It (1914)

Regulators encourage public whistleblowing to deter wrongdoing by firms. For example, the Securities and Exchange Commission (SEC)'s Dodd-Frank Whistleblower Program is designed to improve stakeholder protection and provide incentives to whistleblowers in financial markets.[1] Since inception, the program has awarded more than \$2.2 billion to 444 whistleblowers, and has led to SEC enforcement actions with remedies in excess of \$6 billion. In fiscal year 2024 alone, the SEC's Office of the Whistleblower received over 24,980 tips and awarded over \$255 million as part of the program.

While these policies are arguably effective at detecting unlawful behavior of firm insiders, they can also have unintended consequences. By increasing the likelihood that internal communication is leaked publicly, these policies may distort how a firm's employees communicate with each other within the firm. Salient examples include executives engaging in the "spiral of silence" at Enron, policies discouraging written documentation of consumer complaints at Tesla, and changes to internal communication policies at Meta in response to the leak of the Facebook Files by whistleblower Frances Haugen.[2] Because internal communication and information sharing are critical for firm performance, the overall impact of higher whistleblower incentives on real (allocative) efficiency and firm value is not clear.

We show that while increasing whistleblower rewards and disciplinary penalties reduce the incidence of misconduct, they worsen the firm's internal information environment. Im-

[1] The SEC Whistleblower Program is a recent addition to a number of other, similar financial market regulations (e.g., US Whistleblower Protection Act (1989), Section 806 of Sarbanes-Oxley Act (2002), Whistleblower Protection Enhancement Act (2007)). See the U.S. Securities and Exchange Commission, 2023 Annual Report to Congress on the Dodd-Frank Whistleblower Program (https://www.sec.gov/files/fy23-annual-report.pdf) for more details.

[2] See Taylor-Neu, Rahaman, Saxton, and Neu (2024) for an analysis of senior executive emails from Enron, McDade (2023) about policies at Tesla, and Heath (2021) about the information "lockdown" at Meta.

portantly, this can lead to lower real efficiency, especially when alignment within the firm and, consequently, effective internal communication are sufficiently important for firm value. Our analysis also suggests that such policies can have different implications for public information quality: while larger rewards improve public information quality, harsher penalties can worsen it.

Model Overview and Intuition. There is a single firm with a manager (he) and an employee (she). The manager chooses how much of the firm's internal cash to divert for private benefit. This misconduct reduces the firm's productivity, on average, and generates an internal noisy, "warning" signal. Importantly, the manager can engage in costly manipulation of this warning signal to cover up evidence of his misconduct.

Given this internal information, the employee takes an action to maximize firm value by increasing the alignment between her action and the firm's fundamentals. Moreover, she can "blow the whistle" by leaking the warning signal to the regulator at a private cost.[3] Conditional on a leak, the regulator (i) penalizes the manager in proportion to the expected impact of his misconduct, and (ii) provides a whistleblower reward to the employee. The firm's terminal value depends on both productivity and the degree of alignment between fundamentals and the employee's action.

In this setting, we show that increasing whistleblower rewards for the employee and penalties for the manager both lead to (i) lower misconduct by the manager and (ii) worse internal information quality. However, we show that their impact on public information quality is different: while an increase in rewards leads to more public information on average, harsher penalties lead to less informative public information. Moreover, the overall effect of these policies can be to reduce real efficiency, as measured by expected firm value.

To see why, note that the manager's cash diversion decision trades off the private benefit from stealing cash against the expected cost from whistleblower penalties, conditional on a leak. As a result, whistleblower policies have two effects. First, harsher penalties and larger

[3] The cost can arise from adverse actions against the employee as a result of retaliation or the efforts that she spares to collect evidence for whistleblowing.

rewards (which increase the likelihood of a leak) both lead to an increase in the expected cost of diverting cash, and so lead to less misconduct and higher firm productivity. Second, for a given level of misconduct, both these policies increase the manager's incentive to cover up his wrongdoing by manipulating the internal signal. This leads to less informative decisions by the employee and, consequently, less alignment between her action and fundamentals.

The above implies that either type of policy (harsher penalties or larger rewards) will worsen internal communication. However, the two policies have different effects on the incidence of leak and informativeness of the information, which constitutes public information in our setting. Specifically, while harsher penalties for the manager induce more manipulation, they do not directly affect the probability of whistleblowing by the employee.[4] As a result, such changes always reduce the quality of public information. In contrast, larger whistleblower rewards increase the likelihood of leaks, in addition to increasing the extent of manipulation. We show that the first effect dominates, so that higher whistleblower rewards always lead to more informative public information.

We then characterize how such policies affect real efficiency, as measured by expected firm value. When the impact of internal alignment on overall firm value is relatively small, the first effect dominates, and so firm value increases with whistleblower incentives. However, when internal alignment is sufficiently important for firm value, the latter effect can dominate and real efficiency decreases with stronger whistleblower rewards and penalties, when these incentives are sufficiently large. In the latter case, there is a continuum of optimal policy choices (of reward-penalty pairs) which maximize expected firm value.

A common feature of whistleblower rewards in practice is that they are proportional to the regulators monetary sanctions collected from enforcement actions.[5] Our analysis implies

[4]In our main model, the likelihood of whistleblowing depends only on the whistleblower reward, which is constant, and the personal cost incurred by the employee. However, in Section 6.2, we consider a setting in which the whistleblower reward, and consequently, the likelihood of a leak, depend on the expected misconduct, given the internal communication. However, we show our results are qualitatively similar in this case.

[5]For instance, under the SEC Whistleblower Program, whistleblowers are entitled to an award between 10 - 30% of the collected sanctions — see www.sec.gov/whistleblower.

that in such cases, the optimal policy pair which maximizes expected firm value is uniquely pinned down. Specifically, if the fraction of penalties that are paid out as rewards increases, the optimal level of penalty should be lowered. Moreover, the optimal intensity of penalties also varies across firms: it is lower for firms with more volatile fundamentals and for those in which the relative value of internal alignment is higher.

Given the potential adverse effects of whistleblower penalties on real efficiency and public informativeness, we then consider a setting in which the manager does not face any regulatory penalties, conditional on a leak. Instead, we assume that the manager's compensation depends on the firm's stock price, which reflects the market's expectation of terminal cash-flows, conditional on public information (including the outcome of an employee leak, if there is one). We show that the disciplinary role of stock prices is analogous to whistleblower penalties in our main analysis. Specifically, an increase in stock price sensitivity leads to less misconduct, but worse internal and public communication, and can lead to lower expected firm value when the costs of misalignment are sufficiently high. This suggests that the qualitative implications of our analysis are robust to specific assumptions on managerial payoffs. Moreover, the analysis implies that larger whistleblower bounties are more likely to be counterproductive especially for firms with high-powered, stock-based managerial incentives.

Our analysis suggests a novel channel through which regulatory policy towards whistleblowers can affect firm behavior. We show that while stronger whistleblower incentives can lead to less misconduct by firm managers, this need not always lead to higher firm value or improve real efficiency because it can worsen internal communication. Moreover, our analysis highlights that excessive whistleblower incentives can have a negative impact especially for firms in which internal alignment is an important component of value. This implies that such policies can have qualitatively different effects on firms across different sectors and industries and care must be taken in accounting for these differences when evaluating the overall impact of such regulations.

The rest of the paper is organized as follows. The next section briefly discusses the related literature and the paper's contribution. Section 3 presents the model and discusses the key assumptions, while Section 4 characterizes the equilibrium. Section 5 presents the analysis of how whistleblower rewards and penalties affect various measures of efficiency, including the informativeness of internal and public information and expected firm value. Section 6 explores the robustness of our results by considering (i) a setting in which the manager does not face any penalties, but is instead compensated by the stock price, and (ii) a setting in which the employee's whistleblower reward depends on the extent of managerial misconduct. Section 7 provides a discussion of the model's implications and concludes. All proofs and additional analysis are in the Appendix.

2 Related Literature

Our model builds on the seminal work of Dye (1988) and Fischer and Verrecchia (2000), and the subsequent contributions of Frankel and Kartik (2019) and Ball (2025), who study a sender's incentives to engage in costly manipulation when communicating with a receiver. Unlike this existing literature, the manager in our model engages not only in communication, but also in cash-flow diversion, which affects the distribution of fundamentals (firm productivity).[6] As a result, the incentives for communication distortion and misconduct are interdependent and both depend endogenously on the expected incidence of whistleblowing. Importantly, our results show that these incentives can move in opposite directions: policies that deter misconduct can lead to more distortion in communication, and the overall impact on real outcomes depends on the relative impact of the two.

Our work contributes to the growing literature on the impact of whistleblower incentives on firm decisions.[7] The most closely related paper is Nan and Zheng (2023) where a manager probabilistically detects a product defect and decides whether to share this information

[6] One could instead have assumed that the internal communication was in the form of cheap talk. An earlier version of this paper considered such a case, and showed qualitatively similar results about the trade-off between misconduct and internal communication, but the resulting model was far less tractable.

[7] There is a related but distinct literature on leniency programs and whistleblowing incentives to deter collusion in cartels (e.g., see Motta and Polo (2003), Spagnolo (2004), and Aubert, Rey, and Kovacic (2006)).

with an employee. In turn, the informed employee chooses either to fix the defect, thereby improving welfare, or to blow the whistle about the defect, which negatively impacts the stock price. The paper shows that when whistleblower incentives are very strong, the employee is very likely to blow the whistle, which discourages the manager from reporting defects and, therefore, reduces welfare.

We view our analysis as complementary and highlighting a different mechanism. As in their setting, stronger whistleblowing incentives can reduce firm value. In our setting, this occurs because such policies lead the manager to distort his internal communication, which leads to less informed decisions by the employee, and consequently, potentially reduces firm value. However, unlike their setting, higher whistleblowing incentives in our model also reduce the manager's incentives to divert cash flows, which can enhance firm value. Our analysis characterizes how the overall impact of whistleblower incentives on firm value and real efficiency depends on the relative importance of these effects. We also examine how such incentives affect public informativeness, which is particularly relevant, given that investor protection is the main motivation of the SEC's enactment of the whistleblower program.[8]

More broadly, our paper adds to the recent literature that points out how stronger whistleblower incentives can have negative effects on social welfare. For instance, Friebel and Guriev (2012) show that in a setting where the manager can bribe the employee to deter whistleblowing, higher whistleblower rewards can lead to less effort by the employee and lower real efficiency. In contrast, the manager manipulates internal communication in our setting to mitigate the adverse impact of a leak, but this reduces real efficiency by worsening the information available to the employee. Nan, Tang, and Zhang (2024) show that increasing whistleblower incentives leads insiders to leak information more frequently, which can make such reports less informative about the actual incidence of a fraud.[9] Nan, Tang, and

[8]See the U.S. Securities and Exchange Commission, 2017 Annual Report to Congress on the Dodd-Frank Whistleblower Program (https://www.sec.gov/files/sec-2017-annual-report-whistleblower-program.pdf).

[9]In related work, Givati (2016) considers a setting in which the whistleblower incurs a personal cost and can make false reports which succeed with some probability. The paper shows that the optimal level of whistleblower rewards is increasing and then decreasing in both the personal cost and the probability of a false report succeeding. Similarly, Buccirossi, Immordino, and Spagnolo (2021) characterize how the interaction

Ye (2025) extends this framework to study how an increase in whistleblower incentives can reduce audit quality, misstatement detection, enforcement by regulators and, consequently, social welfare. This is because, in their setting, the insider chooses to disclose less negative information when incentives are stronger. In contrast, while the likelihood of whistleblowing also increases with rewards in our setting, the quality of the information that is leaked is unaffected (it only depends on how much manipulation the manager engages in). As a result, higher rewards lead to more informative public information in our setting but can lead to less informative public information in theirs.

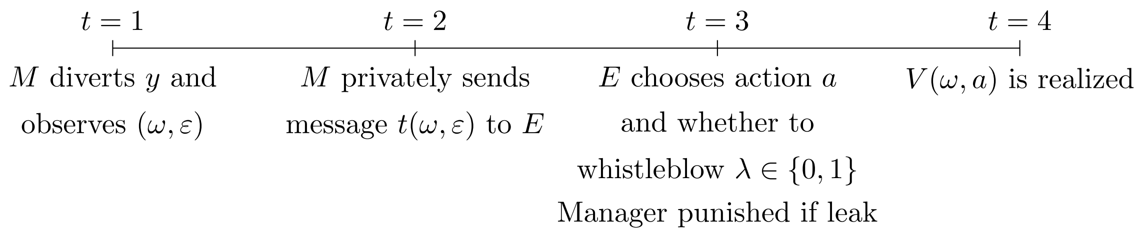
More generally, while we abstract from the impact of misconduct realization on the likelihood of whistleblowing in our main analysis, we uncover a novel channel through which stronger whistleblower incentives can have opposite effects on measures of internal and public informativeness, and real efficiency. As such, our paper relates to the larger literature that distinguishes between forecasting price efficiency and revelatory price efficiency (e.g., Bond, Edmans, and Goldstein (2012)). Specifically, our paper highlights a novel mechanism through which a higher incidence of leaks can lead to lower firm value, even while making prices (and public information) more informative about fundamentals on average.

Our paper also relates to the literature on communication with multiple receivers that compares the effectiveness of private versus public communication.[10] Much of this literature focuses on settings in which the sender can explicitly choose whether to send a private or public message to a given receiver. In contrast, in our setting, the sender (manager) accounts for the possibility that his private communication with the employee might become public (to the regulator / market), and strategically distorts his communication accordingly.

of whistleblower rewards, penalties for false reports, the probabilities of false positives and negatives in the judicial process, and the standard of proof affect the effectiveness of a whistleblower program.

[10]See Farrell and Gibbons (1989); Levy and Razin (2004); Johns (2007); Koessler (2008); Goltsman and Pavlov (2011) for general communication games. Newman and Sansing (1993) and Gigler (1994) consider the problem that a firm's disclosure of information about product demand may be simultaneously observed by the capital market, shareholders and competitors. Gertner, Gibbons, and Scharfstein (1988) examine the choice of financial structure when the financing contract is observed by both the capital market and a competing firm in a signaling model. Spiegel and Spulber (1997) consider the audiences as the capital market and a regulator.

Figure 1: Timeline of events



3 Model

We consider a setting in which the manager (he) chooses how much of the firm's cash to divert for private benefit, and this misconduct reduces the productivity of the firm on average. An employee (she) chooses the optimal action to exert based on her perception of the firm's productivity. The manager can influence this perception by engaging in costly manipulation of the employee's information. Finally, the employee chooses whether or not to blow the whistle on the manager's misconduct by leaking this internal information at a private cost.

Specifically, there are four dates (i.e., $t \in \{1, 2, 3, 4\}$) and three participants: (i) the firm manager M (ii) the firm employee E, and (iii) a regulator. The firm initially begins with \$1 of internal cash. The terminal (date four) value of the firm $V(\omega, a)$ depends on the manager's action through its impact on productivity, ω, and the employee's action, a, as described below.

Timing. The timing of events is summarized in Figure 1.

At date one, the manager diverts a fraction y of the internal cash, which generates private benefits of $y - (1/2)y^2$. The diversion generates a signal $\omega \sim N(y, \sigma_\omega^2)$, which also affects the terminal cash flows of the firm, as described below.

At date two, the manager can engage in a costly "cover-up" of the warnings signal ω. Specifically, by incurring a cost $\frac{\tau}{2}\frac{(t-\omega)^2}{\varepsilon}$, the manager can conceal ω and instead send a signal t to the employee. The manager privately observes the cost parameter $\varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon^2)$, which is independent of ω, but this is unobserved by the employee and the regulator.[11]

At date three, the employee takes an action a to maximize the value of the firm $V(a, \omega)$. Moreover, she can incur a cost $c \geq 0$ to leak the private communication t to the regulator, where $c \sim F(\cdot)$ is the private cost of whistleblowing she incurs. We assume that the CDF $F(c)$ is sufficiently well-behaved, i.e., it is continuous and differentiable for all c. This personal cost reflects both direct costs of whistleblowing (e.g., collecting and providing relevant information to the regulator, and directing them appropriately) and indirect costs (e.g., a loss of reputation as being loyal to the firm, costs due to retaliation by the manager or the firm). If she does blow the whistle, she receives a whistleblower bounty or (expected) reward R from the regulator. Let us denote her decision to leak by $\lambda \in \{0, 1\}$, where $\lambda = 0$ implies she did not blow the whistle, and $\lambda = 1$ denotes that she did.

Conditional on a leak, the manager faces a penalty Π that depends on the loss in productivity ω that results from his stealing. For tractability, we assume $\Pi = \pi \times \mathbb{E}[\omega|t]$, so that the total penalty is proportional to the expected impact of the misconduct, conditional on the leak, and the intensity π of the penalty reflects the severity of the regulator's whistleblower policy and the likelihood of success of prosecution.

At date four, the firm's terminal cash flows, given by

$$V(a, \omega) = (1 - \omega) - \beta(a - (1 - \omega))^2, \tag{1}$$

are realized, where ω reflects the reduction in firm productivity as a result of managerial misconduct. A key parameter for our analysis is β, which reflects the relative importance of internal alignment for firm value. As we discuss in more detail below, this parameter determines the impact of internal communication on real efficiency.

Payoffs. The employee chooses $\{a, \lambda\}$ to maximize her payoff:

$$u_E = \mathbb{E}[V(a, \omega)|t] + \lambda(R - c), \tag{2}$$

[11] To address the case when $\varepsilon = 0$, we adopt the convention that $0/0 = 0$.

given the signal t available to her. The manager chooses $\{y, t\}$ to maximize:

$$u_M = y - (1/2)y^2 - \frac{\tau}{2}\frac{(t-\omega)^2}{\varepsilon} - \lambda \times \pi \times \mathbb{E}[\omega|t]. \tag{3}$$

We make the following assumption to restrict the size of the penalty, so that the manager is not liable for more than the implied loss to the company due to stealing.

Assumption 1. *Assume that* $0 \leq \pi \leq 1$*.*

Equilibrium. We focus on pure strategy, subgame perfect equilibrium. In particular, an equilibrium is characterized by: (i) an optimal choice of action a and whistleblowing λ by the employee which maximize (conditional) expectation of u_E in (2); (ii) an optimal choice of diversion y and internal message $t(\omega, \varepsilon)$ by the manager which maximize (conditional) expectation of u_M in (3); and, (iii) participants' beliefs that satisfy Bayes' rule wherever it is well-defined.

3.1 Discussion of Assumptions

The manager in our setting engages in misconduct (by diverting cash), which reduces firm productivity. Moreover, we assume that this behavior generates an internal signal, or warning ω that is observable to the employee. However, the manager can engage in a "cover-up" by manipulating this internal communication. Our specification provides a parsimonious model which maps closely to notion of managerial misconduct in practice, but abstracts from any value enhancing roles for the manager. Equivalently, one could instead assume that the manager exerts costly effort $1 - y$ to increase the fundamentals of the firm's cash flows, and ω is a signal about how much "shirking" he engages in. Importantly, our setting is different from traditional settings in which the sender provides a signal about exogenous fundamentals. Instead, the manager manipulates information about *endogenous* firm productivity, the distribution of which he exerts an influence on.[12] Importantly, the impact

[12]Misinformation is false or inaccurate information, while disinformation is false information which is intended to mislead. As such, one could interpret our model as one in which the manager feeds disinformation

of the cash flow diversion on firm value is through ω which depends stochastically on the diversion y — this ensures that there is a role for learning by the employee and regulator and a corresponding role for manipulation by the manager.[13]

Firm value. The specification of firm value in (1) reflects a key feature of our analysis — namely, that firm value depends, in part, on the informativeness of internal communication to the employee. We capture this in a tractable and parsimonious way by assuming that firm value is increasing in the alignment of the employee's action with interim cash flows. The parameter β captures the relative importance of this effect on firm value. When β is low, firm value is primarily driven by interim cash flows, and consequently, the amount of cash flow diversion by the manager. However, when β is sufficiently high, the alignment between the employee's action and interim cash flows plays an important role, and so does the precision of internal communication.[14] Existing work, including Bolton, Brunnermeier, and Veldkamp (2013) and Dessein and Santos (2021), has explored the role of such alignment components of firm value in determining corporate culture and organizational structure. However, we expect other specifications of firm value that capture the key feature to yield similar implications.

Managerial payoffs. The specifications for the manager and employee payoffs are primarily for tractability. In our main analysis, we highlight that while the regulator penalty serves to discipline the manager's misconduct, it can have unintended consequences on firm value. We assume that the penalty depends on the regulator's posterior beliefs about misconduct, given the manipulated signal t (and not directly on the realized misconduct ω or cash flow diversion y). This captures the feature that, in practice, regulators often impose penalties based on the assessed damages based on the evidence presented, and ensures that the manager has an

to the employee, so as to benefit from the misinformation that the employee reports to the regulator. We are grateful to Lin Nan for making this connection.

[13] Specifically, if firm value depended directly on y, then this would be perfectly inferred by the regulator and the employee in equilibrium, and so the manager would have no incentive to manipulate his internal communication. Similarly, if we had instead extended the model of Dye and Sridhar (2004) to capture manipulation by the manager (i.e., if the cost of the message t was proportional to $(t - \omega - \varepsilon)^2$), the informativeness of the internal communication would not depend on the likelihood of whistleblowing or the accompanying penalties.

[14] We discuss strategies for identifying empirical proxies of β in Section 7.

incentive to manipulate his internal communication with the employee.

In Section 6.1, we characterize how managerial behavior and firm value change if, instead, the market can serve a similar disciplinary role in a setting where the manager is compensated by the firm's stock price. This analysis allows us to capture the notion that most executives receive stock based compensation in practice and implies that the manager has an incentive to distort his communication to the employee and the market.[15]

Employee payoffs. We assume that the employee receives a constant reward from blowing the whistle, but show in Section 6.2 that the main results from our analysis are qualitatively similar if, instead, we assume that the whistleblower bounty is proportional to the expected misconduct. Similarly, the assumption that the employee's payoff is sensitive to the terminal value ensures that she has an incentive to align her action to fundamentals, that is, choose a as close to $1-\omega$ as possible. Making her payoff also depend on the firm's stock price would not qualitatively change our results, but would reduce tractability. We abstract from other motivations for whistleblowing (e.g., ethical or fairness considerations) for expositional clarity and tractability.[16] However, one could interpret the personal cost c as reflecting a *net* cost of whistleblowing after adjusting for such preferences. Moreover, in our benchmark analysis we assume that the employee can leak the internal message $t(\omega)$ perfectly, and the probability with which she does so increases with the whistleblower reward R and decreases in the private cost c. One could instead assume that the employee always leaks a noisy signal of t, where the precision of the leak increases in her whistleblower bounty — this would yield qualitatively similar results.[17]

Distributional assumptions. The assumption that ω and ε are normally distributed

[15]If instead, the manager's payoff only depended on the terminal cash-flow V, then the manager would have no incentive to distort his communication (i.e., $t=\omega$) and would not divert any internal cash (i.e., $y^*=0$). We expect our results to be qualitatively similar if, instead, the manager's payoff was driven by a weighted average of P and V, although the analysis would be more cumbersome.

[16]For instance, one could assume that the cost c is lower when perceived misconduct $\mathbb{E}[\omega|t]$ is higher if the employee has an ethical motive for whistleblowing, but this would make the analysis less tractable.

[17]In fact, the benchmark model in the previous version of this paper considered this case. While similarly tractable, this alternate version implies that there is always a leak, which is inconsistent with the empirical observation that whistleblowing is observed relatively infrequently.

implies that the conditional distribution of ω, given the signals, is also normal, which ensures that characterizing agents' updating of beliefs is tractable. As in Fischer and Verrecchia (2000), this implies that the cost of manipulating signals can sometimes be negative (since ε can take on negative values). However, by setting μ_ε appropriately, the likelihood of this can be made arbitrarily small. One could alternatively ensure non-negativity by assuming that the shocks are elliptically distributed (as in Frankel and Kartik (2019) and Ball (2025)), but this makes the characterization of real efficiency in our analysis less tractable.[18]

4 Analysis

We solve the model by working backwards.

Date three. Given her objective, the employee's optimal choice of whistleblowing is given by

$$\lambda^* = \mathbf{1}_{R>c}, \tag{4}$$

where $\mathbf{1}_.$ is an indicator function, and her optimal choice of action a is given by:

$$a^* = \arg\max_a \mathbb{E}[(1-\omega) - \beta(a-(1-\omega))^2|t] = 1 - \mathbb{E}[\omega|t]. \tag{5}$$

We shall conjecture and verify that $\mathbb{E}[\omega|t] = b_0 + b_t t$, where the coefficients b_0 and b_t will be determined in equilibrium and characterized below.

Date two. The manager conditions on the realizations of ω and ε when choosing how to distort the internal signal — specifically, he chooses $t(\omega, \varepsilon)$ to maximize:

$$\mathbb{E}[u_M|\omega,\varepsilon] = y - (1/2)y^2 - \frac{\tau}{2}\frac{(t-\omega)^2}{\varepsilon} - F(R) \times \pi \times (b_0 + b_t t)$$

[18] Specifically, unless the shocks are normally distributed, the conditional variance of ω depends on the realization of the signals (e.g., see Foster and Viswanathan (1993)), which implies the expected firm value is no longer analytically tractable.

which implies

$$t^*(\omega, \varepsilon) = \omega - x\varepsilon, \quad \text{where} \quad x \equiv \frac{b_t \pi F(R)}{\tau} \tag{6}$$

This is intuitive: the manager's choice of x reflects the *endogenous* intensity of manipulation. Consider the natural case where $\varepsilon > 0$, so that the manager has an incentive to bias downwards the signal $t(\omega)$ of his misconduct. The above implies that the intensity of manipulation increases in (i) the intensity π of the penalty, (ii) the likelihood $F(R)$ of whistleblowing, and (iii) the sensitivity b_t of the conditional expectation of wrongdoing ω, given the leak t, but decreases in the manager's cost τ of manipulation.

Let $\hat{y}$ denote the employee and regulator's conjecture about the manager's misconduct. Then,

$$\mathbb{E}[\omega|t] = \frac{\frac{\hat{y}}{\sigma_\omega^2} + \frac{t + x\mu_\varepsilon}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}} \quad \text{and} \quad \mathbb{V}[\omega|t] = \frac{1}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}} \tag{7}$$

which verifies the conjecture $\mathbb{E}[\omega|t] = b_0 + b_t t$.

Date one. Given these coefficients, the manager's diversion choice y maximizes:

$$\mathbb{E}[u_M|y] = y - (1/2)y^2 - \frac{\tau}{2}\mathbb{E}\left[\left.\frac{(x\varepsilon)^2}{\varepsilon}\right| y\right] - F(R) \times \pi \times \left(b_0 + b_t \mathbb{E}\left[\omega - x\varepsilon \middle| y\right]\right)$$

which implies the manager's optimal diversion is given by:

$$y^* = 1 - \tau x = 1 - b_t \pi F(R). \tag{8}$$

Again, this is intuitive. The manager steals less (i.e., y^* is lower) when the intensity π of the penalty, the likelihood $F(R)$ of whistleblowing and the informativeness b_t of the leak t about misconduct are higher. The following proposition establishes that there is a unique equilibrium in our setting.

Proposition 1. *Suppose Assumption 1 holds. Then there exists a unique equilibrium in which optimal choices are characterized by* (4)*,* (5)*,* (6) *and* (8)*, and the employee's / regulator's conditional beliefs about* ω *given* t *are given by* $\mathbb{E}[\omega|t] = b_0 + b_t t$ *and* $\mathbb{V}[\omega|t] = \sigma_\omega^2(1 - b_t)$*, where*

$$b_0 = \frac{\frac{\hat{y}}{\sigma_\omega^2} + \frac{x\mu_\varepsilon}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}}, \quad b_t = \frac{\tau x}{\pi F(R)},$$

and the intensity of manipulation $x \in (0, 1/\tau)$ *is the unique solution to:*

$$x = \frac{\pi F(R)\sigma_\omega^2}{\tau\sigma_\omega^2 + \tau x^2 \sigma_\varepsilon^2}. \tag{9}$$

The proof establishes that there is a unique solution x to the fixed point characterized by (9), which pins down the equilibrium. This expression reflects the fact that the weight $b_t = \frac{\tau x}{\pi F(R)}$ that the employee and the regulator put on the internal signal t depends on the variance of the error in the equilibrium communication t^*, which depends on the intensity of manipulation by the manager x, which in turn, increases with the weight b_t.

5 Measures of Efficiency

In this section, we characterize how changes in whistleblower rewards R and the intensity of regulatory penalties π affect the various measures of efficiency. We consider three measures of efficiency: (i) the informativeness of internal communication, (ii) the informativeness of public information, and (iii) real efficiency.

We measure the informativeness of **internal communication**, $\mathcal{I}_{IC}$, as follows:

$$\mathcal{I}_{IC} \equiv \mathbb{E}[\sigma_\omega^2 - \mathbb{V}[\omega|t]].$$

This captures how informative the internal signal t is about the impact of misconduct, ω, on average.[19]

Similarly, the informativeness $\mathcal{I}_P$ of **public information** captures how informative public information is about ω. In our setting, the only source of public information is leaks by the employee.[20] Accordingly, we measure $\mathcal{I}_P$ as follows:

$$\mathcal{I}_P \equiv \mathbb{E}[\sigma_\omega^2 - \mathbb{V}[\omega|\lambda, \psi]],$$

where $\psi \in \{\varnothing, t\}$ denotes the public information conditional on whether there is a leak i.e.,

$$\psi \equiv \begin{cases} \varnothing & \text{if } \lambda = 0 \\ t & \text{if } \lambda = 1. \end{cases}$$

The following result characterizes how whistleblower incentives can have differential effects on the informativeness of internal communication and public information.

Proposition 2. *(1) The informativeness of internal communication $\mathcal{I}_{IC}$ always decreases in whistleblower penalties π and whistleblower rewards R.*
(2) The informativeness of public information $\mathcal{I}_P$ always decreases in whistleblower penalties π but increases in whistleblower rewards R.

The impact of whistleblower rewards and penalties on the informativeness of internal communication is intuitive. Note that an increase in whistleblower rewards leads to more leaks (higher $F(R)$) all else equal, while an increase in whistleblower penalties π penalizes the manager more. While these changes decrease the manager's incentives to divert cash flows (i.e., reduces y^*), they also increase the manager's incentives to distort the internal communication. As a result, this makes t a noisier signal about ω, and consequently, the

[19]An alternate measure of informativeness would be how informative the internal signal is about terminal cash flows i.e., $(\mathbb{V}[V] - \mathbb{V}[V|t])$. We prefer our measure of price informativeness because it is more analytically tractable and intuitive - this is because the terminal cash flows are a non-linear function of ω.

[20]In some instances, the whistleblower leaks information publicly (e.g., Frances Haugen and the Facebook files). In other settings, the leaked information becomes public when the regulator initiates litigation, even if in such cases the identity of the whistleblower may remain secret to preserve confidentiality.

informativeness of internal communication decreases.

In contrast, the impact of the two whistleblower incentives on the informativeness of public information is different. To see why, note that one can express $\mathcal{I}_P$ as:

$$\mathcal{I}_P = \mathbb{E}[\sigma_\omega^2 - \mathbb{V}[\omega|\lambda, \psi]] = \sigma_\omega^2 - \big((1 - F(R))\sigma_\omega^2 + F(R)\mathbb{V}[\omega|t]\big)$$

As discussed above, higher penalties and higher rewards both lead to the (leaked) signal t being less informative (i.e., increase $\mathbb{V}[\omega|t]$), which tend to decrease the informativeness of public information. Moreover, an increase in penalties does not affect the employee's decision to blow the whistle (i.e., leaves $F(R)$ unaffected) and so the overall effect of such a change is to reduce $\mathcal{I}_P$. However, an increase in whistleblower rewards increases the probability of a leak (i.e., increases $F(R)$), which in turn, tends to increase the informativeness of public information. In the proof of the above result, we show that the latter effect dominates and, consequently, the overall effect of higher whistleblower rewards is to increase the informativeness of public information.

Finally, the third measure of efficiency we consider is **real efficiency**, which reflects the unconditional expected value of the firm, i.e.,

$$RE \equiv \mathbb{E}[V] = \mathbb{E}[1 - \omega] - \beta\mathbb{E}[\mathbb{V}[1 - \omega|t]].$$

As the next result illustrates, the impact of whistleblower incentives on real efficiency depends crucially on the relative importance of coordination within the firm.

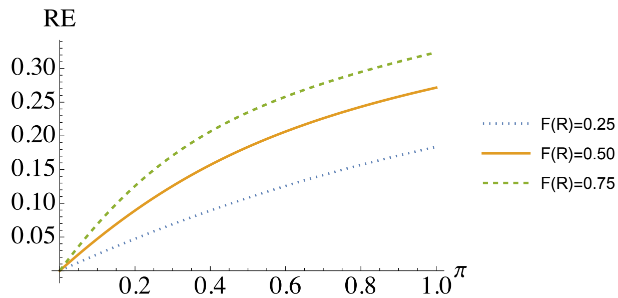
Proposition 3. *There exists $0 < \underline{\beta} < \bar{\beta}$, such that:*

(i) when $\beta \leq \underline{\beta}$, real efficiency RE increases in both R and π, and

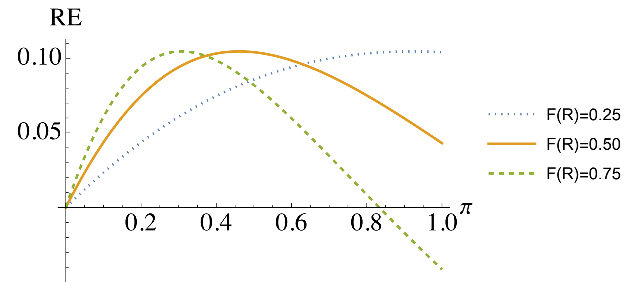
(ii) when $\beta > \overline{\beta}$, real efficiency RE is hump-shaped in both R and π.

To gain some intuition for the impact of whistleblower policies on real efficiency, note that RE consists of two components. The first component $\mathbb{E}[1 - \omega] = 1 - y^*$ decreases with

Figure 2: Real efficiency RE as a function of π and $F(R)$
Unless otherwise specified, other coefficients are set to $\sigma_\omega = \sigma_\varepsilon = \tau = 1$.



(a) RE vs. π when $\beta = 1$



(b) RE vs. π when $\beta = 2.5$

the magnitude of managerial misconduct y^* which, in turn, increases with an increase in either whistleblower rewards or penalties. As a result, this component increases in both R and π.

The second component $\beta\mathbb{E}[\mathbb{V}[1-\omega|t]]$, reflects the loss in firm value due to the misalignment between fundamentals and the employee's action. This loss is proportional to the average posterior uncertainty that the employee faces, after observing the internal communication t. As discussed above, an increase in either whistleblower rewards or penalties decreases informativeness, and consequently, leads to a larger loss in firm value due to misalignment.

When the relative impact of the loss from misalignment is sufficiently low (i.e., $\beta < \underline{\beta}$), the first component dominates and real efficiency always increases with both R and π. Panel (a) of Figure 2 provides an illustration. However, when β is sufficiently high, the latter effect can dominate when whistleblower incentives are sufficiently large. In this case, real efficiency can first increase but then eventually decrease in R and π, as illustrated in panel (b) of Figure 2.

To gain some intuition for this hump-shape, note that one can express real efficiency as

$$RE = \tau x - \frac{\beta}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}},$$

where $x = \frac{b_t \pi F(R)}{\tau}$ is increasing in π and R. Moreover, the above implies that the benefit from increasing π or R through higher productivity (captured by the $\mathbb{E}[1-\omega]$ term in RE) is linear in x. However, the loss from misalignment (captured by $\beta\mathbb{E}[\mathbb{V}[1-\omega|t]]$) is initially convex in x.[21]

This implies that when π or R, and consequently x, is very low, the increase in RE due to lower cash-flow diversion dominates the decrease in RE due to more misalignment. However, as π or R (and x) increase, the rate of increase in RE due to lower diversion is constant, but the rate of decrease due to misalignment increases. Beyond a certain level of π and/or R under a sufficiently high β, we show that decrease due to misalignment dominates and so RE decreases with further increases in whistleblower incentives.

This implies that when β is sufficiently high, there exists an interior optimum level of x^* which maximizes real efficiency. One can interpret this as the level of x^* that the regulator should target, if their goal is to maximize firm value. Note that since $x = \frac{b_t \pi F(R)}{\tau}$, this corresponds to a continuum of optimal policy choices $\{\pi^*, R^*\}$ for which the regulator is indifferent.

In practice, the reward for whistleblowing is tied to the penalties incurred by the manager. For instance, The Dodd-Frank Whistleblower Program rewards whistleblowers between 10 and 30 percent of the government's monetary sanctions collected from enforcement actions.[22] In line with this, suppose that whistleblower reward R is set as a fraction of π i.e, $R = \alpha\pi$ for some $\alpha \in (0, 1]$. Then, the above analysis implies that when β is sufficiently large, there is a unique pair of $\{\pi^*, R^*\}$ which maximizes firm value, characterized implicitly by $x^* = \frac{b_t \pi^* F(\alpha\pi^*)}{\tau}$. The next proposition characterizes how this optimal pair of policy choices varies with the other parameters.

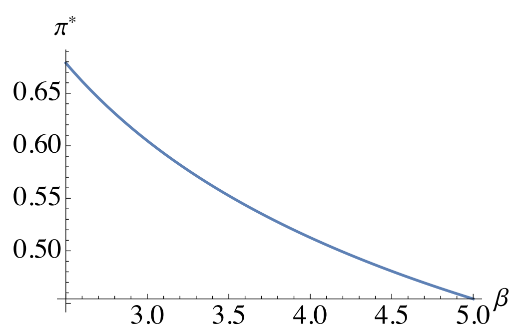
Proposition 4. *Suppose $\beta > \overline{\beta}$ and the whistleblower reward R is proportional to the penalty*

[21] As we show in the proof of Proposition 3, the penalty from misalignment can become concave in the right tail of x depending on the other parameters. However, even in that case, we always have a hump-shaped real efficiency *locally*.

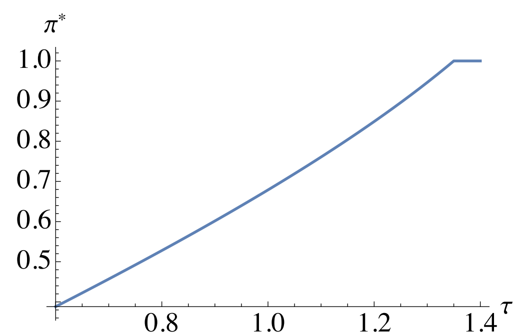
[22] Similar rewards are specified by the False Claims Act and the IRS whistleblower law - see https://www.whistleblowers.org/whistleblower-protections-and-rewards/.

π, i.e., $R = \alpha\pi$. Then, the optimal level of penalty π^*, when interior, is increasing in τ, decreasing in β, σ_ε, σ_ω and α.

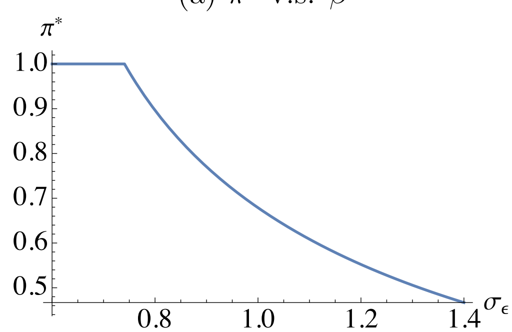
Figure 3: Optimal level of whistleblower penalty π^* as a function of model parameters. We plot the optimal level of whistleblower penalty π^* when the reward is proportional to the penalty i.e., $R = \alpha\pi$, and the private cost from whistleblowing is uniformly distributed between zero and one i.e., $F(c) = c$. Unless otherwise specified, other coefficients are set to $\sigma_\omega = \sigma_\varepsilon = \tau = 1$, $\beta = 2.5$, and $\alpha = 0.5$.



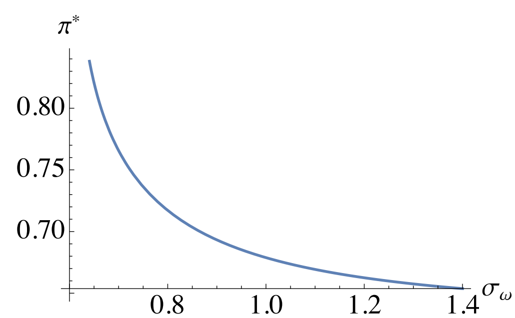
(a) π^* v.s. β



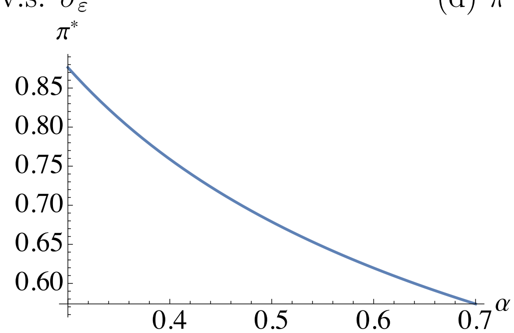
(b) π^* v.s. τ



(c) π^* v.s. σ_ε



(d) π^* v.s. σ_ω



(e) π^* v.s. α

Figure 3 provides an illustration of these results. In each of these cases, a change in the underlying parameter differentially affects the marginal benefit (i.e., $\frac{\partial \mathbb{E}[1-\omega]}{\partial \pi}$) and the marginal cost (i.e., $-\frac{\partial \beta \mathbb{E}[\mathbb{V}[1-\omega|t]]}{\partial \pi}$) of increasing π. For instance, an increase in the sensitivity to misalignment β increases the marginal cost of increasing π, but leaves the marginal benefit unchanged, and as a result, the optimal choice of π^* is lower. Similarly, an increase in τ makes it more costly for the manager to manipulate the internal signal, making the employee and the regulator better informed. While this has no direct impact on the marginal benefit, it indirectly improves the marginal benefit (through the impact on b_t) and reduces the marginal cost, and consequently, the optimal choice of π^* increases in τ.

An increase in σ_ε makes the internal communication noisier and so decreases the marginal benefit and increases the marginal cost, which leads to a decrease in the optimal choice of π^*. An increase in the volatility σ_ω of the internal warning has more subtle effects. Such an increase implies that the employee and the regulator put relatively more weight on the manipulated signal t (i.e., b_t is higher). On the one hand, this increases the marginal benefit from increasing π, since the disciplinary role of whistleblowing is more effective. However, a higher b_t also increases the manager's incentive to manipulate the internal communication, which increases the marginal cost of increasing π. When β is sufficiently large, the increase in the marginal cost due to misalignment is higher than the increase in the marginal benefit, and so the optimal π^* decreases.

Finally, an increase in α implies that a given level of penalty π corresponds to larger whistleblower rewards. Since π and R have complementary effects on the manager's optimal behavior (i.e., they always affect the manager's action through $\pi F(R)$), this implies a higher α requires a smaller penalty π to induce the optimal action (i.e., induce x^*). As a result, the optimal whistleblower penalty decreases with α.

Importantly, the above results imply that one must be cautious about evaluating the impact of such policies on the cross-section of firms, since they can have qualitatively different effects across different firms. Specifically, whistleblower incentives should be weaker for firms in which internal alignment is more important (i.e., β is higher) and cash flows are more volatile (i.e., σ_ω is higher), but stronger when it is costlier for the manager to manipulate internal information (i.e., τ is higher).

Moreover, the above results highlight the importance of coordinating policy instruments. An increase in whistleblower incentives (e.g., by increasing the fraction α of the penalties that are awarded to the whistleblower) should be accompanied by a reduction in the intensity or severity of penalties imposed. Otherwise, these polices can lead to too much managerial manipulation of internal information, and consequently, lower firm value.

Given the potential adverse effects of penalizing the manager too much, one might wonder how the impact of whistleblower incentives changes if, instead, the manager were disciplined by market forces. In the next section, we explore this issue by considering a setting in which the manager's payoff depends on the firm's stock price instead of a regulatory penalty.

6 Robustness

In this section, we explore the robustness of our analysis to alternate specifications. In Section 6.1, we assume that the manager does not incur a penalty conditional on a leak; instead, he is compensated based on the firm's stock price. In Section 6.2, we relax the assumption that the whistleblower reward is independent of misconduct by assuming that it depends proportionally on the expected misconduct. In both cases, we show that the resulting equilibrium and the impact of whistleblower incentives remain qualitatively similar to our main analysis.

6.1 The Disciplinary Role of Prices

In our benchmark analysis, we assume that the regulator penalizes the manager for misconduct when there is a leak. In this section, we explore a setting in which, instead, the manager's behavior is disciplined by market prices.

Specifically, we assume that in addition to the manager, the employee, and the regulator, there is a representative, risk neutral investor who observes all the available public information at date 3, and then sets the firm's stock price P. Specifically, the firm's stock price reflects the investor's conditional expectation of firm's terminal value $V(\omega, a)$ given the employee's decision to blow the whistle (i.e., λ) and the signal t, conditional on a leak.

Moreover, we assume that the manager's compensation depends on the date-three price, but that he does not incur a penalty when there is a leak. In particular, we assume that the manager chooses $\{y, t\}$ to maximize:

$$u_M = y - (1/2)y^2 + \delta P - \frac{\tau}{2}\frac{(t-\omega)^2}{\varepsilon}. \tag{10}$$

We assume that $0 \leq \delta \leq 1$ so that the moral hazard problem faced by the manager remains relevant. Specifically, note that the above implies that investing one dollar in the firm yields him δ dollars in expectation. Consequently, one could increase δ arbitrarily high to ensure that the manager never engages in cash flow diversion. Finally, we assume that the employee's payoffs remain the same as in our benchmark model.

6.1.1 Analysis and Equilibrium

As before, we can solve the model by working backwards.

Date three. Note that since the employee's payoffs remain unchanged, her optimal actions are still characterized by (4) and (5).

Moreover, since the employee's decision of whether or not to whistle-blow is independent of the realization of the internal signal t, the market price at date 3 can be expressed as:

$$P = \begin{cases} \mathbb{E}[V(a^*, \omega)] \equiv P_0 & \text{if } \lambda = 0 \\ \mathbb{E}[V(a^*, \omega)|t] \equiv P_1(t) & \text{if } \lambda = 1 \end{cases}, \tag{11}$$

where P_0 is independent of t, but can depend on the conjecture about managerial misconduct $\hat{y}$, while $P_1(t)$ depends on the leaked internal signal t. We conjecture and verify that the price is linear in the signal, i.e., $P_1 = p_0 - p_t t$.[23]

Date two. Given the price above, the manager conditions on the realizations of ω and ε when choosing how to distort the internal message – specifically, he chooses $t(\omega, \varepsilon)$ to maximize:

$$\mathbb{E}[u_M|\omega, \varepsilon] = y - (1/2)y^2 + \delta\left((1 - F(R))P_0 + F(R)(p_0 - p_t t)\right) - \frac{\tau}{2}\frac{(t-\omega)^2}{\varepsilon}$$

which implies

$$t^*(\omega, \varepsilon) = \omega - \frac{\delta p_t F(R)}{\tau}\varepsilon. \tag{12}$$

This is analogous to the expression in our main analysis. Since t is a signal of managerial misconduct, we expect the price to decrease with t, i.e., $p_t > 0$. In this case, when $\varepsilon > 0$ (so that it is costly for the manager to manipulate t), the above implies that the manager has an incentive to bias the signal downwards.

Let $\hat{y}$ denote the conjecture about the manager's misconduct, and define $x \equiv \frac{\delta p_t F(R)}{\tau}$. Then, we can express $t(\omega, \varepsilon) = \omega - x\varepsilon$, which implies

$$\mathbb{E}[\omega|t] = \frac{\frac{\hat{y}}{\sigma_\omega^2} + \frac{t + x\mu_\varepsilon}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}} \quad \text{and} \quad \mathbb{V}[\omega|t] = \frac{1}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}}. \tag{13}$$

Next, note that conditional on the leaked signal t, the price is given by:

$$\begin{aligned} P_1(t) &= \mathbb{E}\left[V(\omega, a^*)|t\right] \\ &= \mathbb{E}\left[1 - \omega|\, t\right] - \beta\mathbb{E}\left[\mathbb{E}\left[(a^* - (1-\omega))^2\big|\, t\right]\big|\, t\right] \\ &= \mathbb{E}\left[1 - \omega|\, t\right] - \beta\mathbb{E}\left[\mathbb{V}\left[(1-\omega)|\, t\right]\big|\, t\right], \end{aligned}$$

[23] As we shall see below, the assumption that ω and ε are normally distributed ensures that the misalignment component of firm value does not depend on the realization of ω and ε in expectation, and so the price is linear in t when there is a leak.

where the final equality follows from the observation that $a^* = \mathbb{E}[1-\omega|t]$. But since $\mathbb{E}[\omega|t]$ is linear in t and $\mathbb{V}[1-\omega|t] = \mathbb{V}[\omega|t]$ is independent of t, the above verifies the conjectured form: $P_1(t) = p_0 - p_t t$, where $p_t > 0$.

Date one. Given these coefficients, the manager's diversion choice y maximizes:

$$\mathbb{E}[u_M|y] = y - (1/2)y^2 + \delta\left((1-F(R))P_0 + F(R)(p_0 - p_t\mathbb{E}[t|y])\right) - \frac{\tau}{2}\mathbb{E}\left[\left.\frac{(x\varepsilon)^2}{\varepsilon}\right| y\right]$$

which implies the manager's optimal diversion is given by:

$$y^* = 1 - \tau x = 1 - \delta p_t F(R). \tag{14}$$

Again, this is intuitive. The manager steals less (i.e., y^* is lower) when his compensation is more sensitive to the price (i.e., δ is higher), the likelihood $F(R)$ of whistleblowing is higher and the price is more sensitive to the leaked signal (i.e., p_t is higher). The following proposition establishes that there is a unique equilibrium in our setting.

Proposition 5. *Suppose $0 \leq \delta \leq 1$. Then there exists a unique equilibrium in which optimal choices are characterized by (4), (5), (12) and (14), the employee's conditional beliefs about ω given t are given by (13), and the price is given by (11) as follows,*

$$P_0 = \tau x - \beta\sigma_\omega^2(1-p_t), \qquad P_1(t) = p_0 - p_t t,$$

where

$$p_0 = 1 - \frac{\frac{\hat{y}}{\sigma_\omega^2} + \frac{x\mu_\varepsilon}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}} - \beta\sigma_\omega^2(1-p_t) \quad \text{and} \quad p_t = \frac{\tau x}{\delta F(R)},$$

and where $x \in (0, 1/\tau)$ is the unique solution to

$$x = \frac{\delta F(R)\sigma_\omega^2}{\tau\sigma_\omega^2 + \tau x^2 \sigma_\varepsilon^2}. \tag{15}$$

The proof establishes that there is a unique solution x to the fixed point characterized by (15) which pins down the equilibrium. This expression reflects the fact that the weight $p_t = \frac{\tau x}{\delta F(R)}$ that the market puts on the leaked signal t depends on the variance of the error in the equilibrium communication t^*, which depends on the extent to which the manager distorts his internal communication (as given by equation (12)), which in turn, increases with the weight p_t.

It is worth noting that the fixed point in (15), which characterizes the equilibrium in this case, is analogous to the fixed point in (9), which characterized the equilibrium in our benchmark setting, The key difference is that the intensity of the whistleblower penalty, π, in the benchmark analysis is replaced by the manager's payoff sensitivity to price, δ. This, in turn, implies that one can establish analogous results regarding how whistleblower rewards and the price sensitivity of the manager's payoffs affect various measures of efficiency.

Specifically, let $\mathcal{I}_{IC} = \mathbb{E}[\sigma_\omega^2 - \mathbb{V}[\omega|t]]$ denote the informativeness of internal communication, $\mathcal{I}_P = \mathbb{E}[\sigma_\omega^2 - \mathbb{V}[\omega|\lambda, P]]$ denote the informativeness of public information and $RE = \mathbb{E}[V]$ denote real efficiency as before. The following result establishes that the analysis in Section 5 extends to the current setting.

Proposition 6. *(1) The informativeness of internal communication $\mathcal{I}_{IC}$ always decreases in the manager's price sensitivity δ and whistleblower rewards R.*
(2) The informativeness of public information $\mathcal{I}_P$ always decreases in the manager's price sensitivity δ but increases in whistleblower rewards R.
(3) There exists $0 < \underline{\beta} < \bar{\beta}$, such that:

(i) when $\beta \leq \underline{\beta}$, real efficiency RE increases in both R and δ, and

(ii) when $\beta > \overline{\beta}$, real efficiency RE is hump-shaped in both R and δ.

The proof is analogous to Propositions 2 and 3, where we replace the penalty π with the payoff sensitivity δ and the sensitivity b_t by p_t. Intuitively, while the manager's misconduct is disciplined by exposure to the market price, this also induces him to manipulate the internal information. As a result, a higher sensitivity δ to price leads to worse internal communication and worse public information (since, conditional on whistleblowing, the leaked information is less informative). Moreover, as in our main analysis, when β is sufficiently high, real efficiency is hump-shaped in both the whistleblower reward R and the price sensitivity δ.

In addition to establishing the robustness of our main results, this analysis provides additional testable implications of our analysis. Specifically, the adverse impact of whistleblowing on public information and internal communication is likely to be more severe for firms in which managers have higher-powered stock-based incentives (e.g., more stock / options based compensation). Moreover, the above analysis also suggests that in response to an increase in whistleblower rewards, firms should reduce the sensitivity of managerial stock-based compensation in order to mitigate the negative impact on information quality and firm value.[24]

6.2 Alternate Whistleblower Rewards

In our main analysis, we assume that the employee's reward from whistleblowing, and consequently, the likelihood of a leak, does not depend on the magnitude of misconduct. This assumption is made for tractability. In this section, we consider a setting in which the whistleblower reward is increasing in the expected value of the misconduct.

Specifically, suppose conditional on leaking the internal signal t, the employee receives a reward $R = \rho\mathbb{E}[\omega|t]$. To ensure that the analysis remains tractable, we make two additional assumptions. First, we assume that the employee's private cost of whistleblowing is distributed uniformly between 0 and 1, i.e., $c \sim U[0, 1]$.

Second, suppose that the penalty incurred by the manager, conditional on a leak, is constant, i.e., $\Pi = \pi_0$. As the analysis below illustrates, maintaining the assumption that the penalty incurred by the manager is also proportional to the expected misconduct (as in our main analysis) would imply that the manager's signal $t(\omega, \varepsilon)$ is no longer a linear

[24] As noted in Proposition 4, without additional conditions, real efficiency only dictates an optimal level of the product $\delta F(R)$, not δ and R individually.

function of ω. In turn, this violates the conjecture that the conditional expectation of ω given t (i.e., $\mathbb{E}[\omega|t]$) is linear in t, which renders the analysis intractable.

The remaining assumptions are as in the main model described in Section 3.

Given the above, the employee's optimal action is still given by (5). Moreover, given the whistleblowing reward, she only blows the whistle if

$$\lambda^* = \mathbf{1}_{\rho\mathbb{E}[\omega|t]>c}. \tag{16}$$

This implies that, from the manager's perspective, the probability of a leak is given by:

$$\Pr(\lambda = 1) = F(\rho\mathbb{E}[\omega|t]) = \rho\mathbb{E}[\omega|t].$$

As before, we conjecture and then verify that $\mathbb{E}[\omega|t] = b_0 + b_t t$. Given this, the manager chooses the internal signal, $t(\omega, \varepsilon)$, to maximize:

$$\mathbb{E}[u_M|\omega, \varepsilon] = y - (1/2)y^2 - \frac{\tau}{2}\frac{(t-\omega)^2}{\varepsilon} - \pi_0\rho(b_0 + b_t t)$$

which implies

$$t^*(\omega, \varepsilon) = \omega - \frac{\rho\pi_0 b_t}{\tau}\varepsilon. \tag{17}$$

Comparing the above expression with (6), it is immediate that setting $x = \frac{\rho\pi_0 b_t}{\tau}$ and proceeding the analysis as in Section 4 yields an analogous equilibrium. In particular, we can verify that beliefs about ω conditional on t are given by (7), and the manager's optimal level of diversion is given by

$$y^* = 1 - \tau x = 1 - \rho\pi_0 b_t. \tag{18}$$

The following proposition summarizes the characterization of the equilibrium in this setting.

Proposition 7. *Suppose $\pi_0 < 1$ and $\rho < 1$. Then there exists a unique equilibrium in which optimal choices are characterized by (5), (16), (17) and (18), the employee's conditional beliefs about ω given t are given by (7), where*

$$b_0 = \frac{\frac{\hat{y}}{\sigma_\omega^2} + \frac{x\mu_\varepsilon}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}}, \quad b_t = \frac{\tau x}{\rho \pi_0},$$

where $x \in (0, 1/\tau)$ is the unique solution to:

$$x = \frac{\pi_0 \rho \sigma_\omega^2}{\tau \sigma_\omega^2 + \tau x^2 \sigma_\varepsilon^2}.$$

We omit the proof of the above result, since it follows immediately from the proof of Proposition 1 by replacing the expression for x.

The above characterization suggests that the results of our main analysis remain qualitatively similar in this alternative specification. The key difference is that the relevant results now pertain to the sensitivity of the employee's reward to misconduct, as parametrized by ρ. Specifically, one can show that an increase in the sensitivity of whistleblower rewards, ρ, leads to a decrease in misconduct and the informativeness of internal communication, an increase in the informativeness of public information, and, when β is sufficiently high, a potential decline in real efficiency.

7 Discussion and Conclusion

We develop a stylized model to study how changes in whistleblower policies affect a manager's incentives for misconduct and manipulation of internal information, and consequently, how they affect the quality of internal and public information and firm value. We show that while larger whistleblower rewards and harsher penalties lead to less managerial misconduct, they can lead to a worsening of internal communication and lower firm value, especially when the costs from internal misalignment are sufficiently high.

The growing empirical literature on the impact of whistleblowing documents evidence consistent with our predictions. For example, Berger and Lee (2022) document that the introduction of the Dodd-Frank Whistleblower Program reduced the likelihood of accounting fraud by 12-22%, consistent with our prediction that such policies reduce the extent of misconduct. Similarly, Dey, Heese, and Pérez-Cavazos (2021) and Heese and Pérez-Cavazos (2021) show that larger whistleblower rewards and reduction in retaliation costs, respectively, lead to more whistleblowing by employees — these changes correspond to an increase in $F(R)$ in our main analysis.

Novel Predictions. However, our model makes a number of novel predictions that have not been directly tested in the empirical literature to our knowledge. First, all else equal, stronger whistleblower incentives should be associated with less informative internal communication. There is anecdotal evidence of such behavior at firms in response to whistleblowers (e.g., see Heath (2021) and Mak (2021) for recent examples from technology firms like Apple, Google, and Meta). However, empirical tests of this prediction are confounded by the fact that following (the incidence of) a leak, firms usually make changes to the internal governance and communication policies. As such, a test of this prediction requires identifying firms that are ex-ante more likely to have leaks, and comparing their internal communication to a control group. Bowen, Call, and Rajgopal (2010) provide some preliminary evidence consistent with this prediction: in their sample, targets of employee whistleblowing allegations are more likely to have unclear internal communication channels.

Second, different aspects of whistleblower policy can have different implications for internal and public information quality. Specifically, an increase in whistleblower rewards should lead to an increase in public information quality, but a decrease in internal information quality, while an increase in penalties should lead to a decrease in both. However, a key challenge in testing such a prediction is to separate the impact of these two dimensions empirically, since these changes often coincide with each other.

Third, and perhaps most importantly, our model highlights a novel channel through which whistleblower incentives affect real efficiency. While the existing empirical literature has focused on the efficacy of such policies on deterring misconduct, there is relatively less work on its broader impact on firm value. Our model predicts that the impact of whistleblower incentives is qualitatively different across firms: they unambiguously improve real efficiency for firms where internal alignment is not very important, but can reduce firm value when misalignment is very costly.

Proxies for β. Testing this central prediction of our model requires identifying a proxy for β, which captures the relative importance of internal employee alignment for firm value. While we are unaware of specific empirical proxies, we conjecture that, all else equal, β should be higher for firms in which (i) human capital and intellectual property are relatively large drivers of firm value (e.g., technology firms, pharmaceuticals, and healthcare), (ii) manufacturing processes or supply chains are more complex (e.g., automotive and aerospace industries), or (iii) employee collaboration is critical (e.g., financial, consulting, and professional services). On the other hand, we expect β should be lower for firms in which (i) physical capital and external factors are more important drivers of value (e.g., oil and commodity industries, small scale farms) and (ii) operations are decentralized (e.g., restaurant or retail franchises). Our analysis further suggests that the negative impact of whistleblower incentives for high β firms is likely to be exacerbated when the manager's compensation is more sensitive to the (short-term) stock price.

Policy implications. Our analysis also has implications for regulatory policy. When the impact of misalignment is sufficiently high, our analysis implies that there is an optimal level of whistleblower incentives which maximizes expected firm value. We show that the optimal level of rewards / penalties should be higher for firms in which the manager's ability to manipulate internal information is more limited (e.g., when internal control or auditing systems are more stringent). However, the optimal level of whistleblower incentives should be lower for firms in which the relative importance of employee alignment is high and for firms or industries where cash flows (or firm productivity) are more volatile. Our analysis also suggests that the optimal level of employee rewards and whistleblower penalties are substitutes: in order to mitigate the negative impact on firm value, an increase in whistleblower rewards should be accompanied by a decrease in the intensity of penalties (and vice versa).

Future work. Our model is stylized for expositional clarity and analytical tractability. However, our analysis suggests a number of directions for future work. For instance, it would be interesting to consider how whistleblower policies interact with other regulatory policies (e.g., mandatory disclosure requirements) to affect managerial behavior and firm value. It would also be interesting to consider settings in which the employee can manipulate the content of the leaked information, and characterize how larger rewards can lead to "false claims" and affect the effectiveness of whistleblower policies. We hope to explore these directions in future work.

References

Aubert, C., P. Rey, and W. E. Kovacic (2006). The impact of leniency and whistle-blowing programs on cartels. *International Journal of industrial organization 24*(6), 1241–1266. 7

Ball, I. (2025). Scoring strategic agents. *American Economic Journal: Microeconomics 17*(1), 97–129. 2, 3.1

Berger, P. G. and H. Lee (2022). Did the Dodd–Frank whistleblower provision deter accounting fraud? *Journal of Accounting Research 60*(4), 1337–1378. 7

Bolton, P., M. K. Brunnermeier, and L. Veldkamp (2013). Leadership, coordination, and corporate culture. *Review of Economic Studies 80*(2), 512–537. 3.1

Bond, P., A. Edmans, and I. Goldstein (2012). The real effects of financial markets. *Annu. Rev. Financ. Econ. 4*(1), 339–360. 2

Bowen, R. M., A. C. Call, and S. Rajgopal (2010). Whistle-blowing: Target firm characteristics and economic consequences. *The Accounting Review 85*(4), 1239–1271. 7

Buccirossi, P., G. Immordino, and G. Spagnolo (2021). Whistleblower rewards, false reports, and corporate fraud. *European Journal of Law and Economics 51*(3), 411–431. 9

Dessein, W. and T. Santos (2021). Managerial style and attention. *American Economic Journal: Microeconomics 13*(3), 372–403. 3.1

Dey, A., J. Heese, and G. Pérez-Cavazos (2021). Cash-for-information whistleblower programs: Effects on whistleblowing and consequences for whistleblowers. *Journal of accounting research 59*(5), 1689–1740. 7

Dye, R. A. (1988). Earnings management in an overlapping generations model. *Journal of Accounting research*, 195–235. 2

Dye, R. A. and S. S. Sridhar (2004). Reliability-relevance trade-offs and the efficiency of aggregation. *Journal of Accounting Research 42*(1), 51–88. 13

Farrell, J. and R. Gibbons (1989). Cheap talk with two audiences. *The American Economic Review 79*(5), 1214–1223. 10

Fischer, P. E. and R. E. Verrecchia (2000). Reporting bias. *The Accounting Review 75*(2), 229–245. 2, 3.1

Foster, F. D. and S. Viswanathan (1993). The effect of public information and competition on trading volume and price volatility. *The Review of Financial Studies 6*(1), 23–56. 18

Frankel, A. and N. Kartik (2019). Muddled information. *Journal of Political Economy 127*(4), 1739–1776. 2, 3.1

Friebel, G. and S. Guriev (2012). Whistle-blowing and incentives in firms. *Journal of Economics & Management Strategy 21*(4), 1007–1027. 2

Gertner, R., R. Gibbons, and D. Scharfstein (1988). Simultaneous signalling to the capital and product markets. *The RAND Journal of Economics*, 173–190. 10

Gigler, F. (1994). Self-enforcing voluntary disclosures. *Journal of Accounting Research 32*(2), 224–240. 10

Givati, Y. (2016). A theory of whistleblower rewards. *The Journal of Legal Studies 45*(1), 43–72. 9

Goltsman, M. and G. Pavlov (2011). How to talk to multiple audiences. *Games and Economic Behavior 72*(1), 100–122. 10

Heath, A. (2021). Meta goes into lockdown. *The Verge*. 2, 7

Heese, J. and G. Pérez-Cavazos (2021). The effect of retaliation costs on employee whistleblowing. *Journal of Accounting and Economics 71*(2-3), 101385. 7

Johns, L. (2007). A servant of two masters: communication and the selection of international bureaucrats. *International Organization 61*(2), 245–275. 10

Koessler, F. (2008). Lobbying with two audiences: Public vs private certification. *Mathematical Social Sciences 55*(3), 305–314. 10

Levy, G. and R. Razin (2004). It takes two: an explanation for the democratic peace. *Journal of the European Economic Association 2*(1), 1–29. 10

Mak, A. (2021). How leakers get dealt with in Silicon Valley. *Slate*. 7

McDade, A. (2023). Tesla instructed employees to only communicate verbally about complaints so there was no written record, leaked documents show. *Business Insider*. 2

Motta, M. and M. Polo (2003). Leniency programs and cartel prosecution. *International journal of industrial organization 21*(3), 347–379. 7

Nan, L., C. Tang, and M. Ye (2025). Auditing with a chance of whistleblowing. *Management Science*. 2

Nan, L., C. Tang, and G. Zhang (2024). Whistleblowing bounties and informational effects. *Journal of Accounting and Economics 77*, 101616. 2

Nan, L. and R. Zheng (2023). Whistleblowing and internal communication. *Working Paper*. 2

Newman, P. and R. Sansing (1993). Disclosure policies with multiple users. *Journal of Accounting Research*, 92–112. 10

Spagnolo, G. (2004). Divide et impera: Optimal leniency programs. *Available at SSRN 716143*. 7

Spiegel, Y. and D. F. Spulber (1997). Capital structure with countervailing incentives. *The RAND Journal of Economics*, 1–24. 10

Taylor-Neu, K., A. S. Rahaman, G. D. Saxton, and D. Neu (2024). Tone at the top, corporate irresponsibility and the Enron emails. *Accounting, Auditing & Accountability Journal 37*(9), 336–364. 2

A Proofs

A.1 Proof of Proposition 1

The optimal choices of whistleblowing and action of the employee follow from optimizing her payoff function (2). The manager's optimal distortion and diversion are given by the first order condition of his payoff:

$$\begin{aligned}
&-\frac{\tau}{\varepsilon}(t-\omega) - F(R)\pi b_t = 0 \\
&1 - y - F(R)\pi b_t = 0.
\end{aligned}$$

By matching the coefficients of $\mathbb{E}[\omega|t] = b_0 + b_t t$, we have

$$b_t = \frac{\frac{1}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}} \quad \text{and} \quad b_0 = \frac{\frac{\hat{y}}{\sigma_\omega^2} + \frac{x\mu_\varepsilon}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}}.$$

It follows from the definition of x that

$$b_t = \frac{\tau x}{\pi F(R)} = \frac{\frac{1}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}}.$$

So we have

$$x = \left(\frac{\pi F(R)}{\tau}\right)\left(\frac{\sigma_\omega^2}{\sigma_\omega^2 + x^2\sigma_\varepsilon^2}\right).$$

We next show that there is a unique solution x to $G(x) = 0$, where

$$G(x) \equiv \frac{\pi F(R)\sigma_\omega^2}{\tau\sigma_\omega^2 + \tau x^2\sigma_\varepsilon^2} - x.$$

Note that

$$\begin{aligned}
G(0) &= \frac{F(R)\pi}{\tau} > 0, \\
G(1/\tau) &= \frac{\pi F(R)\tau\sigma_\omega^2}{\tau^2\sigma_\omega^2 + \sigma_\varepsilon^2} - \frac{1}{\tau} < \frac{\pi F(R)}{\tau} - \frac{1}{\tau} \leq \frac{1}{\tau} - \frac{1}{\tau} = 0 \\
G'(x) &= -\frac{2\pi F(R)x\sigma_\omega^2\sigma_\varepsilon^2}{\tau\left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)^2} - 1 < 0,
\end{aligned}$$

where the inequality in the equation of $G(1/\tau)$ follows from the fact that $\pi F(R) \leq 1$ by Assumption 1. This implies there exists a unique solution $x \in (0, 1/\tau)$ to $G(x) = 0$. We then have $y^* = 1 - \tau x \in (0,1)$. Moreover, because t is normally distributed,

$$\mathbb{V}[\omega|t] = \frac{1}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}} = \frac{1}{\left(1 + \frac{b_t}{1-b_t}\right)\frac{1}{\sigma_\omega^2}} = \sigma_\omega^2\left(1 - b_t\right).$$

$\square$

A.2 Proof of Proposition 2

Recall that the equilibrium is pinned down by the value of $x = \frac{b_t F(R)\pi}{\tau}$ which solves the fixed point $G(x) = 0$, where

$$G(x) = \frac{\pi F(R)\sigma_\omega^2}{\tau\sigma_\omega^2 + \tau x^2\sigma_\varepsilon^2} - x.$$

Applying the chain rule, this implies:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}\pi} &= -\frac{\partial G/\partial \pi}{\partial G/\partial x} = \frac{F(R)\sigma_\omega^2\left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)}{2\pi F(R)x\sigma_\omega^2\sigma_\varepsilon^2 + \tau\left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)^2} > 0, \quad \text{and}\\ \frac{\mathrm{d}x}{\mathrm{d}R} &= -\frac{\partial G/\partial R}{\partial G/\partial x} = \frac{\pi\sigma_\omega^2\left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)}{2\pi F(R)x\sigma_\omega^2\sigma_\varepsilon^2 + \tau\left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)^2}F'(R) > 0.\end{aligned}$$

Since $\mathcal{I}_{IC} = \mathbb{E}[\sigma_\omega^2 - \mathbb{V}[\omega|t]] = \sigma_\omega^2 - \mathbb{V}[\omega|t] = \sigma_\omega^2 - \frac{1}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}}$ is decreasing in x, this implies $\mathcal{I}_{IC}$ is *decreasing* in both π and R.

Next, note that

$$\begin{aligned}\mathcal{I}_P &= \mathbb{E}[\sigma_\omega^2 - \mathbb{V}[\omega|\lambda,\psi]]\\ &= \sigma_\omega^2 - ((1 - F(R))\sigma_\omega^2 + F(R)\sigma_\omega^2(1 - b_t))\\ &= \sigma_\omega^2 F(R) b_t.\end{aligned}$$

This implies that

$$\frac{\mathrm{d}\mathcal{I}_P}{\mathrm{d}R} = \frac{\mathrm{d}}{\mathrm{d}R}\sigma_\omega^2 F(R) b_t = \sigma_\omega^2 \frac{\mathrm{d}}{\mathrm{d}R}\left(\frac{\tau x}{\pi}\right) > 0.$$

Similarly, we have

$$\begin{aligned}\frac{\mathrm{d}\mathcal{I}_P}{\mathrm{d}\pi} &= \sigma_\omega^2 \frac{\mathrm{d}}{\mathrm{d}\pi}\left(\frac{\tau x}{\pi}\right)\\ &= \sigma_\omega^2 \tau\left(\frac{1}{\pi}\left(\frac{\mathrm{d}x}{\mathrm{d}\pi}\right) - \frac{x}{\pi^2}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sigma_\omega^2\tau\left(\frac{1}{\pi}\left(\frac{F(R)\sigma_\omega^2\left(\sigma_\omega^2+x^2\sigma_\varepsilon^2\right)}{2\pi F(R)x\sigma_\omega^2\sigma_\varepsilon^2+\tau\left(\sigma_\omega^2+x^2\sigma_\varepsilon^2\right)^2}\right)-\frac{x}{\pi^2}\right)\\
&= \frac{\sigma_\omega^2\tau}{\pi^2}\left(-\frac{2\pi F(R)\sigma_\omega^2x^2\sigma_\varepsilon^2}{2\pi F(R)x\sigma_\omega^2\sigma_\varepsilon^2+\tau\left(\sigma_\omega^2+x^2\sigma_\varepsilon^2\right)^2}\right)\\
&< 0.
\end{aligned}
$$

□

A.3 Proof of Proposition 3

Let $\bar{x}$ be the unique value that solves $x=\frac{\sigma_\omega^2}{\tau\sigma_\omega^2+\tau x^2\sigma_\varepsilon^2}$, which is the unique root of $G(x)$ defined in the proof of Proposition 1 when $\pi F(R)=1$. Because x is strictly increasing in π and R as shown in the proof of Proposition 2, $\bar{x}\geq x$ for any π and $F(R)$. Moreover, because $\max\pi=1$ and $\sup_R F(R)=1$,[25] $\bar{x}$ is the supremum of the function $x(\pi,R)$. Let

$$
\underline{\beta}\equiv\frac{\tau}{2\bar{x}\sigma_\varepsilon^2}\qquad\text{and}\qquad\overline{\beta}\equiv\frac{\tau\left(\sigma_\omega^2+\bar{x}^2\sigma_\varepsilon^2\right)^2}{2\sigma_\omega^4\bar{x}\sigma_\varepsilon^2}.
$$

It is clear that $\underline{\beta}<\overline{\beta}$.

Recall that real efficiency is given by

$$
RE=(1-y^*)-\beta\left(\frac{1}{\frac{1}{\sigma_\omega^2}+\frac{1}{x^2\sigma_\varepsilon^2}}\right)=\tau x-\frac{\beta}{\frac{1}{\sigma_\omega^2}+\frac{1}{x^2\sigma_\varepsilon^2}}
$$

which is a univariate function of x. It follows that

$$
\frac{\partial RE}{\partial x} = \tau-\frac{2\beta x\sigma_\omega^4\sigma_\varepsilon^2}{\left(\sigma_\omega^2+x^2\sigma_\varepsilon^2\right)^2}. \tag{19}
$$

When $\beta\leq\underline{\beta}$, for all $x\leq\bar{x}$,

$$
\begin{aligned}
\frac{\partial RE}{\partial x} &= \tau-\frac{2\beta x\sigma_\omega^4\sigma_\varepsilon^2}{\left(\sigma_\omega^2+x^2\sigma_\varepsilon^2\right)^2}\\
&\geq \tau-\frac{\tau\frac{x}{\bar{x}}\sigma_\omega^4}{\left(\sigma_\omega^2+x^2\sigma_\varepsilon^2\right)^2}\\
&\geq \tau-\frac{\tau\sigma_\omega^4}{\left(\sigma_\omega^2+x^2\sigma_\varepsilon^2\right)^2}\\
&= \tau\left(1-\frac{1}{\left(1+\frac{x^2\sigma_\varepsilon^2}{\sigma_\omega^2}\right)^2}\right)
\end{aligned}
$$

[25] R can span over the support of the whistleblowing cost c.

$$> \quad 0,$$

where the first inequality follows from $\beta \leq \underline{\beta}$ and the definition of $\underline{\beta}$, and the second inequality follows from $\bar{x} \geq x$. It follows from $\mathrm{d}x/\mathrm{d}\pi, \mathrm{d}x/\mathrm{d}R > 0$ that real efficiency RE increases in both π and R.

Next, we show that RE is hump-shaped in π and R when $\beta > \overline{\beta}$. By (19), $\partial RE/\partial x > 0$ if and only if

$$\tau \left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)^2 > 2\beta x \sigma_\omega^4 \sigma_\varepsilon^2.$$

Let $H(x) \equiv \tau\left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)^2 - 2\beta x\sigma_\omega^4\sigma_\varepsilon^2$. First, note that

$$H(0) = \tau\sigma_\omega^4 > 0.$$

Also, we have

$$\frac{\partial H}{\partial x} = 2\sigma_\varepsilon^2\left(2\tau\left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)x - \beta\sigma_\omega^4\right).$$

Because $2\tau\left(\sigma_\omega^2 + x^2\sigma_\varepsilon^2\right)x - \beta\sigma_\omega^4$ is strictly increasing in x, is negative when $x = 0$ and positive when $x \to \infty$, H is first strictly decreasing and then strictly increasing in x. When $\beta > \overline{\beta}$, we have

$$\begin{aligned} H(\bar{x}) &= \tau\left(\sigma_\omega^2 + \bar{x}^2\sigma_\varepsilon^2\right)^2 - 2\beta\bar{x}\sigma_\omega^4\sigma_\varepsilon^2 \\ &< \tau\left(\sigma_\omega^2 + \bar{x}^2\sigma_\varepsilon^2\right)^2 - \tau\left(\sigma_\omega^2 + \bar{x}^2\sigma_\varepsilon^2\right)^2 \\ &= 0, \end{aligned}$$

where the inequality follows from $\beta > \overline{\beta}$ and the definition of $\overline{\beta}$. Because $\bar{x}$ is the supremum of x, there exists some π and R such that $H(x(\pi, R)) < 0$ by the continuity of H. It follows that there is $\hat{x} \in (0, \bar{x})$ such that $H(\hat{x}) = 0$ by the Intermediate value theorem. Moreover, because H is first strictly decreasing and then strictly increasing in x, $H(0) > 0$, and $H(\bar{x}) < 0$, $\hat{x}$ is the unique solution to $H(x) = 0$ in $[0, \bar{x}]$. It follows that $H(x) > 0$ and $\partial RE/\partial x > 0$ for $0 \leq x < \hat{x}$; while $H(x) < 0$ and $\partial RE/\partial x < 0$ for $\hat{x} < x < \bar{x}$. By $\mathrm{d}x/\mathrm{d}\pi, \mathrm{d}x/\mathrm{d}R > 0$, we conclude that real efficiency RE is hump-shaped in both π and R. □

A.4 Proof of Proposition 4

When $\beta > \overline{\beta}$, the maximum of real efficiency is attained at $x = \hat{x}$ as shown in Proposition 3. If the whistleblower reward R is given by $R = \alpha\pi$, the optimal level of penalty π^*, when interior, is determined by the following equation:

$$\hat{x} = \frac{\pi^* F(\alpha\pi^*)\sigma_\omega^2}{\tau\sigma_\omega^2 + \tau\hat{x}^2\sigma_\varepsilon^2}.$$

Let

$$Q \equiv \frac{\pi F(\alpha\pi)\sigma_\omega^2}{\tau\sigma_\omega^2 + \tau\hat{x}^2\sigma_\varepsilon^2} - \hat{x}.$$

It follows from the Implicit function theorem that

$$\frac{\partial\pi^*}{\partial x} = -\frac{\partial Q/\partial x}{\partial Q/\partial\pi}\bigg|_{\pi=\pi^*} = \frac{2\pi F(\alpha\pi)\sigma_\omega^2\hat{x}\sigma_\varepsilon^2 + \tau(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2)^2}{(\alpha\pi F'(\alpha\pi) + F(\alpha\pi))\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)\sigma_\omega^2}\bigg|_{\pi=\pi^*} > 0.$$

So π^* is strictly increasing in x. Next, we consider how $\hat{x}$ changes with the parameters $\tau, \beta, \sigma_\varepsilon, \sigma_\omega$.

Recall that $H(\hat{x}) = 0$ and H is decreasing at $x = \hat{x}$. That is,

$$H(\hat{x}) = \tau\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)^2 - 2\beta\hat{x}\sigma_\omega^4\sigma_\varepsilon^2 = 0$$

and

$$\frac{\partial H}{\partial x}\bigg|_{x=\hat{x}} = 2\sigma_\varepsilon^2\left(2\tau\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)\hat{x} - \beta\sigma_\omega^4\right) < 0.$$

This implies that

$$2\tau\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)\hat{x} - \beta\sigma_\omega^4 = \frac{\beta\sigma_\omega^4\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)}{\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2} < 0.$$

It follows that the second order condition is satisfied at $x = \hat{x}$:

$$\frac{\partial^2 RE}{\partial x^2}\bigg|_{x=\hat{x}} = \frac{2\beta\sigma_\omega^4\sigma_\varepsilon^2\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)}{\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)^3} < 0.$$

Given that $(\partial RE/\partial x)|_{x=\hat{x}} = 0$, the derivatives then follow from the Implicit function theorem:

$$\begin{aligned}
\frac{\mathrm{d}\hat{x}}{\mathrm{d}\tau} &= -\frac{\partial^2 RE/\partial x\partial\tau}{\partial^2 RE/\partial x^2}\bigg|_{x=\hat{x}} = -\frac{\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)^3}{2\beta\sigma_\omega^4\sigma_\varepsilon^2\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)} > 0 \\
\frac{\mathrm{d}\hat{x}}{\mathrm{d}\beta} &= -\frac{\partial^2 RE/\partial x\partial\beta}{\partial^2 RE/\partial x^2}\bigg|_{x=\hat{x}} = \frac{\hat{x}\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)}{\beta\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)} < 0 \\
\frac{\mathrm{d}\hat{x}}{\mathrm{d}\sigma_\varepsilon} &= -\frac{\partial^2 RE/\partial x\partial\sigma_\varepsilon}{\partial^2 RE/\partial x^2}\bigg|_{x=\hat{x}} = \frac{2\hat{x}\left(\sigma_\omega^2 - \hat{x}^2\sigma_\varepsilon^2\right)}{\sigma_\varepsilon\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)} = \frac{4\hat{x}^3\sigma_\varepsilon}{3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2} - \frac{2\hat{x}}{\sigma_\varepsilon} < 0
\end{aligned}$$

$$\frac{\mathrm{d}\hat{x}}{\mathrm{d}\sigma_\omega} = -\frac{\partial^2 RE/\partial x \partial \sigma_\omega}{\partial^2 RE/\partial x^2}\bigg|_{x=\hat{x}} = \frac{4\hat{x}^3\sigma_\varepsilon^2}{\sigma_\omega\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)} < 0.$$

Finally, we prove the comparative statics. By the chain rule and the Implicit function theorem, we have

$$\begin{aligned}
\frac{\mathrm{d}\pi^*}{\mathrm{d}\tau} &= \frac{\partial \pi^*}{\partial \tau} + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\tau}\right) \\
&= -\frac{\partial Q/\partial \tau}{\partial Q/\partial \pi}\bigg|_{\pi=\pi^*} + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\tau}\right) \\
&= \frac{1}{\alpha\pi F'(\alpha\pi) + F(\alpha\pi)}\left(\frac{\pi F(\alpha\pi)}{\tau} - \frac{(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2)^2\left(2\pi F(\alpha\pi)\sigma_\omega^2\hat{x}\sigma_\varepsilon^2 + \tau(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2)^2\right)}{2\beta\sigma_\omega^6\sigma_\varepsilon^2\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)}\right)\bigg|_{\pi=\pi^*} \\
&> 0,
\end{aligned}$$

since $3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2 < 0$. So π^* is increasing in τ. Note that

$$\begin{aligned}
\frac{\mathrm{d}\pi^*}{\mathrm{d}\beta} &= \frac{\partial \pi^*}{\partial \beta} + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\beta}\right) \\
&= -\frac{\partial Q/\partial \beta}{\partial Q/\partial \pi} + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\beta}\right) \\
&= 0 + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\beta}\right) \\
&< 0.
\end{aligned}$$

So π^* is decreasing in β. Note that

$$\begin{aligned}
\frac{\mathrm{d}\pi^*}{\mathrm{d}\sigma_\varepsilon} &= \frac{\partial \pi^*}{\partial \sigma_\varepsilon} + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\sigma_\varepsilon}\right) \\
&= -\frac{\partial Q/\partial \sigma_\varepsilon}{\partial Q/\partial \pi} + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\sigma_\varepsilon}\right) \\
&= \frac{2\hat{x}\left(\pi F(\alpha\pi)\sigma_\omega^2\hat{x}\sigma_\varepsilon^2\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right) + \left(2\pi F(\alpha\pi)\sigma_\omega^2\hat{x}\sigma_\varepsilon^2 + \tau(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2)^2\right)\left(\sigma_\omega^2 - \hat{x}^2\sigma_\varepsilon^2\right)\right)}{\left(\alpha\pi F'(\alpha\pi) + F(\alpha\pi)\right)\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)\sigma_\omega^2\sigma_\varepsilon\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)}\Bigg|_{\pi=\pi^*} \\
&= \frac{2\hat{x}\left(\pi F(\alpha\pi)\sigma_\omega^2\hat{x}\sigma_\varepsilon^2\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right) + \tau(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2)^2\left(\sigma_\omega^2 - \hat{x}^2\sigma_\varepsilon^2\right)\right)}{\left(\alpha\pi F'(\alpha\pi) + F(\alpha\pi)\right)\left(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2\right)\sigma_\omega^2\sigma_\varepsilon\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)}\Bigg|_{\pi=\pi^*} \\
&< 0,
\end{aligned}$$

which follows from $\sigma_\omega^2 - \hat{x}^2\sigma_\varepsilon^2 > \sigma_\omega^2 - 3\hat{x}^2\sigma_\varepsilon^2 > 0$. So π^* is decreasing in σ_ε. Note that

$$\begin{aligned}
\frac{\mathrm{d}\pi^*}{\mathrm{d}\sigma_\omega} &= \frac{\partial \pi^*}{\partial \sigma_\omega} + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\sigma_\omega}\right) \\
&= -\frac{\partial Q/\partial \sigma_\omega}{\partial Q/\partial \pi} + \left(\frac{\partial \pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\sigma_\omega}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \left.\frac{2\hat{x}^2\sigma_\varepsilon^2\left(\pi F(\alpha\pi)\sigma_\omega^2 + 2\tau\hat{x}(\sigma_\omega^2 + \hat{x}^2\sigma_\varepsilon^2)\right)}{\left(\alpha\pi F'(\alpha\pi) + F(\alpha\pi)\right)\sigma_\omega^3\left(3\hat{x}^2\sigma_\varepsilon^2 - \sigma_\omega^2\right)}\right|_{\pi=\pi^*} \\
&< 0.
\end{aligned}
$$

So π^* is decreasing in σ_ω. Note that

$$
\begin{aligned}
\frac{\mathrm{d}\pi^*}{\mathrm{d}\alpha} &= \frac{\partial\pi^*}{\partial\alpha} + \left(\frac{\partial\pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\alpha}\right) \\
&= -\frac{\partial Q/\partial\alpha}{\partial Q/\partial\pi} + \left(\frac{\partial\pi^*}{\partial x}\right)\left(\frac{\mathrm{d}\hat{x}}{\mathrm{d}\alpha}\right) \\
&= -\frac{\pi^2 F'(\alpha\pi)}{\alpha\pi F'(\alpha\pi) + F(\alpha\pi)} + 0 \\
&< 0.
\end{aligned}
$$

So π^* is decreasing in α. □

A.5 Proof of Proposition 5

By matching the coefficients, $x = \frac{\delta p_t F(R)}{\tau}$ solves

$$
\frac{\frac{1}{x^2\sigma_\varepsilon^2}}{\frac{1}{\sigma_\omega^2} + \frac{1}{x^2\sigma_\varepsilon^2}} = p_t = \frac{\tau x}{\delta F(R)}.
$$

Equivalently, x solves $\tilde{G}(x) = 0$, where

$$
\tilde{G}(x) \equiv \frac{\delta F(R)\sigma_\omega^2}{\tau\sigma_\omega^2 + \tau x^2\sigma_\varepsilon^2} - x.
$$

This is identical to $G(x)$ in Proposition 1, except that π is replaced with δ. Using the same argument as before, we have the results as stated in the proposition. □

A.6 Proof of Proposition 6

The results follow from the same argument as the proofs of Propositions 2 and 3 by replacing the penalty π with the payoff sensitivity to price δ and the coefficient of informativeness b_t with p_t. The definitions of $\underline{\beta}$ and $\overline{\beta}$ are the same as in Proposition 3. Furthermore, the price P has a one-to-one correspondence with the public information ψ conditional on whether there is a leak, leading to a similar expression for $\mathcal{I}_P$. □