

SELF-REPORTING SCHEMES AND CORPORATE CRIME*

Charles Angelucci[†] Martijn Han[‡]

December 2015

Abstract

We study the design of self-reporting schemes for corporate crimes. We model a welfare-maximizing authority and a continuum of firms, where each firm has one employer and one employee. Employees decide whether to take an action beneficial to themselves but harmful to their employer and society. Employers design contracts that either prevent the harmful act or tolerate it. We investigate the optimal corporate and individual sanctions to impose on firms whose employees confess to having committed the harmful act. In an extension, we let employers monitor their employees, and investigate the sanctions to impose when employers report their own employees.

Keywords: corporate crime, white-collar crime, self-reporting, leniency, antitrust

JEL Classification Numbers: K21, K42, L40

*This paper formerly circulated as “Private and Public Control of Management.” We are grateful to many for extremely useful comments, including Maria Bigoni, Giuseppe Dari-Mattiacci, Bernhard Ganglmair, Oliver Hart, Giovanni Immordino, Bruno Jullien, David Martimort, Massimo Motta, Patrick Rey, Daniel Sokol, Maarten Pieter Schienkel, Giancarlo Spagnolo, Kathryn Spier, Roland Strausz, Jan Tuinstra, and Wouter Wils. We are grateful to participants at CRESSE 2011, CLEEN 2011, IIOC 2011, AEA 2011, the University of Amsterdam, Toulouse, Harvard, MIT, the Max Planck Institute in Munich, and UFC in Besançon. This paper received the “Robert F. Lanzillotti” Prize in 2011. Usual disclaimers apply.

[†]Division of Finance and Economics, Columbia Business School. Email: ca2630@gsb.columbia.edu

[‡]Institute for Economic Theory I, Humboldt-Universität zu Berlin. Email: martijn.alexander.han@hu-berlin.de

1 Introduction

Corporate crime frequently makes the headlines and invariably triggers debates among policy makers, economists, and legal scholars alike. Employee misconduct ranges from violations of environmental or safety rules, to asset misappropriation, self-dealing, the leaking or misuse of information, and antitrust infringements. The harm to firms and society is often significant, and new measures are regularly introduced to increase deterrence.¹ In the past 20 years, many regulatory agencies have implemented self-reporting schemes, whereby firms or individuals voluntarily disclose incriminating information in exchange for a reduced sanction.² These schemes have been at the forefront of the fight against corporate crime. For instance, the Environmental Protection Agency reports it “continues to receive hundreds of new disclosures every year,” and the Department of Justice states its “Leniency Program is its most important investigative tool for detecting cartel activity.”³ In this paper, we provide a simple theoretical framework to analyze self-reporting schemes in the context of corporate crime, and propose adjustments to existing schemes.

We model the interaction between a welfare-maximizing authority and a continuum of firms. Each firm consists of one employer and one employee, and employees decide whether to take an action beneficial to themselves (e.g., not complying with time-consuming safety rules) but harmful to their employers and the rest of society (e.g., because it increases the chances of an accident). The gain from the harmful act differs across employees, and employers design employment contracts that either prevent misconduct or let it occur. Employers tolerate the harmful act when preventing it is too costly. The authority audits (at some cost) a subset of randomly chosen firms whose activities led to some publicly observable event (say an accident), and these audits result in either a conviction—in which case, individual and corporate sanctions are imposed—or an acquittal.

We first analyze the version of the model in which employees who commit the harmful act can report their behavior to the authority. We take as given the sanctions employers and employees face in the absence of a report, and investigate whether the authority should reduce either or both

¹Using survey data for the year 2014, auditor firm PricewaterhouseCoopers reports that 46% of U.S. firms suffered employee fraud in the 2013-14 period, with costs exceeding of \$100,000 in 54% of the cases and \$5 million in 8% of the cases (<http://www.pwc.com/gx/en/economic-crime-survey/>).

²See, for instance, the EPA’s *Audit Policy*, the Department of Justice’s *Corporate Leniency Program*, the European Commission’s revised *Leniency Program*, the Department of Defense’s *Contractor Disclosure program*, and the Federal Energy Regulatory Commission’s *Self-reporting Scheme*.

³For the EPA’s Audit Policy, visit <https://www.gpo.gov/fdsys/pkg/FR-2015-12-09/pdf/2015-30928.pdf>. For the DOJ’s Leniency Program, visit <http://www.justice.gov/atr/leniency-program>.

sanctions in case an employee comes forward—prior to an accident—with incriminating evidence. Our analysis highlights the distinct role of the corporate and individual sanctions. The corporate sanction enters the employers’ payoff, and thus affects their willingness to design contracts that prevent the harmful act in a simple manner. By contrast, the individual sanction affects the employees’ incentives, and thus has only an indirect effect—through wages—on employers.

We find granting partial amnesty to the employers whose employees report their misconduct is optimal. This policy induces employers who tolerate the harmful act to encourage their employees to “self-report,” which reduces enforcement costs. Moreover, the partial nature of the reduction ensures deterrence is not sacrificed in the process. In addition, we show that promising employees who report their misconduct a sanction strictly lower than the expected sanction they face when they remain silent allows the authority to raise deterrence without having to increase the frequency of its audits, that is, without having to sustain higher enforcement costs. This positive effect on deterrence occurs even though the reduction in the sanction makes the harmful act more tempting to employees. When employees are risk neutral, granting full individual amnesty is optimal.

We then extend the baseline model by allowing employers to monitor their employees and report evidence of misconduct to the authority. We find the authority implements both a *corporate reporting scheme* and an *employee self-reporting scheme*. In particular, we show that partially reducing the sanction imposed on employers who report their employees, while fully punishing these employees, is optimal because it increases the expected sanction employees face when they commit the harmful act but do not self-report. In addition, much like in the baseline model, granting partial amnesty to the employers whose employees report their own misconduct remains optimal because it reduces enforcement costs.

These results are in contrast to existing practice in both the United States and Europe. Corporate reporting schemes specify reduced corporate sanctions, but often also shield employees involved in the wrongdoing from individual sanctions. Moreover, employee reporting schemes are rare and have no systematic approach to dealing with corporate sanctions.⁴

Related Literature. Our paper is related to the literature that analyzes corporate crime and liability rules using the principal agent framework (see, e.g., Segerson and Tietenberg (1992),

⁴See, for instance, the Department of Justice’s *Individual Leniency Program*.

Polinsky and Shavell (1993), Arlen (1994, 2012), Shavell (1997), and Hiriart and Martimort (2006)). We take the maximum sanctions firms and individuals face as given, and investigate whether reducing these sanctions when employees report their misconduct is desirable. We thus also relate to the literature on self-reporting schemes (see, e.g., Kaplow and Shavell (1994), Innes (1999), Motta and Polo (2003), Spagnolo (2005), Harrington (2008, 2013), Miller (2009), Marvo and Spagnolo (2014), and Marshall, Marx, and Mezzetti (2015)). A large strand of this literature focuses on cartels. Similar to the case of cartels, our self-reporting scheme possesses a deterrent dimension because it makes coordination more difficult for players. However, unlike in cartels, employment contracts that act as coordination devices link our players. Aubert, Rey, and Kovacic (2006) argue whistleblowers should be rewarded when reporting their firm’s involvement in a cartel. Also, Aubert (2009) analyzes the impact of individual leniency programs on shareholders’ ability to affect managers’ incentives to exert effort and collude on prices. Focusing on “ordinary” crimes, Kaplow and Shavell (1994) show self-reporting schemes allow a judicial authority to save on enforcement costs without sacrificing deterrence. We extend their analysis to the case of corporate crime, and find that not only do self-reporting schemes allow to save on enforcement costs without weakening deterrence, they also allow to increase deterrence without raising costs. Finally, see Dyck, Morse, and Zingales (2011) for an empirical analysis of the detection of corporate fraud, and Abrantes-Metz and Sokol (2013) on corporate governance and antitrust violations.

We present the setting in section 2. In section 3, we solve the baseline model. We analyze corporate reporting schemes and discuss other extensions in section 4. We discuss the policy implications of our results in section 5. Section 6 concludes.

2 The Setting

Players, Actions, and Information

A population of firms of size normalized to one exists, as does an authority. Each firm is composed of one employer (she) and one employee (he). Employees choose whether to commit an act beneficial to themselves but detrimental to their employers and the rest of society (e.g., not abiding by safety rules, selling confidential information, over-prescribing a drug in return for a

payment, etc). The authority audits firms and, if necessary, imposes individual as well as corporate sanctions. Employees are risk averse, whereas employers are risk neutral. We assume employers find hiring employees to be profitable for reasons not modeled here.

Each employee privately chooses whether to commit ($e = 1$) or not commit ($e = 0$) the “harmful act.” If an employee commits the harmful act, he obtains a gain $G \in [0, +\infty)$.⁵ The gain G differs among employees (e.g., because of different costs of abiding by safety rules) and has density $f(\cdot)$ with cumulative distribution $F(\cdot)$. The gain to all employees in choosing the safe action is zero. We assume employers know their employees’ gain G .

An employee’s choice of e determines the distribution of a firm-specific public signal $\pi \in \{\underline{\pi}, \bar{\pi}\}$, where $\pi = \underline{\pi}$ ($\pi = \bar{\pi}$) denotes the occurrence (nonoccurrence) of an accident.⁶ Specifically, we assume $\Pr(\pi = \bar{\pi} | e = 0) = \Pr(\pi = \underline{\pi} | e = 1) = \rho > \frac{1}{2}$. Whether an accident occurs is publicly observable, and each occurrence is associated with one firm only. In other words, for any given accident, the authority knows which firm could have caused it (e.g., it knows which hospital/doctor was handling the treatment of a patient who unexpectedly died). However, we assume the authority must gather evidence of actual misconduct to impose sanctions. An alternative interpretation is as follows. Because we assume π enters the employers’ payoffs (see below), $\pi = \underline{\pi}$ can be interpreted as the automatic sanctioning of a firm that caused an accident. The authority may then wish to supplement these sanctions with additional ones, but must examine firms to do so.

Prior to the possible occurrence of accidents, employees have the opportunity to report to the authority that they committed the harmful act. Specifically, employees who committed the harmful act can either “self-report” or remain silent, whereas employees who did not commit the harmful act can only remain silent. Reports are publicly observable. For the firms whose employees remained silent, the authority observes whether an accident occurs, and audits with probability β each firm that caused an accident. Let $\tau \in \{a, c\}$ denote the outcome of the audit: the employer/employee pair is either convicted ($\tau = c$) or acquitted ($\tau = a$). Assume $\Pr(\tau = c | e = 1) = \alpha$ and $\Pr(\tau = a | e = 0) = 1$; the authority never convicts by mistake, but may acquit by mistake.

In section 4, we allow employers to monitor employees and report evidence to the authority.

⁵Depending on the application, G can be interpreted as the saved cost associated with abiding by safety rules, the money an employee receives in exchange for selling confidential information, the money a patient pays a doctor who is over-prescribing a drug, and so on.

⁶For certain applications, interpreting π as the occurrence of consumer complaints, bad publicity, and so on is more natural.

Payoffs

Employees. Let t denote a transfer an employee receives from his employer. Also, let s and S denote the monetary sanctions the authority imposes on, respectively, employees and employers in case of a conviction following an audit; s_r and S_r are the corresponding sanctions in case of a report. We assume s and S are exogenously given, but let the authority choose $s_r \geq 0$ and $S_r \geq 0$.

Employee *ex post* payoffs, denoted $U(G)$, are additively separable in the gain eG associated with action e and the utility from wealth $u(\cdot)$, whose argument is equal to the transfer t minus the individual sanction (if any). For instance, $U(G) = eG + u(t - s)$ if the authority imposes sanction s on the employee following an audit. We take $u(\cdot)$ to be increasing and concave, with $u(0) = 0$. To help intuition, we sometimes use the notation $u(-x) = -l(x)$, and refer to $l(\cdot)$ as a loss. Also, to obtain closed-form expressions, we often analyze the case in which employees' utility functions are linear in wealth (i.e., $u(x) = x$), that is, the case of risk neutral employees.

Employers. The realization of the firm-specific signal $\pi \in \{\underline{\pi}, \bar{\pi}\}$ enters employers' payoffs. Specifically, we assume the employers' *ex post* payoffs, denoted $\Pi(G)$, are separable in the realization π , the transfer t , and the corporate sanction (if any). For instance, $\Pi(G) = \pi - t - S$ if the authority imposes corporate sanction S on the employer following an audit. In the following, let $\bar{\pi} - \underline{\pi} = \Delta\pi > 0$, and let $\pi_0 = \mathbb{E}[\pi | e = 0]$ and $\pi_1 = \mathbb{E}[\pi | e = 1]$. An employer's expected gross profit is lower when $e = 1$ than when $e = 0$, because $\pi_0 - \pi_1 = (2\rho - 1)\Delta\pi > 0$.

Authority. Let h be the harm imposed onto third parties whenever an accident occurs (i.e., whenever $\pi = \underline{\pi}$), and let k be the authority's cost of each single audit.⁷ Collecting sanctions is costless, and they are transferred to third parties whose utility functions are linear in wealth. The authority chooses the scheme (β, s_r, S_r) to maximize social welfare, defined as the sum of all the employers' and employees' expected payoffs plus the expected total amount of collected sanctions, minus the expected level of harm, as well as the expected cost of audits.

To illustrate, when $u(x) = x$, and anticipating that employers prevent the harmful act if and

⁷To simplify expressions, we assume handling reports involves no administrative costs. What matters for our results is that the cost of handling a report is lower than that of launching an audit.

only if their employees' gain G is lower than a threshold $\hat{G}(\beta, s_r, S_r)$, social welfare is given by:⁸

$$W(\beta, s_r, S_r) = \int_0^{\hat{G}(\beta, s_r, S_r)} (\pi_0 - (1 - \rho)(h + \beta k)) dF(G) + \int_{\hat{G}(\beta, s_r, S_r)}^{\infty} (\pi_1 + G - \rho(h + \beta k)) dF(G). \quad (1)$$

When employees are risk averse, wages appear in the welfare function and the authority takes into account the consequent *wealth effects* of its decisions. Finally, note that, because $G \in [0, \infty)$, deterring all employees from committing the harmful act will not be socially optimal.⁹

Employment Contracts and Timing

Contracts. Each employer offers her employee a contract that specifies a transfer \bar{t} in case $\pi = \bar{\pi}$, and, in case $\pi = \underline{\pi}$, (i) a transfer t_r when the employee reports that he committed the harmful act, (ii) a transfer t_τ if an audit occurs (where $\tau = a, c$), and (iii) a transfer \underline{t} otherwise. We assume transfers must be nonnegative, and that employees must be guaranteed a nonnegative expected payoff. We also assume the authority does not regulate contracts ex ante. We discuss this assumption in Section 4. During the analysis, employers will sometimes be indifferent between several contracts. In these instances, we report the simplest contract in the main body and leave the complete characterization to the Appendix.

Note that we allow employers to *indemnify* their employees (i.e., $t_c, t_r \geq 0$). The extent to which corporations are able to indemnify their employees varies with the jurisdiction and the nature of the violation. However, it appears the discretion firms enjoy is somewhat ample (see, e.g., Mullin and Snyder (2010)). In practice, corporations may wish to promise indemnifications in exchange for having a say in the preparation of their employees' defense, for instance, to induce their employees to cooperate with the judicial system (or, on the contrary, to avoid the disclosure of sensitive information). In section 4, we discuss the cases in which indemnifications are infeasible.

Timing. The authority moves first by announcing the sanctions it will impose on employees and employers in case of a report. Employers then design their employment contracts—in essence, they choose whether to prevent or tolerate the harmful act, and all employees who accepted an employment offer choose whether to commit the harmful act. Employees who committed the

⁸Expression (1) also implicitly assumes no employee self-reports in equilibrium.

⁹If we had assumed $G \in [0, \bar{G}]$, deterring all harm might still not be optimal, because of enforcement costs.

harmful act choose whether to report it to the authority, in which case, the authority imposes sanctions s_r and S_r . For the remaining firms, accidents possibly occur and the authority audits a fraction of the firms that caused an accident (and possibly imposes sanctions s and S).¹⁰

We end the description with two last assumptions. We assume employees who are indifferent between various strategies adopt their employer's preferred strategy. Similarly, we assume employers who are indifferent between various strategies adopt the authority's preferred strategy. These assumptions are unimportant insofar as both the authority and the employers are able to break, respectively, the employers' and the employees' indifference at an arbitrarily small cost.

3 Solving the Model

The Employers' Problem

We solve separately for the contracts chosen by the employers who wish to prevent the harmful act, and those chosen by the employers who let it occur.

Preventing the Harmful Act

Consider a given firm, and suppose the employer wishes to prevent the harmful act. To design her contract, the employer chooses the transfers $\{\bar{t}, \underline{t}, t_c, t_a, t_r\}$ to maximize her expected payoff:

$$\rho(\bar{\pi} - \bar{t}) + (1 - \rho)(\underline{\pi} - \beta t_a - (1 - \beta)\underline{t}), \quad (2)$$

subject to ensuring the employee does not commit the harmful act:

$$\begin{aligned} \rho u(\bar{t}) + (1 - \rho)(\beta u(t_a) + (1 - \beta)u(\underline{t})) \geq \\ G + \rho(\beta(\alpha u(t_c - s) + (1 - \alpha)u(t_a)) + (1 - \beta)u(\underline{t})) + (1 - \rho)u(\bar{t}), \end{aligned} \quad (3)$$

or commit the harmful act and subsequently report to the authority that he did:

$$\rho u(\bar{t}) + (1 - \rho)(\beta u(t_a) + (1 - \beta)u(\underline{t})) \geq G + u(t_r - s_r). \quad (4)$$

¹⁰We assume a report by an employee does not affect the likelihood of an accident. Assuming a report decreases the chances of an accident can only increase the desirability of self-reporting schemes (see, e.g., Innes (1999)).

The employer must also guarantee the employee a nonnegative expected payoff, and set all transfers greater than or equal to zero. The next lemma states the optimal contract when employees are *not* tempted to report that they committed the harmful act. Its proof as well as all other proofs are in the appendix.

Lemma 1 *Suppose $l(s_r) \geq \rho\beta\alpha l(s)$, and consider a given employer preventing the harmful act.*

1. *If the employee were to commit the harmful act, he would be better off not reporting it.*
2. *It is optimal for the employer to set*
 - (a) *All transfers equal to zero if $G \leq \rho\beta\alpha l(s)$,*
 - (b) *And otherwise either*
 - i. *$\bar{t}(G) > 0$, and all other transfers equal to zero if $\alpha \leq \tilde{\alpha} = \frac{2\rho-1}{\rho}$, or*
 - ii. *$\bar{t}(G), t_a(G) > 0$, and all other transfers equal to zero if $\alpha > \tilde{\alpha}$.*
3. *The associated expected transfer is equal to zero when $G \in [0, \rho\beta\alpha l(s)]$, and is strictly and continuously increasing in G for $G \in [\rho\beta\alpha l(s), \infty)$.*

To gain intuition, observe that, for all employers who prevent the harmful act, setting $t_c = 0$ and $t_r = 0$ to relax, respectively, (3) and (4) is optimal; employees are punished with a transfer equal to zero when convicted, or after having reported misconduct.¹¹ Employees' payoff when reporting their misconduct is thus equal to $-l(s_r)$, so that, when $l(s_r) \geq \rho\beta\alpha l(s)$, they prefer to remain silent and enjoy both a lower expected sanction and possible transfers from their employer.

All employees whose gain G is lower than $\rho\beta\alpha l(s)$ have no desire to commit the harmful act. It is then sufficient for their employers to set all transfers equal to zero. By contrast, employees whose gain G is above $\rho\beta\alpha l(s)$ have private incentives to misbehave. Preventing the harmful act is then costly, and the optimal way to prevent it consists of making one positive transfer in the absence of an accident, and possibly another one in case of an acquittal (if the audit technology is sufficiently accurate; i.e., if $\alpha > \tilde{\alpha}$).¹² Finally, the cost to an employer of preventing the harmful act is (weakly) increasing in her employee's gain G from the harmful act.

¹¹Setting $t_c = 0$ and $t_r = 0$ is, in some cases, only *weakly* optimal. We focus on this solution to help intuition and keep lemmas short. The complete characterization can be found in the appendix.

¹²Building on the medical malpractice example, this contract would amount to hospitals ensuring doctors' pay is negatively correlated with, say, the mortality rate of their patients. In case of a death that triggers an enquiry, one might think the doctor should not be penalized in her pay when cleared of any suspicion of negligence.

In what follows, we use the superscript “1” when referring to the transfers chosen by the employers who prevent the harmful act when $l(s_r) \geq \rho\beta\alpha l(s)$ (e.g., \bar{t}^1).

The next lemma states the properties of the optimal contract when employees who misbehave may be tempted to report they committed the harmful act.

Lemma 2 *Suppose $l(s_r) < \rho\beta\alpha l(s)$, and consider a given employer preventing the harmful act. A threshold $\tilde{G}(\beta, s_r) > l(s_r)$ exists such that it is optimal for the employer to set:*

1. All transfers equal to zero if $G \leq l(s_r)$,
2. $\bar{t}(G, s_r) > 0$, $t_a(G, s_r) \geq 0$, $\underline{t}(G, s_r) \geq 0$, and $t_c = t_r = 0$ such that inequality (4) binds if $l(s_r) < G \leq \tilde{G}(\beta, s_r)$,
3. And otherwise either
 - (a) $\bar{t}^1(G) > 0$, and all other transfers equal to zero if $\alpha \leq \tilde{\alpha}$, or
 - (b) $\bar{t}^1(G), t_a^1(G) > 0$, and all other transfers equal to zero if $\alpha > \tilde{\alpha}$.

The associated expected transfer is equal to zero when $G \in [0, l(s_r)]$, and is strictly and continuously increasing in G for $G \in [l(s_r), \infty)$.

Setting s_r so that $l(s_r) < \rho\beta\alpha l(s)$ does not imply that all employees—upon deviating and committing the harmful act—find reporting the act to the authority to be profitable. Because employees who remain silent can hope to receive positive transfers from their employer in some states of nature (e.g., in the absence of an accident), as opposed to when reporting to the authority, the loss $l(s_r)$ must be sufficiently smaller than the expected loss $\rho\beta\alpha l(s)$ to trigger a report.

To illustrate, consider, for instance, an employer offering the contract stated in Lemma 1, for the case in which $G > \rho\beta\alpha l(s)$ and $\alpha \leq \tilde{\alpha}$. The employee, if he deviates and commits the harmful act, is better off remaining silent if:

$$(1 - \rho) u(\bar{t}^1(G)) - \rho\beta\alpha l(s) \geq -l(s_r), \quad (5)$$

that is, if the reduction in the expected loss from a sanction is lower than the expected utility derived from his employer’s transfers. From Lemma 1, we know the transfer $\bar{t}^1(G)$ is increasing in

G . It follows that inequality (5) holds for all employees whose gain G is higher than some threshold, which we denote $\tilde{G}(\beta, s_r)$. A qualitatively equivalent condition holds for the case in which $\alpha > \tilde{\alpha}$.

Setting $l(s_r)$ below $\rho\beta\alpha l(s)$ does not hurt all employers who prevent the harmful act. Specifically, employers whose employees' gain $G \geq \tilde{G}(\beta, s_r)$ are unaffected: they offer the same contract they would offer when $l(s_r) \geq \rho\beta l(s)$. This is not true, however, for the employers whose employees' gain from committing the harmful act is intermediate (i.e., $l(s_r) < G < \tilde{G}(\beta, s_r)$). These employers are concerned their employees will commit the harmful act and subsequently report it.¹³ The least costly way of discouraging such behavior consists of setting transfers in a way that guarantees a payoff equal to $G - l(s_r)$ in equilibrium, while promising transfers equal to zero in case of a report or a conviction. Finally, employees whose gain G is lower than or equal to the loss $l(s_r)$ have no private incentives to misbehave, and preventing the harmful act is costless. Once again, the cost to an employer of preventing the harmful act is (weakly) increasing in the gain G .

We again use the superscript "1" for the transfers chosen by the employers whose employees' gain $G \geq \tilde{G}(\beta, s_r)$ when $l(s_r) < \rho\beta l(s)$. We use the superscript "2" for those chosen by the employers whose employees' gain $G < \tilde{G}(\beta, s_r)$, and denote by $\mathbb{E}[t^2(G, s_r)]$ the associated expected transfer.

Risk Neutrality. We state the solution to the employers' problem when $u(x) = x$. For the sake of brevity, we focus on the case in which $\alpha \leq \frac{\rho^2 - (1-\rho)^2}{\rho^2}$. This restriction ensures it is optimal for the employers who prevent the harmful act to reward employees in the absence of accidents. We maintain it throughout the paper whenever we concentrate on the case of risk neutrality. Results are qualitatively identical when $\alpha > \frac{\rho^2 - (1-\rho)^2}{\rho^2}$. Finally, the contracts we state are in some cases only weakly optimal; the proofs in the Appendix provide a complete characterization.

When $s_r \geq \rho\beta\alpha s$, employees have no incentives to self-report, and the optimal contract consists of setting $\bar{t}^1(G) = \max\left[\frac{G - \rho\beta\alpha s}{(2\rho - 1)}, 0\right]$, and all other transfers to zero. This contract best satisfies (3) and verifies (4) when $s_r \geq \rho\beta\alpha s$. The associated expected transfer $\frac{\rho}{(2\rho - 1)} \max[(G - \rho\beta\alpha s), 0]$ is increasing in G . When $s_r < \rho\beta\alpha s$, employees may or may not have incentives to self-report. Specifically, it is optimal for the employers to set $\bar{t}^1(G) = \frac{G - \rho\beta\alpha s}{(2\rho - 1)}$ and all other transfers to zero when their employees' gain G is such that:

$$\frac{(1 - \rho)}{(2\rho - 1)} (G - \rho\beta\alpha s) - \rho\beta\alpha s \geq -s_r, \quad (6)$$

¹³When $l(s_r) < G < \tilde{G}(\beta, s_r)$, it can also happen that employers must deter *both* deviations at the same time.

that is, when $G \geq \tilde{G}(\beta, s_r) = \frac{1}{1-\rho} (\rho^2 \beta \alpha s - (2\rho - 1) s_r) > \rho \beta \alpha s$. When $G \geq \tilde{G}(\beta, s_r)$, satisfying inequality (3) implies inequality (4). In other words, even though $s_r < \rho \beta \alpha s$, an employee deviating and committing the harmful act may choose to remain silent, hoping to pocket the transfer $\frac{G - \rho \beta \alpha s}{(2\rho - 1)}$. Further, employers whose employees' gain G is lower than $\tilde{G}(\beta, s_r)$ but higher than s_r set $\bar{t}^2(G, s_r) = \frac{G - s_r}{\rho}$ and all other transfers equal to zero. This contract is such that (3) holds when $G \leq \tilde{G}(\beta, s_r)$. Employees are promised an expected transfer equal to their net gain from committing the harmful act *and* self-reporting. Finally, employers whose employees' gain $G \leq s_r$ set all transfers equal to zero. Much like the case in which $s_r \geq \rho \beta \alpha s$, the expected transfer is increasing in G .

Tolerating the Harmful Act

Because preventing misconduct can be costly, some employers may be tempted to let it occur despite the higher probability of an accident and the prospect of legal sanctions. Further, employers who tolerate the harmful act must decide whether to encourage or discourage self-reporting. Consider first the problem faced by an employer who tolerates the harmful act but discourages her employee from self-reporting. The employer chooses transfers $\{\bar{t}, \underline{t}, t_c, t_a, t_r\}$ to maximize:

$$\rho(\underline{\pi} - \beta(\alpha t_c + (1 - \alpha) t_a) - (1 - \beta) \underline{t}) + (1 - \rho)(\bar{\pi} - \bar{t}) - \rho \beta \alpha S, \quad (7)$$

subject to ensuring the employee does not take the safe action:

$$G + \rho(\beta(\alpha u(t_c - s) + (1 - \alpha) u(t_a)) + (1 - \beta) u(\underline{t})) + (1 - \rho) u(\bar{t}) \geq \rho u(\bar{t}) + (1 - \rho)(\beta u(t_a) + (1 - \beta) u(\underline{t})), \quad (8)$$

or report to the regulator after having committed the harmful act:

$$G + \rho(\beta(\alpha u(t_c - s) + (1 - \alpha) u(t_a)) + (1 - \beta) u(\underline{t})) + (1 - \rho) u(\bar{t}) \geq G + u(t_r - s_r). \quad (9)$$

As before, the employer must also guarantee the employee a nonnegative expected payoff, and set all transfers greater than or equal to zero.

Because employers suffer from lower expected profits as well as possible legal sanctions when their employees misbehave, we can safely anticipate no employer would prefer her employee to commit the harmful act when preventing it is costless. If $l(s_r) \geq \rho \beta \alpha l(s)$, preventing misconduct

is costless when $G \leq \rho\beta\alpha l(s)$. Similarly, if $l(s_r) < \rho\beta\alpha l(s)$, preventing misconduct is costless when $G \leq l(s_r)$. As a result, the next lemma focuses on the case in which $G > \min[l(s_r), \rho\beta\alpha l(s)]$.

Lemma 3 *Take a given firm and suppose the employer tolerates the harmful act. Suppose also the employer discourages her employee from self-reporting.*

1. *If $l(s_r) \geq \rho\beta\alpha l(s)$, setting all transfers equal to zero is optimal.*
2. *If $l(s_r) < \rho\beta\alpha l(s)$, setting all transfers equal to zero is optimal, except for $t_c(s_r) > 0$, which is set so that the employee does not self-report, i.e., $\rho\beta\alpha u(t_c - s) = u(-s_r)$.*

The employee's expected payoff is equal to $G - \min[l(s_r), \rho\beta\alpha l(s)]$.

When $l(s_r) \geq \rho\beta\alpha l(s)$, employees have no interest in reporting to the authority that they committed the harmful act. Because $G > \rho\beta\alpha l(s)$, setting all transfers equal to zero is optimal.

In case $l(s_r) < \rho\beta\alpha l(s)$, employers “bribe” their employees into remaining silent, which they achieve by indemnifying employees in case of a conviction, that is, by setting $\rho\beta\alpha u(t_c - s) = u(-s_r)$.¹⁴ Note the size of the transfer $t_c(s_r)$ depends on s_r but not G : setting $l(s_r) < \rho\beta\alpha l(s)$ therefore equally hurts all employers who tolerate the harmful act and discourage reports.

We use the superscript “3” to refer to the transfers chosen by the employers who tolerate the harmful act and discourage their employees from self-reporting (e.g., $t_c^3(s_r)$).

Risk Neutrality. The employers who discourage their employees from self-reporting set all transfers equal to zero if $s_r \geq \rho\beta\alpha s$, and otherwise $t_c^3(s_r) = s - \frac{s_r}{\rho\beta\alpha}$ and all other transfers to zero.

Consider now the employers who tolerate the harmful act and encourage their employees to self-report. The problem these employers solve is identical to that above, except they maximize expected payoff $\rho\underline{\pi} + (1 - \rho)\bar{\pi} - t_r - S_r$ and inequality (9) is replaced by:

$$G + u(t_r - s_r) \geq G + \rho(\beta(\alpha u(t_c - s) + (1 - \alpha)u(t_a)) + (1 - \beta)u(\underline{t})) + (1 - \rho)u(\bar{t}). \quad (10)$$

Inequality (10) ensures employees prefer to report the harmful act to the authority.

Again, we anticipate no employer would prefer the employee to commit the harmful act when preventing it is costless. The next lemma thus focuses on the case in which $G > \min[l(s_r), \rho\beta\alpha l(s)]$.

¹⁴In section 4, we discuss the case in which indemnifications are unfeasible, for instance, because they would attract bad publicity.

Lemma 4 *Take a given firm and suppose the employer tolerates the harmful act. Suppose also the employer encourages her employee to self-report.*

1. *If $l(s_r) \leq \rho\beta\alpha l(s)$, setting all transfers equal to zero is optimal.*
2. *If $l(s_r) > \rho\beta\alpha l(s)$, setting all transfers equal to zero is optimal, except for $t_r(s_r) > 0$, which is set so that the employee self-reports, i.e., $u(t_r - s_r) = \rho\beta\alpha u(-s)$.*

The employee's expected payoff is equal to $G - \min[l(s_r), \rho\beta\alpha l(s)]$.

When $l(s_r) \leq \rho\beta\alpha l(s)$, employees have private incentives to self-report, and setting all transfers equal to zero is therefore optimal. In case $l(s_r) > \rho\beta\alpha l(s)$, however, employees do not report to the authority unless their employers induce them to do so. To encourage a report, employers commit to indemnifying employees, that is, commit to a transfer $t_r(s_r) > 0$ set so that $u(t_r - s_r) = \rho\beta\alpha u(-s)$. Again, note the transfer $t_r(s_r)$ depends on s_r but not G .

We use the superscript “4” to refer to the transfers chosen by the employers who tolerate the harmful act and encourage their employees to self-report (e.g., $t_r^4(s_r)$).

Risk Neutrality. The employers who encourage their employees to self-report set all transfers equal to zero if $s_r < \rho\beta\alpha s$, and otherwise $t_r^4(s_r) = s_r - \rho\beta\alpha s$ and all other transfers equal to zero.

The next lemma states the condition under which an employer who tolerates the harmful act is better off encouraging her employee to report the act to the authority. In what follows, let s_0 denote the value of s_r such that $\rho\beta\alpha S = t_r^4(s_0)$. Because $t_r^4(s_r)$ is increasing in s_r , $\rho\beta\alpha S \geq t_r^4(s_r)$ for $\forall s_r \leq s_0$. Also, because $t_r^4(l^{-1}(\rho\beta\alpha l(s))) = 0$, $l(s_0) > \rho\beta\alpha l(s)$ necessarily.

Lemma 5 *Suppose $s_r \leq s_0$. An employer who tolerates the harmful act encourages her employee to self-report if and only if:*

$$S_r \leq \tilde{S}(\beta, s_r) = \begin{cases} \rho\beta\alpha S - t_r^4(s_r) & \text{if } l(s_0) \geq l(s_r) \geq \rho\beta\alpha l(s), \\ \rho\beta\alpha (S + t_c^3(s_r)) & \text{if } \rho\beta\alpha l(s) > l(s_r) \geq 0. \end{cases} \quad (11)$$

Now suppose $s \geq s_r > s_0$. An employer who tolerates the harmful act discourages her employee from self-reporting for $\forall S_r \geq 0$.

Proof. Suppose first $l(s_r) \geq \rho\beta\alpha(s)$. From Lemma 3, the expected payoff of an employer who tolerates the harmful act but discourages a report is equal to $\pi_1 - \rho\beta\alpha S$. From Lemma 4, the expected payoff of an employer who tolerates the harmful act and encourages her employee to self-report is equal to $\pi_1 - t_r^4(s_r) - S_r$. Comparing both expressions, we find the employer is better off encouraging a report if and only if $S_r \leq \tilde{S}_r(\beta, s_r) = \rho\beta\alpha S - t_r^4(s_r)$, where the right-hand side of the inequality is positive if and only if $s_r \leq s_0$. In case $s \geq s_r > s_0$, there does not exist a nonnegative value of S_r that induces the employer to encourage her employee to self-report. The case in which $l(s_r) < \rho\beta\alpha(s)$ is almost identical and thus left out. ■

The corporate sanction S_r must be sufficiently low (i.e., $S_r \leq \tilde{S}(\beta, s_r)$) for an employer to be better off designing an employment contract that induces her employee to self-report. Note the threshold $\tilde{S}(\beta, s_r)$ depends on the individual sanction s_r , because an employer who tolerates the harmful act takes into account not only the difference between $\rho\beta\alpha S$ and S_r when deciding whether to encourage a report, but also the difference in terms of expected transfers. In case s_r is very high (i.e., in case $s_r > s_0$), it is too expensive for an employer to encourage her employee to self-report even when $S_r = 0$. Finally, because $t_c^3(s_r)$ and $t_r^4(s_r)$ are independent of G , either all employers who tolerate the harmful act encourage their employees to self-report or none of them do.

When $u(x) = x$, $s_0 = \rho\beta\alpha(s + S)$ and (11) simplifies to $s_r + S_r \leq \rho\beta\alpha(s + S)$. Employer and employee coordinate on the jointly optimal reporting strategy when utility is perfectly transferable.

Finally, note that, in equilibrium, the expected payoff of an employee who commits the harmful act is equal to $G - \min[l(s_r), \rho\beta\alpha(s)]$, that is, his gain from the harmful act *under his preferred reporting strategy*. To see this, recall that the employers who tolerate misconduct only make transfers in case of disagreement regarding which reporting strategy to adopt, for instance, because $S_r > \tilde{S}(\beta, s_r)$ and $l(s_r) < \rho\beta\alpha(s)$. Moreover, in these cases, transfers are chosen so that employees are indifferent between their preferred reporting strategy and that of their employers.

The Level of Deterrence

Having computed the optimal employment contracts, we are in a position to determine the fraction of employers who choose to prevent the harmful act, for any given scheme (β, s_r, S_r) .

Lemma 6 *For any (β, s_r, S_r) , a threshold $\hat{G}(\beta, s_r, S_r) > \min[\rho\beta\alpha(s), l(s_r)]$ exists such that an*

employer prevents the harmful act if and only if her employee's gain G is lower than or equal to $\hat{G}(\beta, s_r, S_r)$. In equilibrium, a fraction $F(\hat{G}(\beta, s_r, S_r))$ of employers thus prevent the harmful act.

Only employers whose employees' gain G is not too large prevent misconduct. This follows from the fact that, for any (β, s_r, S_r) , (i) the expected transfer necessary to prevent misconduct is increasing in G , whereas (ii) the transfers made when tolerating it are independent of G . For all the employers whose employees' gain $G > \hat{G}(\beta, s_r, S_r)$, the benefits to preventing the harmful act are not large enough to justify the increase in wages.¹⁵ We refer to $\hat{G}(\beta, s_r, S_r)$ as the *deterrence threshold*, and to the employers who are payoff-indifferent between preventing and tolerating the harmful act as the *marginal employers*. We assume marginal employers prevent the harmful act.

The Authority's Problem

The authority specifies the individual sanction s_r and the corporate sanction S_r when designing the self-reporting scheme. The analysis to come highlights the distinct role of these sanctions. The sanction S_r enters directly the *employers'* payoff, and thus affects their willingness to prevent misconduct in a straightforward manner. By contrast, the sanction s_r affects the *employees'* incentives, and thus has only an indirect effect on employers.

A self-reporting scheme can be socially desirable in at least two ways. First, if employees who commit the harmful act can be made to self-report without weakening deterrence, the authority's cost of enforcement is lowered because these employees no longer need to be audited. Second, a self-reporting scheme can also be beneficial if it allows the authority to raise deterrence without increasing the frequency of its audits.

Saving on Enforcement Costs through the Corporate Sanction S_r

We analyze how S_r affects the fraction $F(\hat{G}(\beta, s_r, S_r))$ of employers who prevent misconduct.

Lemma 7 *For $\forall(\beta, s_r)$ such that $0 \leq \tilde{S}(\beta, s_r)$, the deterrence threshold $\hat{G}(\beta, s_r, S_r)$ is strictly increasing in S_r on $[0, \tilde{S}(\beta, s_r)]$ and independent of S_r for $\forall S_r \geq \tilde{S}(\beta, s_r)$.*

¹⁵Also, $\hat{G}(\beta, s_r, S_r) > \min[\rho\beta\alpha l(s), l(s_r)]$ because (i) preventing the harmful act is costless—and thus strictly optimal—for all the employers whose employees' gain $G \leq \min[\rho\beta\alpha l(s), l(s_r)]$ and (ii) some employers are willing to pay a positive expected transfer to prevent the harmful act and avoid the associated net loss $(2\rho - 1)\Delta\pi + \rho\beta\alpha S > 0$.

Proof. These results follow from Lemmas 5 and 6, and the fact that S_r negatively affects only the employers who tolerate the harmful act *and* encourage their employees to self-report. We disregard the possibility that $\tilde{S}(\beta, s_r) < 0$ (i.e., that $s > s_0$) because we show below that it cannot be optimal for the authority to induce such an outcome. ■

For any choice of (β, s_r) such that $0 \leq \tilde{S}(\beta, s_r)$, it is sufficient for the authority to impose a partially reduced sanction S_r equal to $\tilde{S}(\beta, s_r)$ to induce the highest possible level of deterrence. Increasing S_r beyond $\tilde{S}(\beta, s_r)$ does not increase deterrence, because employers who tolerate misconduct are either payoff-indifferent whether to encourage their employees to self-report (when $S_r = \tilde{S}(\beta, s_r)$) or are strictly better off discouraging their employees from self-reporting (when $S_r > \tilde{S}(\beta, s_r)$). By contrast, when $S_r < \tilde{S}(\beta, s_r)$, employers are strictly better-off encouraging their employees to self-report; increasing S_r thus hurts them, which raises deterrence.

Proposition 1 *Setting $S_r \leq \tilde{S}(\beta, s_r)$ is socially optimal. As a result, in equilibrium, all employees who commit the harmful act report their misconduct to the authority.*

We prove Proposition 1 in two steps. First, we establish the authority can replicate *at a lower cost* the level of deterrence achieved by any scheme (β, s_r, S_r) , where $S_r > \tilde{S}(\beta, s_r)$, with the scheme $(\beta, s_r, \tilde{S}(\beta, s_r))$. Second, we show this alternative scheme $(\beta, s_r, \tilde{S}(\beta, s_r))$ leaves the employers' and employees' expected payoffs unchanged, so that no detrimental *wealth effects* arise.

The reason deterrence can be preserved despite lower enforcement costs is twofold. First, when $S_r = \tilde{S}(\beta, s_r)$, the employers who tolerate the harmful act encourage their employees to self-report, which, all else being equal, reduces enforcement costs. Second, deterrence is unchanged because setting $S_r = \tilde{S}(\beta, s_r)$ provides as strong an incentive to prevent misconduct as setting $S_r > \tilde{S}(\beta, s_r)$.

Setting $S_r = \tilde{S}(\beta, s_r)$ instead of $S_r > \tilde{S}(\beta, s_r)$ affects the contracts designed by the employers who tolerate misconduct.¹⁶ With risk-averse employees, such changes could lead to ambiguous wealth effects. This, however, does not occur, because the policy change affects neither the employees' nor the employers' expected payoffs. Indeed, recall the expected payoff of the employees who misbehave is always equal to their net gain from the harmful act under *their* preferred reporting strategy, which is independent of S_r . Finally, by definition of $\tilde{S}(\beta, s_r)$, the expected payoff of the employers who tolerate the harmful act is identical when $S_r = \tilde{S}(\beta, s_r)$ and $S_r > \tilde{S}(\beta, s_r)$.

¹⁶Employers who prevent the harmful act do not adjust their contracts, because they are unaffected by S_r .

Because social welfare is strictly higher when $S_r = \tilde{S}(\beta, s_r)$ instead of $S_r > \tilde{S}(\beta, s_r)$, for $\forall(\beta, s_r)$, we know setting $S_r \in [0, \tilde{S}(\beta, s_r)]$ is optimal. We can also infer from this result that self-reporting schemes are strictly welfare-improving: by setting $s_r = s$ and choosing the same probability of audit absent the self-reporting scheme, social welfare is higher when $S_r = \tilde{S}(\beta, s)$. In fact, this argument understates the benefits of self-reporting schemes, because the authority would typically choose a different pair (β, s_r) , and may even set $S_r < \tilde{S}_r(\beta, s_r)$.

Without specifying $u(\cdot)$, characterizing the value S_r should take in $[0, \tilde{S}(\beta, s_r)]$ is not possible. Below, we show that, for the case of quasilinear utility functions, setting S_r as high as possible (i.e., $S_r = \tilde{S}(\beta, s_r)$) so as to lower the probability of audit β as much as possible is strictly optimal.

Increasing Deterrence through the Individual Sanction s_r

We now show that the self-reporting scheme also allows the authority to increase deterrence without having to raise the frequency of its audits β . To focus on deterrence, we ignore, for the moment, the potential welfare consequences of wealth effects.

Proposition 2 *Suppose $S_r = \tilde{S}(\beta, s_r)$ and fix any probability of audit β . The authority can induce a larger fraction $F(\hat{G}(\beta, s_r, S_r))$ of employers to prevent the harmful act by setting $l(s_r) < \rho\beta\alpha(s)$ instead of $l(s_r) \geq \rho\beta\alpha(s)$.*

In the Appendix, we prove Proposition 2 by showing that, when $S_r = \tilde{S}(\beta, s_r)$, setting $l(s_r)$ slightly below $\rho\beta\alpha(s)$ increases the fraction of employers who prevent the harmful act.

First, when $S_r = \tilde{S}(\beta, s_r)$, reducing s_r decreases the expected payoff of *all* the employers who tolerate the harmful act. The threshold $\tilde{S}(\beta, s_r)$ drives this effect. Recall $\tilde{S}(\beta, s_r)$ is the value of S_r that makes employers who tolerate the harmful act payoff-indifferent between encouraging and discouraging their employees to self-report. Recall also that (i) decreases in s_r decrease the expected payoff of the employers who discourage their employees from self-reporting and (ii) decreases in s_r —if S_r is held constant— increase the expected payoff of the employers who encourage their employees to self-report. As a result, if $S_r = \tilde{S}(\beta, s_r)$, that is, if the authority is to preserve the employers' indifference, the threshold sanction $\tilde{S}(\beta, s_r)$ must increase as s_r decreases.¹⁷ This

¹⁷This relation can be immediately seen for the case of quasilinear utility functions, where $\tilde{S}(\beta, s_r) = \rho\beta\alpha(s + S) - s_r$.

property of $\tilde{S}(\beta, s_r)$ implies that, in equilibrium, the employers who tolerate the harmful act are hurt by decreases in s_r even though they actually encourage their employees to self-report.

Second, setting $l(s_r) < \rho\beta\alpha l(s)$ decreases the payoff of *some* employers who prevent the harmful act, but these employers do not switch to letting it occur. To see this, consider the employers with the lowest incentives to prevent misbehavior when $l(s_r) \geq \rho\beta\alpha l(s)$; that is, consider the marginal employers. These employers need to make relatively large transfers to deter the harmful act, so that their employees—should they deviate and misbehave—would require a loss $l(s_r)$ significantly smaller than $\rho\beta\alpha l(s)$ to report their misbehavior and give up on these possible transfers. Therefore, when the authority sets $l(s_r)$ only slightly lower than $\rho\beta\alpha l(s)$, the decrease does not hurt these employers and they continue to prevent the harmful act. By contrast, the employers that are hurt are those whose employees' gain G is close to $\rho\alpha\beta l(s)$, that is, the employers with the strongest incentives to prevent misbehavior when $l(s_r) \geq \rho\beta\alpha l(s)$. When $l(s_r) \geq \rho\beta\alpha l(s)$, these employers need to make relatively small transfers (if any) to deter the harmful act. Therefore, their employees, if they were to deviate and misbehave, would report their misbehavior even if $l(s_r)$ is only slightly lower than $\rho\beta\alpha l(s)$. Setting $l(s_r) < \rho\beta\alpha l(s)$ thus increases these employees' incentives to commit the harmful act, and their employers have no choice but to raise transfers to continue deterring the misconduct. However, because $l(s_r)$ is set only slightly lower than $\rho\beta\alpha l(s)$, the increase in transfers is modest and these employers continue to prevent the harmful act.

To conclude, when $S_r = \tilde{S}(\beta, s_r)$, the authority is able to set $l(s_r) < \rho\beta\alpha l(s)$ in a way that does not hurt the employers who are only slightly better off preventing the harmful act when $l(s_r) \geq \rho\beta\alpha l(s)$, and hardly hurts those who are considerably better off preventing it anyway. Because setting $l(s_r) < \rho\beta\alpha l(s)$ lowers the payoff of all the employers who tolerate the harmful act, a larger fraction of employers must choose to prevent misconduct.

The ability to increase deterrence without increasing β suggests the authority—for any level of deterrence it wishes to achieve—may set $l(s_r) < \rho\beta\alpha l(s)$ so as to reduce β . Whether doing so is socially desirable depends on the function $u(\cdot)$. Indeed, changes in s_r affect the employment contracts and thus trigger ambiguous wealth effects. We next show that, for the case of quasilinear utility functions in which no wealth effects exist, setting $s_r = 0$ is optimal.

The Optimal Self-reporting Scheme under Risk Neutrality

When $u(x) = x$, transfers and sanctions drop out of the social welfare function and:

$$W(\beta, s_r, S_r) = \int_0^{\hat{G}(\beta, s_r, S_r)} (\pi_0 - (1 - \rho)(h + \beta k)) dF(G) + \int_{\hat{G}(\beta, s_r, S_r)}^{\infty} (\pi_1 + G - \rho h) dF(G). \quad (12)$$

As a result, the authority's only concern is to choose the pair of sanctions (s_r, S_r) that minimizes the probability of audit β necessary to achieve its desired level of deterrence $\hat{G}(\beta, s_r, S_r)$. To help intuition, note that we show in the appendix that the level of deterrence is increasing in β .

Proposition 1 established that setting $S_r \in [0, \tilde{S}(\beta, s_r)]$ is optimal because it ensures employers who tolerate misconduct encourage their employees to self-report, which reduces enforcement costs. If $S_r < \tilde{S}(\beta, s_r)$, $\forall s_r$ and $\forall \beta$, increasing S_r reduces the expected payoff of the employers who tolerate misconduct, which, in turn, raises deterrence. As a result, for any level of deterrence it wishes to achieve, the authority finds setting $S_r = \tilde{S}(\beta, s_r) = \rho\beta\alpha(s + S) - s_r$ so as to minimize β is optimal.

Proposition 2 established that, all else equal, a higher level of deterrence can always be achieved by setting $s_r < \rho\beta\alpha s$ instead of $s_r \geq \rho\beta\alpha s$. By a logic similar to that in the previous paragraph, it follows that setting $s_r < \rho\beta\alpha s$ to lower β is optimal. In fact, we find granting *full amnesty* achieves the highest level of deterrence possible, $\forall \beta$, which allows the authority to select the lowest probability β compatible with its desired level of deterrence.

To see this, recall that, when $s_r \leq \rho\beta\alpha s$, decreasing the individual sanction s_r produces several effects. First, it increases the corporate sanction $S_r = \rho\beta\alpha(s + S) - s_r$, which decreases the expected payoff of all the employers who tolerate misconduct. Indeed, their expected payoff is equal to:

$$\pi_1 - \tilde{S}(\beta, s_r) = \pi_1 + s_r - \rho\beta\alpha(s + S). \quad (13)$$

Second, lowering the sanction s_r hurts some but not all the employers who prevent the harmful act. The employers who are hurt by a decrease in s_r are those who make relatively low transfers (if any) when $s_r \geq \rho\beta\alpha s$. As s_r decreases and the harmful act becomes more tempting to employees, these low transfers are no longer sufficient to deter misbehavior, and employers have no choice but to raise wages. Formally, for any s_r , the employers whose expected payoff has decreased as a result

of setting $s_r < \rho\beta\alpha s$ are those whose employees' gain $G \in [s_r, \tilde{G}(\beta, s_r)]$, where:

$$\tilde{G}(\beta, s_r) = \frac{1}{1-\rho} (\rho^2\beta\alpha s - (2\rho-1)s_r). \quad (14)$$

Their expected payoff is equal to $\pi_0 - (G - s_r)$. Not surprisingly, the range $[s_r, \tilde{G}(\beta, s_r)]$ of employers who must raise their transfers expands as s_r decreases. By contrast, the employers whose expected payoff is unchanged are those who make relatively large transfers when $s_r \geq \rho\beta\alpha s$ (i.e., those whose employees' $G \geq \tilde{G}(\beta, s_r)$). Their expected payoff is equal to:

$$\pi_0 - \frac{\rho}{2\rho-1} (G - \rho\beta\alpha s). \quad (15)$$

When $s_r = \rho\beta\alpha s$, the range $[s_r, \tilde{G}(\beta, s_r)]$ shrinks to $\rho\beta\alpha s$ and the marginal employers are found by equating (13) to (15), which yields:

$$\begin{aligned} \hat{G}(\beta, s_r) &= \rho\beta\alpha s + \frac{(2\rho-1)}{\rho} ((2\rho-1)\Delta\pi + \rho\beta\alpha(s+S) - s_r) \\ &= \rho\beta\alpha s + \frac{(2\rho-1)}{\rho} ((2\rho-1)\Delta\pi + \rho\beta\alpha S). \end{aligned} \quad (16)$$

All else equal, a small decrease in s_r (i) raises deterrence (i.e., $\hat{G}(\beta, s_r)$) and (ii) increases the range $[s_r, \tilde{G}(\beta, s_r)]$ of employers for whom preventing the harmful act becomes more costly. The level of deterrence is increasing in the generosity of the amnesty as long as the cost to the marginal employers of preventing the harmful act is unchanged, that is, as long as $\hat{G}(\beta, s_r) > \tilde{G}(\beta, s_r)$. If the cost to the marginal employers of preventing the harmful act is unaffected when full amnesty is granted—formally, if $\hat{G}(\beta, 0) > \tilde{G}(\beta, 0)$ —then setting $s_r = 0$ clearly achieves the highest possible level of deterrence, which, in turn, allows the authority to select the lowest audit probability β compatible with its desired level of deterrence.

If instead the cost to the marginal employers of preventing the harmful act increases once the reduction in the individual sanction exceeds some threshold—that is, if the value of s_r such that $\hat{G}(\beta, s_r) = \tilde{G}(\beta, s_r)$ is positive—further reducing s_r leaves deterrence unchanged. This lack of effect on deterrence occurs because, past some threshold, further decreasing s_r *equally* hurts the employers who prevent the harmful act and those who tolerate it. To see this, set (13) equal to

$\pi_0 - (G - s_r)$, and rearrange to derive the deterrence threshold when $\hat{G}(\beta, s_r) \leq \tilde{G}(\beta, s_r)$:

$$\hat{G}(\beta, s_r) = (2\rho - 1) \Delta\pi + \rho\beta\alpha(s + S).$$

When the self-reporting scheme affects the marginal employers, an interval of values s_r can take, *which includes* $s_r = 0$, exists that ensures the highest level of deterrence is achieved. Again, therefore, granting full amnesty is optimal because it allows the authority to choose the lowest possible probability of audit compatible with its desired level of deterrence.

Proposition 3 *When $u(x) = x$, it is socially optimal (i) to grant full amnesty to the employees who self-report (i.e., $s_r = 0$) and (ii) to partially reduce the sanction imposed on employers whose employees self-report (i.e., $S_r = \rho\beta\alpha(s + S)$).*

When the authority does not take into account the wealth effects of its decisions, it grants full amnesty to the employees who report their own behavior. Although this policy exacerbates the firms' agency problems when they prevent the harmful act, the net effect is desirable because it allows the authority to incur the lowest possible enforcement costs compatible with any given level of deterrence it wishes to achieve. Further, even though setting $s_r < \rho\beta\alpha s$ is strictly optimal, setting $s_r = 0$ can in some cases only be *weakly* optimal. Determining the interval of values that s_r can take would require specifying the function $F(\cdot)$. Finally, the last step in the authority's problem is conceptually simple: it chooses the probability β that maximizes (12) evaluated at $s_r = 0$ and $S_r = \rho\beta\alpha(s + S)$. We omit this last step for brevity's sake, and emphasize that our results regarding the optimal sanctions hold for any level of deterrence if wishes to achieve.

Remark. The model we solved did not allow for the employees to bring evidence of their misconduct to their employers, and for the latter to subsequently report it to the authority. In a setting in which employers can report evidence acquired from their employees, the sanctions imposed in case of a corporate report would be identical to those imposed in case of an individual report. This equivalence exists because, in equilibrium, the employees who self-report are encouraged by their employers to do so. If nothing changes when employers can report the evidence to the authority themselves, one may wonder whether referring to our reporting scheme as an *employee*

self-reporting scheme is appropriate. However, the analysis to come shows that a clear distinction between corporate and individual reports is warranted when employers monitor their employees.

4 Extensions

We first extend the baseline model to analyze corporate reporting schemes, and then discuss some of our modelling assumptions. For the sake of brevity, we suppose $u(x) = x$ throughout.

Corporate Reporting Schemes

Suppose employers are endowed with a monitoring technology that produces evidence of misconduct with probability η in case an employee committed the harmful act, and otherwise produces no evidence.¹⁸ In addition, suppose whether an employer comes into possession of evidence is unobservable to both her employee and the authority. Therefore, for an employer to credibly commit to reporting evidence, reporting must be rational when the possibility presents itself. We introduce this feature to capture a concern many legal scholars raise regarding firms' incentives to report employee misbehavior, namely, that they may be reluctant to do so when corporate sanctions are too high (see, e.g., Arlen (2012)). Finally, for simplicity, we assume employees who wish to self-report are always able to do so *before* their employers can report misconduct to the authority.

The authority now chooses the policy $(\beta, (s_r, S_r), (s_R, S_R))$. The sanctions (s_r, S_r) are imposed when employees report their own conduct, whereas (s_R, S_R) are the sanctions imposed when employers report their employees' misconduct. To avoid unrealistic predictions, we assume $s_R \leq s$. As before, because $u(x) = x$, the authority chooses the sanctions $((s_r, S_r), (s_R, S_R))$ that minimize the probability of audit β necessary to achieve its desired level of deterrence. Finally, employment contracts now specify an extra transfer t_R paid when an employer reports evidence.

To save on space, we only report the weakly optimal choice of sanctions and transfers.

¹⁸Formally, the monitoring technology produces no evidence with probability $(1 - \eta)$ if $e = 1$, and with probability 1 if $e = 0$.

Employment Contracts

We begin by computing the optimal contracts. To do so, we anticipate that it is optimal to (i) set S_R low enough that all employers who prevent misconduct are able to costlessly commit to reporting evidence, and (ii) set $s_R \geq \rho\beta\alpha s$. As a result, the expected sanction employees face when they remain silent is equal to $\eta s_R + (1 - \eta)\rho\beta\alpha s > \rho\beta\alpha s$. In what follows, let $\bar{\eta} = \eta + (1 - \eta)\rho\beta\alpha$.

Preventing the harmful act. Suppose the employers who prevent misconduct are able to credibly and costlessly commit to reporting evidence, and that it is in their interest to do so. The problem these employers solve is identical to that stated in Section 3, except (i) the expected sanction $\eta s_R + (1 - \eta)\rho\beta\alpha s$ replaces $\rho\beta\alpha s$ and (ii) the expected transfer received by an employee who commits the harmful act and remains silent is equal to $\eta t_R + (1 - \eta)(\rho(\beta(\alpha t_c + (1 - \alpha)t_a) + (1 - \beta)\underline{t}) + (1 - \rho)\bar{t})$. If $s_r \geq \eta s_R + (1 - \eta)\rho\beta\alpha s$, the employers set all transfers equal to zero, except possibly for $\bar{t} = \max\left[\frac{G - \eta s_R - (1 - \eta)\rho\beta\alpha s}{\rho - (1 - \rho)(1 - \eta)}, 0\right]$. If instead $s_r < \eta s_R + (1 - \eta)\rho\beta\alpha s$, employers set $\bar{t} = \frac{G - \eta s_R - (1 - \eta)\rho\beta\alpha s}{\rho - (1 - \rho)(1 - \eta)}$ and all other transfers equal to zero if their employees' gain G is higher than or equal to:

$$\tilde{G}(\beta, s_r, s_R) = \frac{(\rho(\eta s_R + (1 - \eta)\rho\beta\alpha s) - (\rho - (1 - \rho)(1 - \eta))s_r)}{(1 - \rho)(1 - \eta)}. \quad (17)$$

Condition (17) is equivalent to condition (6) in the model without corporate reporting schemes. The employers whose employees' $G \leq \tilde{G}(\beta, s_r, s_R)$ set $\bar{t} = \max\left[\frac{G - s_r}{\rho}, 0\right]$, and all other transfers equal to zero. Note the problem the employers face is "relaxed" when s_R increases. Below, we rely on this observation to show that setting $s_R = s$ is optimal.

Tolerating the harmful act. Employers are tempted to tolerate the harmful act if and only if their employees' gain $G > \min[\eta s_R + (1 - \eta)\rho\beta\alpha s, s_r]$.¹⁹ Moreover, employers who tolerate misconduct must decide (i) whether to encourage their employees to self-report and (ii) whether to commit to reporting evidence. Observe that choosing a low S_R may increase the expected payoff of the employers who tolerate the harmful act *and* report evidence to the authority.

Suppose first the employers who choose to tolerate misconduct discourage their employees from self-reporting. Suppose also these employers commit *not* to report evidence to the authority.²⁰

¹⁹When $G \leq \min[\eta s_R + (1 - \eta)\rho\beta\alpha s, s_r]$, preventing the harmful act is costless and, therefore, optimal.

²⁰An employer can commit not to report evidence by, for instance, setting t_R large enough. This can be

Exactly as in section 3, they set all transfers equal to zero if $s_r \geq \rho\beta\alpha s$, and otherwise $t_c = s - \frac{s_r}{\rho\beta\alpha}$ and all other transfers equal to zero. Their expected payoff is equal to:

$$\Pi_1 = \begin{cases} \pi_1 - \rho\beta\alpha S & \text{if } s_r \geq \rho\beta\alpha s, \\ \pi_1 + s_r - \rho\beta\alpha (s + S) & \text{if } s_r < \rho\beta\alpha s. \end{cases} \quad (18)$$

If instead these employers commit to reporting the evidence their monitoring technology produces, they set all transfers equal to zero if $s_r \geq \eta s_R + (1 - \eta) \rho\beta\alpha s$, and otherwise $t_R = t_c = \frac{\eta s_R + (1 - \eta) \rho\beta\alpha s - s_r}{\eta + (1 - \eta) \rho\beta\alpha}$ and all other transfers equal to zero. Below, we show employers never opt for this strategy.

Suppose now the employers who tolerate misconduct encourage their employees to self-report. If they also commit to reporting evidence, they set all transfers equal to zero if $s_r \leq \eta s_R + (1 - \eta) \rho\beta\alpha s$, and otherwise $t_r = s_r - (\eta s_R + (1 - \eta) \rho\beta\alpha s)$ and all other transfers equal to zero. If instead they commit not to report evidence, they set all transfers equal to zero if $s_r \leq \rho\beta\alpha s$, and otherwise $t_r = s_r - \rho\beta\alpha s$ and all other transfers equal to zero. Because these employers are imposed S_r regardless of their own reporting strategy, they select the strategy that minimizes the expected transfer they must pay their employees. A comparison of expected transfers establishes that employers who encourage their employees to self-report prefer to report evidence to the authority.

Choosing the Sanction S_R

Recall that whether an employer comes into possession of evidence is unobservable, so that reporting must be rational when the opportunity presents itself for the commitment to be credible. Consider an employer who designed a contract meant to prevent the harmful act. If in possession of evidence of misbehavior,²¹ this employer reports the evidence if and only if the sanction S_R is lower than or equal to (i) the expected sanction she faces when remaining silent plus (ii) the expected transfer she must pay her employee when pretending not to be in possession of evidence, that is, if and only if:²²

$$S_R \leq (1 - \rho) \max \left[\frac{(G - (\eta s_R + (1 - \eta) \rho\beta\alpha s))}{\rho - (1 - \rho)(1 - \eta)}, 0 \right] + \rho\beta\alpha S. \quad (19)$$

interpreted as the employer not adopting the monitoring technology.

²¹Note this scenario only occurs out of equilibrium. Nevertheless, if in possession of evidence, the employer knows with certainty her employee committed the harmful act.

²²Recall an employee has already given up on the possibility of self-reporting by this stage.

This condition is most stringent for the employers whose employees' $G \leq \eta s_R + (1 - \eta) \rho \beta \alpha s$. It follows setting $S_R \leq \rho \beta \alpha s$ ensures *all* employers who prevent misconduct are able to costlessly commit to reporting evidence. The commitment is costless in that S_R is low enough that the employers do not need to raise their transfers to ensure (19) holds.²³ Suppose then $S_R = \rho \beta \alpha S$.

We anticipated that the employers who tolerate misconduct and discourage their employees from self-reporting prefer not to report evidence. When $S_R = \rho \beta \alpha S$, the expected sanction these employers face is $\rho \beta \alpha S$ regardless of their reporting strategy, so that they choose the strategy that minimizes their expected transfer. Using the contracts identified above, one finds that the employers who discourage their employees from self-reporting prefer not to report evidence.

Choosing the Sanction S_r

An employer who tolerates the harmful act encourages her employee to self-report if and only if the sanction S_r is lower than or equal to some threshold $\tilde{S}(\beta, s_r, s_R, S_R)$, where:

$$\tilde{S}(\beta, s_r, s_R, S_R) = \begin{cases} \rho \beta \alpha S + \eta s_R + (1 - \eta) \rho \beta \alpha s - s_r & \text{if } \eta s_R + (1 - \eta) \rho \beta \alpha s \leq s_r, \\ \rho \beta \alpha S & \text{if } \rho \beta \alpha s \leq s_r \leq \eta s_R + (1 - \eta) \rho \beta \alpha s, \\ \rho \beta \alpha (S + s) - s_r & \text{if } s_r \leq \rho \beta \alpha s. \end{cases} \quad (20)$$

The formula for \tilde{S} is found using the optimal contracts outlined above, substituting $S_R = \rho \beta \alpha S$ into the employers' expected payoffs.²⁴ When $S_r = \tilde{S}$, employers who tolerate misconduct encourage their employees to self-report, and their expected payoff is given by expression (18).

Deterrence. For $\forall(\beta, s_r, S_r, s_R, S_R)$, the expected transfer paid by the employers who prevent misconduct is increasing in G . By contrast, the transfers paid by the employers who tolerate misconduct are independent of G . Therefore, as in section 3, a threshold $\hat{G}(\beta, (s_r, S_r), (s_R, S_R)) > \min[\eta s_R + (1 - \eta) \rho \beta \alpha s, s_r]$ exists such that an employer prevents misconduct if and only if $G \leq \hat{G}$. This threshold \hat{G} is strictly increasing in S_r when $S_r \in [0, \tilde{S}(\beta, s_r, s_R, S_R)]$, and otherwise independent of S_r . It follows setting $S_r = \tilde{S}(\beta, s_r, s_R, S_R)$ is optimal: fixing any (β, s_r, s_R, S_R) , setting $S_r = \tilde{S}$ allows the authority to reach as high a level of deterrence as when $S_r > \tilde{S}$, but at a

²³If S_R were higher than $\rho \beta \alpha S$, some employers may have to increase transfers to make it rational for themselves to report evidence. This, in turn, would make it more costly to prevent the harmful act.

²⁴Condition (20) implicitly assumes s_r is low enough that \tilde{S} is positive. Exactly for the same reasons as those in the baseline model, setting s_r low enough that \tilde{S} is positive is socially optimal (see proof of Proposition 1).

lower cost because employees who commit the harmful act are no longer audited.

The next proposition states this result as well as the optimal choice of (s_R, S_R) .

Proposition 4 *Setting $S_r = \tilde{S}(\beta, s_r, s_R, S_R)$ is socially optimal. In equilibrium, therefore, all employees who commit the harmful act report their misconduct to the authority.*

Moreover, setting $S_R = \rho\beta\alpha S$ and $s_R = s$ is socially optimal. The authority imposes a partially reduced corporate sanction on employers who report employee misconduct, and imposes the full individual sanction on these employers' employees.

Proof. Consider first the employers who tolerate the harmful act. When $S_R \leq \rho\beta\alpha S$, their expected payoff is independent of the sanctions s_R and S_R (see equation (18)), so that the authority—by setting $S_R = \rho\beta\alpha S$ and $s_R = s$ —ensures employers who tolerate misconduct enjoy the lowest possible expected payoff for any given (β, s_r) .²⁵ By contrast, setting $S_r < \rho\beta\alpha S$ weakly increases their expected payoff because the strategy consisting of “discouraging employee reports and reporting evidence” may then dominate the strategy consisting of “discouraging employee reports and not reporting evidence.”

Consider now the employers who prevent the harmful act. Setting $s_R = s$ ensures they enjoy the highest possible expected payoff because, for $\forall (\beta, s_r, S_R)$, their expected payoff is weakly increasing in s_R . Moreover, by setting $S_R = \rho\beta\alpha S$, the authority enables *all* the employers who wish to prevent the harmful act to *costlessly* commit to reporting evidence. This follows from the fact that inequality (19) is slack for $\forall G$, and implies that the employers who prevent misconduct enjoy the highest possible expected payoff for any given (β, s_r) . Indeed, recall the employers are weakly better off when able to costlessly commit to reporting evidence because the resulting expected sanction employees face when committing the harmful act and remaining silent is higher (i.e., it is equal to $\bar{\eta}s > \rho\beta\alpha s$). Choosing a higher S_r would increase the cost to some employers of preventing the harmful act, which can either result in these employers no longer committing to report evidence or, worse, result in these employers preferring to tolerate the harmful act (see (19)). Finally, choosing a lower S_R would not decrease any further the cost of preventing the harmful act to the employers.

It follows setting $S_R = \rho\beta\alpha S$ and $s_R = s$ ensures the highest possible level of deterrence for any

²⁵In other words, the employers who tolerate misconduct would enjoy the same expected payoff if $s_R = +\infty$ and $S_R = +\infty$.

given (β, s_r) , which is optimal because it allows the authority to choose the lowest probability of audit β compatible with its desired level of deterrence. ■

The authority finds it optimal to implement both a corporate and an employee reporting scheme. These two reporting schemes serve distinct purposes, and target different employers. On the one hand, the corporate reporting scheme helps the employers who wish to prevent the harmful act. It makes it rational for them to report evidence they have uncovered through their monitoring activities. This policy is optimal because, when combined with the imposition of a severe sanction on employees, it raises the expected sanction employees face when misbehaving. On the other hand, the employee self-reporting scheme ensures that the employers who tolerate the harmful act encourage their employees to report their behavior to the authority. This policy is desirable because it lowers the authority’s enforcement costs without weakening deterrence.

For the sake of brevity, we stop the analysis at this point, without formally investigating the choice of s_r . However, we wish to make the observation that, unlike in the baseline model, setting $s_r < \bar{\eta}s$, that is, setting the sanction imposed on employees who self-report strictly lower than the expected sanction they face when remaining silent, may or may not allow the authority to increase deterrence. In short, to hurt the employers who tolerate misconduct, the authority needs to set $s_r < \rho\beta\alpha s$ (see equation (18)). However, the cost of preventing the harmful act increases for some employers as soon as $s_r < \bar{\eta}s$, where $\bar{\eta}s > \rho\beta\alpha s$. Intuitively, because the corporate reporting scheme increases the expected individual sanction, it takes a higher value of s_r to make it tempting for employees to commit the harmful act and report it to the authority. Nevertheless, one can show that the deterrence property of employee self-reporting schemes remains, for instance, if employers suffer sufficiently from the harmful act (i.e., if $\Delta\pi$ is large enough).

Remark. Our model assumes employees cannot transmit evidence to their employers. If they could, one may be concerned that employers who tolerate the harmful act “abuse” the corporate scheme by obtaining evidence from their employees so as to subsequently report it themselves. If this was the case, the desirability of the schemes outlined above may be undermined.²⁶ However, when solving the version of the model in which employees can internally self-report, one finds that

²⁶By contrast, one should not be concerned about employers who prevent the harmful act abusing the *employee* self-reporting scheme. In a nutshell, this is because it is optimal for them to commit to the reporting strategy with the highest associated expected *individual* sanction. Because $s_R = s$, the expected individual sanction under the corporate scheme can only be weakly higher than that under the employee scheme.

an employer who tolerates the harmful act *and* abuses the corporate scheme pays her employee an expected transfer $s_R - \min[s_r, \bar{\eta}s]$. This expected transfer makes the employee indifferent between abusing the corporate scheme and his preferred reporting strategy. As a result, the expected payoff of an employer who abuses the corporate scheme is $\pi_1 - S_R - (s_R - \min[s_r, \bar{\eta}s]) = \pi_1 - \rho\beta\alpha S - (s - \min[s_r, \bar{\eta}s])$, which is strictly lower than her payoff when she encourages her employee to self-report (18). Employers who tolerate the harmful act do not abuse the corporate self-reporting scheme because it is too costly for them to compensate their employees for the sanction s_R .

Employee Indemnifications

We assumed employers could offer contracts specifying transfers contingent on judicial decisions and employee reports. We now discuss—focusing on the version of the model without corporate reports—whether the ability for employee self-reporting schemes to raise deterrence without increasing enforcement costs (Proposition 2) remains valid when introducing forms of *contractual incompleteness* meant to capture employers’ difficulty in specifying payments contingent on judicial decisions. In what follows, we take it for granted that inducing employers who tolerate the harmful act to encourage their employees to self-report is optimal (Proposition 1). We choose to comment on the robustness of the former finding because it is, we believe, the most novel result of our paper and the one most specific to corporate crimes (as opposed to “ordinary” crimes).²⁷

We solved the model assuming employers could set $t_c, t_r > 0$, and this feature was exploited by the employers who tolerate the harmful act to induce their preferred reporting strategy. In some instances, however, employers’ ability to *indemnify* employees may be limited, for instance, because it is illegal or because it would attract bad publicity. One could model various kinds of contractual incompleteness to take these limitations into account. Suppose, for instance, that $t_c = t_r = 0$ by rule. This captures a situation in which not only indemnifications are unfeasible, but also employees, when convicted by the authority, lose any positive wage they may have received (either because they must transfer it back to their employers or because the authority appropriates it as part of the sanction). The employers’ problem when preventing the harmful act is unaffected,

²⁷The finding whereby the authority is able to reduce enforcement costs without sacrificing deterrence by inducing employers who tolerate the harmful act to encourage employee reports is very general. Proofs of its robustness to various forms of contractual incompleteness are available upon request.

because setting $t_c = t_r = 0$ is optimal anyhow. Further, because setting $S_r = \tilde{S}(\beta, s_r)$ remains optimal, we only need to analyze the expected payoff of the employers who tolerate the harmful act *and* discourage their employees from self-reporting to determine the level of deterrence. In Lemma 3, we stated that the optimal contract chosen by the employers who tolerate misconduct and discourage their employees from self-reporting consisted of setting $t_c = s - \frac{s_r}{\rho\beta\alpha}$ and all other transfers to zero. However, these employers are actually indifferent between this contract and an alternative one that specifies $\bar{t} = \underline{t} = t_a = \frac{\rho\alpha\beta s - s_r}{(1-\rho\beta\alpha)}$ and $t_c = t_r = 0$. As a result, setting $t_c = t_r = 0$ by rule leaves unaffected the expected payoff of the employers, regardless of the action they induce. It then follows that our findings hold in this modified setup as well.²⁸

Corporate Sanctions and Employment Contracts

We solved our model assuming the authority did not make corporate sanctions contingent on the nature of the employment contracts. Instead, we assumed sanctions—in case of an audit—were imposed with probability α in case an employee committed the harmful act, and otherwise never imposed. As a result, we did not need to specify whether employment contracts were observable to the authority. By contrast, suppose the authority, upon auditing a firm, observes both the contract that links the employer to the employee and the employee’s gain G . Further, for simplicity, maintain the assumption whereby “innocent” employees are never sanctioned and “guilty” ones are sanctioned with probability α .²⁹ Because it knows G , the authority is then perfectly able to distinguish between a contract that prevents the harmful act and one that tolerates it. In such a setting, we conjecture that the authority, for fear of dampening the employers’ incentives to prevent misconduct, would never sanction an employer who designed a contract meant to prevent the harmful act. By a similar logic, if it could, the authority would systematically sanction employers who designed a contract meant to tolerate the harmful act (i.e., with probability $1 > \alpha$).

²⁸When employees are risk averse, setting $t_c > 0$ is *strictly* optimal for the employers who tolerate the harmful act. Other contracts exist that induce their preferred reporting strategy, but not being able to indemnify employees reduces employers’ expected payoff. Nevertheless, our results continue to hold because the salary cost to all employers who tolerate the harmful act becomes positive as soon as $s_r < \rho\beta\alpha s$.

²⁹Intriguingly, a form of “infinite regress” problem arises if one assumes employee sanctions are also contingent on the nature of employment contracts.

5 Policy Discussion

Corporate reporting schemes abound (see, e.g., the EPA’s *Audit Policy*, the DOD’s *Contractor Disclosure Program*, and the FERC’s *Self-reporting Scheme*). As in our model, they specify a partial reduction in the corporate sanction imposed on an applicant firm.³⁰ To deter price-fixing, in both the United States and Europe, the *corporate leniency program* offers the first firm filing for leniency full immunity.³¹ Although our model suggests *partial* immunity is sufficient, the difference is likely due to the fact that our model does not focus on the specificities of cartels.

To the best of our knowledge, however, these corporate reporting schemes allow for all employees within the firm to be granted full amnesty when the firm reports misconduct, including those who physically committed the harmful act.³² Hammond (2004) argues, in the context of price-fixing, that such a “blanket” covering the entire corporation and its employees incentivizes employees to report illegal acts to their superiors in order to file for the corporate leniency program together. This approach is in contrast to our results regarding corporate schemes. Although we find reducing the sanction imposed on the firm is optimal, we also derive that the individuals physically involved in the wrongdoing should be punished. The policy we suggest is desirable because it increases the deterrent effect of individual sanctions. Interestingly, the Department of Justice, in its fight against foreign bribery, has recently emphasized the need for corporations to turn over evidence against their own employees in order to qualify for a reduced corporate sanction.³³

Furthermore, in our view, a striking feature of current practice is the almost complete absence of employee self-reporting schemes. Although the Sarbanes-Oxley Act did foster the protection of individual informants (i.e., whistle blowers), it did not explicitly address the issue of individuals reporting their own misconduct. In addition, even though the United States have an individual leniency program for antitrust violations, this program is rarely used, perhaps because of the argument put forward by Hammond; because corporate leniency applies to everyone within the firm, the individual, eager to retain his/her job, prefers not to engage in actions leading to sanctions

³⁰For a description of the DOD’s *Contractor Disclosure Program*, go to www.dodig.mil/Inspections/IPO/ContractorDisclosure/Contractor%20Program%20Guide%20030509. For a description of the FERC’s policy, go to www.ferc.org/whats-new/commeet/2008/051508/M-1.pdf.

³¹See *Commission Notice on Immunity from fines and reduction of fines in cartel cases*, Official Journal C298/17 (2006); and the U.S. Department of Justice’s *Corporate Leniency Policy* (August 1993).

³²Some corporate reporting schemes do not explicitly offer amnesty to the employees. However, neither do they provide a systematic approach to dealing with the issue of individual sanctions.

³³See <http://www.ft.com/intl/cms/s/0/cdb01524-cb8b-11e5-be0b-b7ece4e953a0.htmlaxzz40HPvoj84>.

being imposed on his/her firm, and instead file for leniency through the firm. By contrast, our model suggests that employee self-reporting schemes can be an important instrument in the fight against corporate crime (even when corporate schemes are already in place). They specifically target the individuals most likely to possess evidence of misconduct—i.e., the perpetrators—and allow the authority to increase deterrence and reduce its enforcement costs.

6 Concluding Remarks

We conclude by discussing possible avenues for future research on self-reporting schemes and corporate crime. During our analysis, we assumed that firms suffered from the breach of the law. The employers who tolerated the misconduct did so only because it was too costly to prevent it. Whereas this captures numerous situations, and in fact is the leading assumption in the literature, it is also plausible that firms sometimes benefit from violations of the law (e.g., price-fixing). A natural path for future research is therefore to analyze the optimal design of self-reporting schemes when firms may take steps to encourage their employees to breach the law.

Further, our analysis does not apply directly to the case of cartels because it ignores the horizontal interactions between firms. However, antitrust agencies being among the most keen to use self-reporting schemes, an extension of our analysis that captures the dynamics of cartels may prove valuable and help agencies adjust their current policies.

Appendix

A Proofs of Lemma 1 and Lemma 2

Observe that setting $\bar{t}, \underline{t}, t_a \geq 0$ guarantees a nonnegative expected payoff to the employee. Because concave functions appear on both sides of (3) and (4), we operate the following change of variables. We specify $\bar{u} = u(\bar{t})$, $\underline{u} = u(\underline{t})$, $u_c = u(t_c - s)$, $u_r = u(t_r - s_r)$, and $u_a = u(t_a)$, and let $h(\bar{u}) = u^{-1}(\bar{u})$, $h(\underline{u}) = u^{-1}(\underline{u})$, and so on. The employer chooses the *ex post* payoffs $\bar{u} \geq 0$, $\underline{u} \geq 0$, $u_c \geq u(-s)$, $u_r \geq u(-s_r)$, and $u_a \geq 0$, to maximize:

$$\rho \bar{\pi} + (1 - \rho) \underline{\pi} - \rho h(\bar{u}) - (1 - \rho) (\beta h(u_a) + (1 - \beta) h(\underline{u})), \text{ subject to:} \quad (21)$$

$$\rho \bar{u} + (1 - \rho) (\beta u_a + (1 - \beta) \underline{u}) \geq G + \rho (\beta (\alpha u_c + (1 - \alpha) u_a) + (1 - \beta) \underline{u}) + (1 - \rho) \bar{u}, \quad (22)$$

$$\rho \bar{u} + (1 - \rho) (\beta u_a + (1 - \beta) \underline{u}) \geq G + u_r. \quad (23)$$

If $G - \min[l(s_r), \rho\beta\alpha l(s)] \leq 0$, setting $\bar{u} = u_a = \underline{u} = 0$, $u_c \in [u(-s), -\frac{G}{\rho\beta\alpha}]$, and $u_r \in [u(-s_r), -G]$ is optimal. Suppose then $G - \min[l(s_r), \rho\beta\alpha l(s)] > 0$. Suppose further inequality (22) is the only binding constraint, and let λ denote its multiplier. Setting $u_c = u(-s)$ and $\underline{u} = 0$ to relax (22) is optimal. Ignoring u_r for now, if \bar{u} and u_a are positive at the optimum, the system of FOCs is:

$$h'(\bar{u}) = \frac{1}{u'(\bar{t})} = \lambda \frac{(2\rho - 1)}{\rho} \quad (24)$$

$$h'(u_a) = \frac{1}{u'(t_a)} = \lambda \frac{((1 - \rho) - \rho(1 - \alpha))}{(1 - \rho)}. \quad (25)$$

Equation (25) can be satisfied if and only if $\alpha > \tilde{\alpha} = \frac{2\rho - 1}{\rho}$. The relevant system of FOCs is thus composed of either (24) and (25) if $\alpha > \tilde{\alpha}$, or (24) if $\alpha \leq \tilde{\alpha}$. The solution when (22) is the only binding constraint is thus such that $\bar{t}, t_a > \underline{t} = t_c = 0$ if $\alpha > \tilde{\alpha}$, and otherwise such that $\bar{t} > t_a = \underline{t} = t_c = 0$.

Further, satisfying (22) implies (23) as long as:

$$\rho (\beta (\alpha u(-s) + (1 - \alpha) u(t_a))) + (1 - \rho) u(\bar{t}) \geq u(t_r - s_r), \text{ if } \alpha > \tilde{\alpha}, \text{ and} \quad (26)$$

$$\rho\beta\alpha u(-s) + (1 - \rho) u(\bar{t}) \geq u(t_r - s_r) \text{ if } \alpha \leq \tilde{\alpha}. \quad (27)$$

A sufficient condition for either inequality to hold is $u(-s_r) \leq \rho\beta\alpha u(-s)$, in which case setting $t_r \in [0, \tilde{t}_r]$ is optimal, where \tilde{t}_r is the highest value of t_r such that the relevant inequality (either (26) or (27)) holds. Suppose instead $u(-s_r) > \rho\beta\alpha u(-s)$, and observe setting $t_r = 0$ is weakly optimal. If (22) is the only binding constraint, then, at the optimum, the positive transfers (i.e., \bar{t}, t_a if $\alpha > \tilde{\alpha}$ and otherwise \bar{t}) are strictly and continuously increasing in G . As a result, there exists a threshold $\tilde{G}(\beta, s_r)$ such that either condition (26), if $\alpha > \tilde{\alpha}$, or condition (27), if $\alpha \leq \tilde{\alpha}$, holds if and only if $G \geq \tilde{G}(\beta, s_r)$. Also, because $u(-s_r) > \rho\beta\alpha u(-s)$, $\tilde{G}(\beta, s_r) > -\rho\beta\alpha u(-s)$ necessarily. We have stated that setting $t_r = 0$ is weakly optimal: specifically, when $G \geq \tilde{G}(\beta, s_r)$, setting $t_r \in [0, \tilde{t}_r]$ is optimal, where \tilde{t}_r is defined as above.

Suppose now $G < \tilde{G}(\beta, s_r)$ and $\rho\beta\alpha u(-s) < u(-s_r)$, so that (23) necessarily binds. Suppose (23) is the only binding constraint. The optimal way to satisfy (23) consists of setting $u_r = u(-s_r)$ and $\bar{u} = \underline{u} = u_a = G + u(-s_r)$. This solution is such that (22) indeed holds if and only if:

$$\rho\beta\alpha(G + u(-s_r)) \geq G + \rho\beta\alpha(t_c - s), \quad (28)$$

where it is weakly optimal to set $t_c = 0$. There are then two cases. If (28) holds when $t_c = 0$, it is optimal to set $t_c \in [0, \hat{t}_c]$, where \hat{t}_c is the highest value of t_c such that (28) holds, and $u(\bar{t}) = u(t_a) = u(\underline{t}) = G - l(s_r)$ and $t_r = 0$.³⁴ In case (28) is violated when $t_c = 0$, both (22) and (23) bind at the optimum. It is then optimal to set $\bar{t} > t_c = t_r = 0$, and it is ambiguous—if one does not specify $u(\cdot)$ —whether t_a and \underline{t} are strictly positive or equal to zero.³⁵ Nevertheless, we know $\rho u(\bar{t}) + (1 - \rho)(\beta u(t_a) + (1 - \beta)u(\underline{t})) = G + u(-s_r)$ at the optimum. Also, by a standard Envelope argument, the expected payoff of the employer is strictly and continuously decreasing in G on $G \in [l(s_r), \tilde{G}(\beta, s_r)]$, and the expected transfer $\rho\bar{t}(G, s_r) + (1 - \rho)(\beta t_a(G, s_r) + (1 - \beta)\underline{t}(G, s_r)) = \mathbb{E}[t^2(G, s_r)]$ is strictly and continuously increasing in G on $G \in [l(s_r), \tilde{G}(\beta, s_r)]$. Finally, $\mathbb{E}[t^2(\tilde{G}, s_r)] = \mathbb{E}[t^1(\tilde{G})]$ because, at $G = \tilde{G}(\beta, s_r)$, maximizing (21) subject to (22) alone must yield the same expected payoff to the employer as when maximizing (21) subject to both (22) and (23). It follows the expected transfer is strictly

³⁴When $u(x) = x$, any $(\bar{t}, t_a, \underline{t})$ such that $\rho\bar{t} + (1 - \rho)(\beta t_a + (1 - \beta)\underline{t}) = G - s_r$ is optimal if $\rho\beta\alpha(G - s_r) \geq G - \rho\beta\alpha s$ holds. In these cases, it is also optimal to set $t_r = 0$ and $t_c \in [0, \frac{\rho\beta\alpha(G - s_r) - (G - \rho\beta\alpha s)}{\rho\beta\alpha}]$. If instead $\rho\beta\alpha(G - s_r) < G - \rho\beta\alpha s$, setting $\bar{t} = \frac{G - s_r}{\rho} > t_a = \underline{t} = 0$ and $t_c \in [0, \frac{1 - \rho}{\rho^2\beta\alpha}(\tilde{G}(\beta, s_r) - G)]$ is optimal.

³⁵If $\alpha > \tilde{\alpha}$, then $t_a > 0$.

and continuously increasing in G for $G \in [l(s_r), \infty)$.

B Proofs of Lemma 3 and 4

Throughout, $G > \min[\rho\beta\alpha l(s), l(s_r)]$. We first prove Lemma 3. Because concave functions appear on both sides of (8) and (9), we operate the same change of variables adopted as above. The employer chooses payoffs $\bar{u} \geq 0$, $\underline{u} \geq 0$, $u_a \geq 0$, $u_c \geq u(-s)$, and $u_r \geq u(-s_r)$ to maximize:

$$\rho\underline{\pi} + (1 - \rho)\bar{\pi} - \rho(\beta(\alpha h(u_c) + (1 - \alpha)h(u_a)) + (1 - \beta)h(\underline{u})) - (1 - \rho)h(\bar{u}) - \rho\beta\alpha S, \text{ subject to} \quad (29)$$

$$G + \rho(\beta(\alpha u_c + (1 - \alpha)u_a) + (1 - \beta)\underline{u}) + (1 - \rho)\bar{u} \geq \rho\bar{u} + (1 - \rho)(\beta u_a + (1 - \beta)\underline{u}), \quad (30)$$

$$G + \rho(\beta(\alpha u_c + (1 - \alpha)u_a) + (1 - \beta)\underline{u}) + (1 - \rho)\bar{u} \geq G + u_r, \quad (31)$$

$$G + (\beta(\alpha u_c + (1 - \alpha)u_a) + (1 - \beta)\underline{u}) + (1 - \rho)\bar{u} \geq 0. \quad (32)$$

When $\rho\beta\alpha u(-s) \geq u(-s_r)$, setting $\bar{u} = u_a = \underline{u} = 0$, $u_c = u(-s)$, and $u_r \in [u(-s_r), \rho\beta\alpha u(-s)]$ is optimal. Suppose now $\rho\beta\alpha u(-s) < u(-s_r)$. Setting all *ex post* payoffs equal to their lower bounds cannot be optimal as inequality (31) would be violated. Suppose (31) is the only binding constraint. It is then strictly optimal to set $u_r = u(-s_r)$, that is, to set $t_r = 0$. If the remaining *ex post* payoffs were strictly above their lower bounds, and denoting by λ the multiplier associated to (31), the FOCs—stated with the transfers as choice variables—would be equal to:

$$\frac{1}{u'(t_c - s)} = \frac{1}{u'(t_a)} = \frac{1}{u'(\bar{t})} = \frac{1}{u'(\underline{t})} = \lambda. \quad (33)$$

Ideally, the employer would like to set $t_c - s = t_a = \bar{t} = \underline{t}$ so as to bind (31). However, this would require negative transfers, which is not feasible. Because $u(\cdot)$ is concave, setting $t_c > 0$ and all other transfers equal to zero is optimal. Constraints (30) and (32) are then also satisfied. When $u(x) = x$, setting $t_r = 0$ and the other transfers so that the equilibrium expected transfer is equal to $\rho\beta\alpha s - s_r$ is optimal.

We now provide the proof of Lemma 4. The optimization problem is identical to that in Lemma 3, except the employer maximizes $\rho\underline{\pi} + (1 - \rho)\bar{\pi} - h(t_r) - S_r$ and inequality (31) is replaced by:

$$G + u_r \geq G + \rho(\beta(\alpha u_c + (1 - \alpha)u_a) + (1 - \beta)\underline{u}) + (1 - \rho)\bar{u}. \quad (34)$$

When $u(-s_r) \geq \rho\beta\alpha u(-s)$, setting $u_r = 0$ and \bar{u} , u_a , \underline{u} , and u_c so that $\rho(\beta(\alpha u_c + (1-\alpha)u_a) + (1-\beta)\underline{u}) + (1-\rho)\bar{u} \in [\rho\beta\alpha u(-s), u(-s_r)]$ is optimal because $G + u(-s_r) > 0$. When $u(-s_r) < \rho\beta\alpha u(-s)$, the proof proceeds exactly as when $\rho\beta\alpha u(-s) < u(-s_r)$ in the proof of Lemma 3. Setting $t_r > 0$ and all other transfers equal to zero is optimal.

C Proof of Lemma 6

For brevity, we focus on the case in which $\alpha \leq \tilde{\alpha}$. The proof when $\alpha > \tilde{\alpha}$ is identical.

Suppose first $S_r > \max[\tilde{S}(\beta, s_r), 0]$ and $u(-s_r) \leq \rho\beta\alpha u(-s)$. Preventing misconduct is costless—and thus optimal—for all employers whose employees' gain $G \leq \rho\beta\alpha l(s)$. When $G > \rho\beta\alpha l(s)$, an employer prevents the harmful act if and only if $\pi_0 - \rho\bar{t}^1(G) \geq \pi_1 - \rho\beta\alpha S$ (Lemmas 1 and 3). Because $\bar{t}^1(G)$ is increasing in G , there exists a threshold $\hat{G}(\beta, s_r, S_r)$ such that employers prevent misconduct if and only if $G \leq \hat{G}(\beta, s_r, S_r)$. Because $(2\rho - 1)\Delta\pi + \rho\beta\alpha S > 0$, $\hat{G} > \rho\beta\alpha l(s)$.

Suppose now $S_r > \tilde{S}(\beta, s_r)$ and $u(-s_r) > \rho\beta\alpha u(-s)$. Preventing misconduct is costless—and thus optimal—for all employers whose employees' gain $G \leq l(s_r)$ (see Lemma 2). If $l(s_r) < G \leq \tilde{G}(\beta, s_r)$, an employer prevents misconduct if and only if $\pi_0 - \mathbb{E}[t^2(G, s_r)] \geq \pi_1 - \rho\beta\alpha t_c^3(s_r) - \rho\beta\alpha S$. Because $\mathbb{E}[\bar{t}^2(G, s_r)]$ is increasing in G , while $t_c^3(s_r)$ is independent of G (Lemma 3), there exists a threshold $\hat{G}'(S)$ such that employers prevent misconduct if and only if $G \leq \hat{G}'(S)$. Note $\hat{G}'(S)$ is increasing in S , implying that $\min[\tilde{G}(\beta, s_r), \hat{G}'(S)] = \tilde{G}(\beta, s_r)$ if and only if S is greater than some threshold $S_0 \leq 0$.³⁶ Similarly, employers whose employees' gain $G > \tilde{G}(\beta, s_r)$ prevent misconduct if and only if $\pi_0 - \rho\bar{t}^1(G) \geq \pi_1 - \rho\beta\alpha t_c^3(s_r) - \rho\beta\alpha S$. Because $\bar{t}^1(G)$ is increasing in G , there exists a threshold $\hat{G}''(S)$ such that employers prevent misconduct if and only if $G \leq \hat{G}''$. Moreover, because the expected transfer incurred by the employers who prevent misconduct is *continuously* increasing in G over $[l(s_r), \infty)$, $\hat{G}'' \geq \tilde{G}(\beta, s_r)$ if and only if $S \geq S_0$. Finally, the relevant deterrence threshold, either \hat{G}' or \hat{G}'' , is *strictly* higher than $l(s_r)$ because $(2\rho - 1)\Delta\pi + \rho\beta\alpha S > 0$.

Consider now the case in which $0 \leq S_r \leq \tilde{S}(\beta, s_r)$. If $l(s_r) \geq \rho\beta\alpha l(s)$, all employers such that $G \leq \rho\beta\alpha l(s)$ prevent the harmful act because doing so is costless. Employers such that their employees' gain $G > \rho\beta\alpha l(s)$ prevent misconduct if and only if $\pi_0 - \rho\bar{t}^1(G) \geq \pi_1 - t_r^4(s_r) - S_r$. Because $\bar{t}^1(G)$ is increasing in G , there exists a threshold \hat{G} such that employers prevent the harmful act if

³⁶Therefore, if $S_0 > 0$, $\min[\tilde{G}(\beta, s_r), \hat{G}'(S)] = \hat{G}'(S)$ if and only if $S \leq S_0$.

and only if $G \leq \hat{G}$. Because $(2\rho - 1) \Delta\pi + \rho\beta\alpha S > 0$, $\hat{G} > \rho\beta\alpha l(s)$.

If $l(s_r) < \rho\beta\alpha l(s)$, the proof proceeds as in the case in which $S_r > \tilde{S}(\beta, s_r)$ and $l(s_r) < \rho\beta\alpha l(s)$, replacing $\rho\beta\alpha S$ with S_r and $t_c^3(s_r)$ with 0.

D Proof of Proposition 1

For brevity, we focus on the case in which $\alpha \leq \tilde{\alpha}$. The proof when $\alpha > \tilde{\alpha}$ is identical. We prove setting $S_r \leq \tilde{S}(\beta, s_r)$ is optimal by showing that welfare is strictly higher when $S_r = \tilde{S}(\beta, s_r)$ compared to when $S_r > \tilde{S}(\beta, s_r)$. Suppose first $\min[l(s_0), l(s)] \geq l(s_r) \geq \rho\beta\alpha l(s)$ and $S_r > \tilde{S}(\beta, s_r) > 0$. Social welfare is given by:

$$\begin{aligned} W(\beta, s_r, S_r) &= \int_0^{\rho\beta\alpha l(s)} (\pi_0 - (1 - \rho)(h + \beta k)) dF(G) \\ &+ \int_{\rho\beta\alpha l(s)}^{\hat{G}(\beta, s_r, S_r)} (u(t^1(G)) + \pi_0 - t^1(G) - (1 - \rho)(h + \beta k)) dF(G) \\ &+ \int_{\hat{G}(\beta, s_r, S_r)}^{\infty} (G - \rho\beta\alpha l(s) + \pi_1 - \rho(h + \beta k) + \beta\alpha s) dF(G). \end{aligned} \quad (35)$$

The first (second) term is associated with the employers for whom it is costless (costly) to prevent the harmful act. The third term is associated with the employers who tolerate the harmful act. The expected payoffs are computed using the transfers identified in Lemmas 1 and 3.

Suppose now the authority keeps β and s_r unchanged, but sets $S_r = \tilde{S}(\beta, s_r) = \rho\beta\alpha S - t_r^4(s_r)$. Also, recall t_r^4 is defined as $u(t_r^4 - s_r) = \rho\beta\alpha u(-s)$. Social welfare is then identical to (35), except for the third term which is replaced by:

$$\int_{\hat{G}(\beta, s_r, S_r)}^{\infty} (G - \rho\beta\alpha l(s) + \pi_1 - \rho h + \beta\alpha s) dF(G).$$

From Lemma 8, we know $\hat{G}(\beta, s_r, S_r)$ is independent of S_r for $\forall S_r \geq \tilde{S}(\beta, s_r)$. It follows welfare is higher when $S_r = \tilde{S}(\beta, s_r)$ instead of $S_r > \tilde{S}(\beta, s_r)$.

Suppose now $l(s) \geq l(s_r) > l(s_0)$, so that $\tilde{S}(\beta, s_r) < 0 \leq S_r$. One verifies $W(\beta, s_r, S_r)$ is independent of s_r for $\forall s_r \in [s_0, s]$. As a result, the authority may as well set $s_r = s_0$ so that $\tilde{S}(\beta, s_r) = 0$. The case then proceeds as above and welfare is strictly higher when $S_r = \tilde{S}(\beta, s_r)$.

The case in which $l(s_r) < \rho\beta\alpha l(s)$ is identical and thus left out.

E Proof of Proposition 2

For brevity, we focus on the case in which $\alpha \leq \tilde{\alpha}$. The proof when $\alpha > \tilde{\alpha}$ is identical. Throughout, $S_r = \tilde{S}(\beta, s_r)$.³⁷ Using (11), the expected payoff of an employer who tolerates the harmful act is equal to:

$$\Pi_1 = \begin{cases} \pi_1 - \rho\beta\alpha S & \text{if } l(s_0) \geq l(s_r) \geq \rho\beta\alpha l(s), \\ \pi_1 - \rho\beta\alpha (t_c^3(s_r) + S) & \text{if } \rho\beta\alpha l(s) > l(s_r) \geq 0, \end{cases} \quad (36)$$

where $t_c^3(s_r)$ is decreasing in s_r .

Suppose s_r is such that $l(s_0) \geq l(s_r) \geq \rho\beta\alpha l(s)$. From Lemma 6, we know $\hat{G}(\beta, s_r, S_r) > \rho\beta\alpha l(s)$, that is, the marginal employers incur a strictly positive expected transfer. It follows inequality (5) evaluated at $G = \hat{G}(\beta, s_r, S_r)$ is slack. There thus exists a nonempty interval $[s', l^{-1}(\rho\alpha\beta l(s))]$ such that inequality (5) holds for any $s_r \in [s', l^{-1}(\rho\alpha\beta l(s))]$. Further, the expected payoff of an employer who tolerates the harmful act is lower the lower s_r is when $l(s_r) \in [0, \rho\beta\alpha l(s)]$ (see (36)). Therefore, setting $s_r \in [s', l^{-1}(\rho\alpha\beta l(s))]$ (i) does not affect the expected payoff of the employers who are marginal when $l(s_r) \geq \rho\beta\alpha l(s)$ and (ii) decreases the expected payoff of the employers who tolerate misconduct. Setting $s' \leq s_r < \rho\alpha\beta l(s)$ thus leads to a strict increase in the fraction of employers who prevent the harmful act.

F Proof of Proposition 3

Proof that Setting $S_r = \tilde{S}(\beta, s_r) = \rho\beta\alpha(s + S) - s_r$ is Optimal

Case 1. Suppose the authority wishes (β, s_r, S_r) to be so that $s_r \geq \rho\beta\alpha s$ and $S \leq \tilde{S}(\beta, s_r)$. Rearranging the corresponding equation that defines the marginal employers, i.e.,

$$\pi_0 - \frac{\rho}{2\rho - 1} (G - \rho\beta\alpha s) = \pi_1 - (s_r - \rho\beta\alpha s) - S_r,$$

we obtain the deterrence threshold:

$$\hat{G}(\beta, s_r, S_r) = \frac{2\rho - 1}{\rho} ((2\rho - 1) \Delta\pi + s_r + S_r) + \frac{1 - \rho}{\rho} \rho\beta\alpha s.$$

³⁷Recall it is weakly optimal for the authority to set $s_r \leq s_0$, so that $\tilde{S}(\beta, s_r) \geq 0$.

Take any (β', s'_r, S'_r) such that $s'_r \geq \rho\beta'\alpha s$ and $S'_r < \tilde{S}(\beta, s'_r)$. There exists a function $\tilde{\beta}(S_r, \beta', S'_r) = \beta' + \frac{(2\rho-1)}{\rho(1-\rho)\alpha s} (S'_r - S_r)$ such that $\frac{\partial}{\partial S_r} \hat{G}(\tilde{\beta}(S_r, \beta', S'_r), s'_r, S_r) = 0$. Let S''_r be defined as $S''_r = \rho\tilde{\beta}(S''_r, \beta', S'_r)\alpha(s+S) - s'_r$. Also, let β'' denote $\tilde{\beta}(S''_r, \beta', S'_r)$. Welfare is higher under (β'', s'_r, S''_r) than under (β', s'_r, S'_r) because (i) $\beta'' < \beta'$ and (ii) $\hat{G}(\beta'', s'_r, S''_r) = \hat{G}(\beta', s'_r, S'_r)$. Therefore, for any (β', s'_r, S'_r) such that $s'_r \geq \rho\beta'\alpha s$ and $S'_r < \tilde{S}(s'_r)$, there exists an alternative policy (β'', s'_r, S''_r) , such that $S''_r = \tilde{S}(\beta'', s'_r)$, which yields higher welfare.

Case 2. Suppose now the authority wishes (β, s_r, S_r) to be so that $s_r \leq \rho\beta\alpha s$ and $S_r \leq \tilde{S}(\beta, s_r)$. The marginal employers are such that their employees' gain $G \leq \tilde{G}(\beta, s_r) = \frac{1}{1-\rho}(\rho^2\beta\alpha s - (2\rho-1)s_r)$ (see Lemma 2). Suppose first $\hat{G}(\beta, s_r, S_r) \geq \tilde{G}(\beta, s_r)$, so that the marginal employers are given by:

$$\pi_0 - \frac{\rho}{2\rho-1}(G - \rho\beta\alpha s) = \pi_1 - S_r,$$

which yields the deterrence threshold $\hat{G}(\beta, s_r, S_r) = \rho\beta\alpha s + \frac{2\rho-1}{\rho}((2\rho-1)\Delta\pi + S_r)$. One verifies indeed $\hat{G}(\beta, s_r, S_r) \geq \tilde{G}(\beta, s_r)$ if and only if $\frac{1-\rho}{\rho}((2\rho-1)\Delta\pi + S_r) + s_r \geq \rho\beta\alpha s$.

Take any (β', s'_r, S'_r) so that $\hat{G}(\beta', s'_r, S'_r) \geq \tilde{G}(\beta', s'_r)$, $s'_r \leq \rho\beta'\alpha s$, and $S'_r < \tilde{S}(\beta', s'_r)$. There exists a function $\tilde{\beta}(S_r, \beta', S'_r) = \beta' + \frac{(2\rho-1)}{\rho^2\alpha s} (S'_r - S_r)$ such that $\frac{\partial \hat{G}}{\partial S_r}(\tilde{\beta}(S_r, \beta', S'_r), s'_r, S_r) = 0$, and let S_r^0 be defined as $s'_r = \rho\tilde{\beta}(S_r^0, \beta', S'_r)\alpha s$. There are two cases to consider depending on S_r^0 . Suppose first $S_r^0 < \rho\tilde{\beta}(S_r^0, \beta', S'_r)\alpha(s+S) - s'_r = \rho\tilde{\beta}(S_r^0, \beta', S'_r)\alpha S$, and let (i) S''_r be defined as $S''_r = \rho\tilde{\beta}(S''_r, \beta', S'_r)\alpha S$ and (ii) s''_r be defined as $s''_r = \rho\tilde{\beta}(S''_r, \beta', S'_r)\alpha s$. Then, $(\tilde{\beta}(S''_r, \beta', S'_r), s''_r, S''_r)$ yields strictly higher welfare than (β', s'_r, S'_r) because (i) $\tilde{\beta}(S''_r, \beta', S'_r) < \beta'$ and (ii) $\hat{G}(\tilde{\beta}(S''_r, \beta', S'_r), s''_r, S''_r) = \hat{G}(\beta', s'_r, S'_r)$.³⁸ Suppose now $S_r^0 \geq \rho\tilde{\beta}(S_r^0, \beta', S'_r)\alpha(s+S) - s'_r$, and let S''_r be defined as $S''_r = \rho\tilde{\beta}(S''_r, \beta', S'_r)\alpha(s+S) - s'_r$. Also, let $\tilde{\beta}(S''_r, \beta', S'_r)$ be denoted by β'' . By assumption, $s'_r \leq \rho\beta''\alpha s$ holds, and welfare is higher under (β'', s'_r, S''_r) than under (β', s'_r, S'_r) because (i) $\beta'' < \beta'$ and (ii) $\hat{G}(\beta'', s'_r, S''_r) = \hat{G}(\beta', s'_r, S'_r)$.³⁹

Suppose now $\hat{G}(\beta, s_r, S_r) \leq \tilde{G}(\beta, s_r)$. The marginal employers are given by:

$$\pi_0 - (G - s_r) = \pi_1 - S_r,$$

which yields threshold $\hat{G}(s_r, S_r) = (2\rho-1)\Delta\pi + s_r + S_r$. One verifies $\hat{G}(s_r, S_r) \leq \tilde{G}(\beta, s_r)$ if and

³⁸Note the equality $s''_r = \rho\tilde{\beta}(S''_r, \beta', S'_r)\alpha s$ implies the condition $\hat{G}(\beta, s_r, S_r) \geq \tilde{G}(\beta, s_r)$.

³⁹Also, $S''_r > S'_r$ and $\beta'' < \beta'$ imply the condition $\hat{G}(\beta, s_r, S_r) \geq \tilde{G}(\beta, s_r)$ holds.

only if $\frac{1-\rho}{\rho} ((2\rho - 1) \Delta\pi + S_r) + s_r \leq \rho\beta\alpha s$. Notice the threshold $\hat{G}(s_r, S_r)$ is independent of β and $s_r \leq \rho\beta\alpha s$ is implied by $\hat{G}(\beta, s_r, S_r) \leq \tilde{G}(\beta, s_r)$.

Take any (β', s'_r, S'_r) so that $\hat{G}(\beta', s'_r, S'_r) \leq \tilde{G}(\beta', s'_r)$ and $S'_r < \tilde{S}(\beta', s'_r)$. Let β^0 be defined as $\frac{1-\rho}{\rho} ((2\rho - 1) \Delta\pi + S'_r) + s'_r = \rho\beta^0\alpha s$. We distinguish between two cases depending on β^0 . Suppose first β^0 is such that $S_r < \rho\beta^0\alpha(s + S) - s_r$. Then, (β^0, s'_r, S'_r) achieves a higher level of welfare than (β', s'_r, S'_r) because (i) $\beta^0 < \beta'$ and (ii) $\hat{G}(\beta^0, s'_r, S'_r) = \hat{G}(\beta', s'_r, S'_r)$. By construction, (β^0, s'_r, S'_r) is such that $\hat{G}(\beta^0, s'_r, S'_r) \geq \tilde{G}(\beta^0, s'_r)$. We then know from the first part of **Case 2** that there exists an alternative policy, such that $S_r = \rho\beta\alpha(s + S) - s_r$, which yields a higher level of welfare than (β^0, s'_r, S'_r) . Suppose now β^0 is such that $S_r \geq \rho\beta^0\alpha(s + S) - s_r$. There then again exists an alternative policy (β'', s'_r, S'_r) , where β'' is such that $S_r = \rho\beta''\alpha(s + S) - s_r$, which yields a higher level of welfare than (β', s'_r, S'_r) because (i) $\beta'' < \beta'$ and (ii) $\hat{G}(\beta'', s'_r, S'_r) = \hat{G}(\beta', s'_r, S'_r)$.

Proof that Setting $s_r < \rho\beta\alpha s$ is Optimal

When $S_r = \tilde{S}(\beta, s_r)$, the deterrence threshold is equal to $\hat{G}(\beta) = \rho\beta\alpha s + \frac{2\rho-1}{\rho} ((2\rho - 1) \Delta\pi + \rho\beta\alpha S)$ for $\forall s_r$ such that $s_r \geq \rho\beta\alpha s$. Therefore, setting $s_r \leq \rho\beta\alpha s$ is w.l.o.g..

Take any (β', s'_r) , where $s'_r = \rho\beta'\alpha s$, and note there exists a function $\tilde{\beta}(s_r, \beta', s'_r) = \beta' + \frac{(2\rho-1)}{\rho} \frac{s_r - s'_r}{(\rho\alpha s + (2\rho-1)\alpha(s+S))}$ such that $\frac{\partial}{\partial s_r} \hat{G}(\tilde{\beta}(s_r, \beta', s'_r), s'_r) = 0$. Moreover, $\hat{G}(\beta', s'_r) > \tilde{G}(\beta', s'_r) = \rho\beta'\alpha s$. There then always exists an alternative policy $(\tilde{\beta}(s''_r, \beta', s'_r), s''_r)$, where $s''_r < s'_r$, which yields higher welfare because (i) $\tilde{\beta}(s''_r, \beta', s'_r) < \beta''$ and (ii) $\hat{G}(\tilde{\beta}(s''_r, \beta', s'_r), s''_r) = \hat{G}(\beta', s'_r)$.⁴⁰

Proof that Setting $s_r = 0$ is Weakly Optimal

Take any (β', s'_r) such that $0 < s'_r \leq \rho\beta'\alpha s'$. Suppose first $\hat{G}(\beta', s'_r) > \tilde{G}(\beta', s'_r)$. Also, recall for $\forall (\beta, s_r)$ such that $\hat{G}(\beta, s_r) > \tilde{G}(\beta, s_r)$, the deterrence threshold is given by:

$$\hat{G}(\beta, s_r) = \rho\beta\alpha s + \frac{2\rho-1}{\rho} ((2\rho - 1) \Delta\pi + \rho\beta\alpha(s + S) - s_r).$$

There exists $\tilde{\beta}(s_r, \beta', s'_r) = \beta' + \frac{(2\rho-1)}{\rho} \frac{s_r - s'_r}{(\rho\alpha s + (2\rho-1)\alpha(s+S))}$ so that $\frac{\partial}{\partial s_r} \hat{G}(\tilde{\beta}(s_r, \beta', s'_r), s'_r) = 0$. Also, when $s_r = 0$, $\tilde{\beta}(s_r, \beta', s'_r) > 0$.⁴¹ Let \tilde{s}_r denote the value of s_r such that $\hat{G}(\beta', s'_r) =$

⁴⁰Note $s''_r < \rho\tilde{\beta}(s''_r, \beta', s'_r)\alpha s$ because $\frac{\partial \rho\tilde{\beta}(s''_r, \beta', s'_r)\alpha s}{\partial (-s_r)} < 1$.

⁴¹To see this multiply $\tilde{\beta}(s_r, \beta', s'_r)$ by $\rho\alpha s$.

$\tilde{G}(\tilde{\beta}(\tilde{s}_r, \beta', s'_r), \tilde{s}_r)$. Then, $(\tilde{\beta}(\max(\tilde{s}_r, 0), \beta', s'_r), \max(\tilde{s}_r, 0))$ yields strictly higher welfare than (β', s'_r) because (i) $\tilde{\beta}(\max(\tilde{s}_r, 0), \beta', s'_r) < \beta'$ and (ii) $\tilde{G}(\tilde{\beta}(\max(\tilde{s}_r, 0), \beta', s'_r), \max(\tilde{s}_r, 0)) = \hat{G}(\beta', s'_r)$.

Suppose $\tilde{s}_r > 0$ and fix $\beta'' = \tilde{\beta}(\tilde{s}_r, \beta', s'_r)$. Then, for $\forall s_r \leq \tilde{s}_r$, the deterrence threshold is $\hat{G}(\beta, s_r) = (2\rho - 1)\Delta\pi + \rho\beta\alpha(s + S)$. The threshold $\hat{G}(\beta, s_r)$ is independent of s_r , so that it follows $(\beta'', 0)$ achieves the same level of welfare as $(\tilde{\beta}(\tilde{s}_r, \beta', s'_r), \tilde{s}_r)$.

Suppose now (β', s'_r) is such that $\hat{G}(\beta', s'_r) \leq \tilde{G}(\beta', s'_r)$. Recall for $\forall(\beta, s_r)$ such that $\hat{G}(\beta, s_r) \leq \tilde{G}(\beta, s_r)$, the deterrence threshold is $\hat{G}(\beta, s_r) = (2\rho - 1)\Delta\pi + \rho\beta\alpha(s + S)$, which is independent of s_r . It follows $(\beta', 0)$ yields the same level of welfare as (β', s'_r) .

References

Abrantes-Metz, R. and Sokol, D. D.. “Antitrust Corporate Governance and Compliance.” *The Oxford Handbook of International Antitrust Economics*, Volume 2 Edited by Roger D. Blair and D. Daniel Sokol (2013).

Arlen, J.. “The Potential Perverse Effects of Corporate Criminal Liability.” *Journal of Legal Studies*: Vol. 23: No. 3, Article 7 (1994).

Arlen, J.. “Economic Analysis of Corporate Criminal Liability: Theory and Evidence.” *Research Handbook on the Economics of Criminal Law*, 144 (Keith Hylton and Alon Harel, eds., Cheltenham, U.K. ; Northampton, MA : Edward Elgar, 2012)

Aubert, C.. “Managerial Effort Incentives and Market Collusion.” *TSE Working Papers* 09-127 (2009).

Aubert, C., Rey, P., and Kovacic, W. E.. “The impact of leniency and whistle-blowing programs on cartels.” *International Journal of Industrial Organization*, vol. 24(6), pages 1241-1266, 2006.

Dyck, A., Morse, A., and Zingales, L.. “Who Blows the Whistle on Corporate Fraud?,” *Journal of Finance*, 65, no. 6 (2010): 2213 - 2253.

Hammond, S.D.. “Cornerstones of an Effective Cartel Leniency Programme.” 4 (2) *Competition Law International* 4 (2008).

Harrington, J. E. Jr. “Optimal Corporate Leniency Programs.” *The Journal of Industrial Economics*, Volume 56, Issue 2, pages 215-246, June 2008.

Harrington, J. E. Jr. “Corporate Leniency Programs when Firms have Private Information: The Push of Prosecution and the Pull of Pre-emption.” *The Journal of Industrial Economics*, Volume 61, Issue 1, pages 1-27, March 2013.

- Hiriart, Y. and Martimort, D.. “The benefits of Extended Liability.” *The RAND Journal of Economics*, RAND Corporation, vol. 37(3), pages 562-582, 09 (2006).
- Innes, R.. “Remediation and Self-Reporting in Optimal Law Enforcement.” *Journal of Public Economics*, Volume 72, Issue 3, June 1999, Pages 379-393.
- Kaplow, L. and Shavell, S.. “Optimal Law Enforcement with Self-Reporting of Behavior,” *Journal of Political Economy*, 1994, 102(3): 583-606.
- Marshall, R. C., Marx, L., and Mezzetti, C.. “Antitrust Leniency with Multi-Market Colluders,” *American Economic Journal: Microeconomics* 7(3), 205-240 (2015).
- Marvo, C. M. P. and Spagnolo, G.. “What Do We Know About the Effectiveness of Leniency Policies? A Survey of the Empirical and Experimental Evidence,” *Mimeo Stockholm School of Economics* (October 1, 2014).
- Miller, N. H.. “Strategic Leniency and Cartel Enforcement. *The American Economic Review*, 99.3 (2009): 750-768.
- Motta, M. and Polo, M.. “Leniency Programs and Cartel Prosecution.” *International Journal of Industrial Organization*, 21, 347-379, 2003.
- Mullin, W. P. and Snyder, C. M.. “Should Firms be Allowed to Indemnify Their Employees for Sanctions?” *Journal of Law, Economics, Organization* Vol. 26, No. 1 (Apr., 2010), pp. 30-53.
- Polinsky, A. M. and Shavell, S.. “Should Employees be Subject to Fines and Imprisonment Given the Existence of Corporate Liability?” *International Review of Law and Economics*, Volume 13, Number 3, September 1993, pages 239-257.
- Segerson, K., and Tietenberg T.. “The Structure of Penalties in Environmental Enforcement: An Economic Analysis.” *The Journal of Environmental Economics and Management* 23, 179-200 (1992).
- Shavell, S.. “The Optimal Level of Corporate Liability Given the Limited Ability of Corporations to Penalize Their Employees,” *International Review of Law and Economics*, 17, 203-213, 1997.
- Spagnolo, G.. “Divide et Impera: Optimal Leniency Programs.” CEPR Discussion Paper (2005).