ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

OPTIMAL LAW ENFORCEMENT WITH ORDERED LENIENCY

Claudia M. Landeo
Kathryn E. Spier

Discussion Paper No. 977

10/2018

Harvard Law School
Cambridge, MA 02138

Optimal Law Enforcement with Ordered Leniency*

Claudia M. Landeo† and Kathryn E. Spier‡

September 13, 2018

Abstract

This paper studies the design of enforcement policies to detect and deter harmful short-term activities committed by groups of injurers. With an ordered-leniency policy, the degree of leniency granted to an injurer who self-reports depends on his or her position in the self-reporting queue. By creating a "race to the courthouse," ordered-leniency policies lead to faster detection and stronger deterrence of illegal activities. The socially-optimal level of deterrence can be obtained at zero cost when the externalities associated with the harmful activities are not too high. Without leniency for self-reporting, the enforcement cost is strictly positive and there is underdeterrence of harmful activities relative to the first-best level. Hence, ordered-leniency policies are welfare improving. Our findings for environments with groups of injurers complement Kaplow and Shavell's (1994) results for single-injurer environments.

KEYWORDS: Law Enforcement; Ordered Leniency; Self-Reporting; Leniency; Harmful Externalities; Non-Cooperative Games; Prisoners' Dilemma Game; Coordination Game; Risk Dominance; Pareto Dominance; Corporate Misconduct; White-Collar Crime; Securities Fraud; Insider Trading; Market Manipulation; Whistleblowers; Plea Bargaining; Tax Evasion; Environmental Policy Enforcement

JEL Categories: C72, D86, K10, L23

*We acknowledge financial support from the National Science Foundation (NSF Grant SES-1155761). We thank Tom Brennan, Dan Coquillette, Nick Feltovich, John Goldberg, Christine Jolls, Louis Kaplow, Max Nikitin, Jack Ochs, Steve Shavell and Abe Wickelgren for insightful discussions and comments. We are grateful for suggestions from participants at the 2018 NBER Summer Institute in Law and Economics and the Harvard Law School Faculty Workshop. We thank Susan Norton for administrative assistance.

†University of Alberta, Department of Economics. Henry Marshall Tory Building 7-25, Edmonton, AB T6G 2H4. Canada. landeo@ualberta.ca, tel. 780-492-2553.

‡Harvard Law School and NBER. 1575 Massachusetts Ave., Cambridge, MA 02138. United States. kspier@law.harvard.edu, tel. 617-496-0019.

1 Introduction

Illegal activities are often committed by groups of people working together rather than by individuals working alone. Common examples in the corporate setting include insider trading and market manipulation schemes. In 2011, the FBI reported 726 corporate fraud cases, several of which involved losses to public investors that individually exceeded $1 billion, and 343 securities fraud cases involving more than 120,000 victims and approximately $8 billion in losses (FBI, 2012). More generally, illegal activities committed by groups of wrongdoers impose considerable costs on society. To combat illegal group activities, law enforcement agencies often grant leniency to wrongdoers who come forward and self-report.

In a typical leniency program, wrongdoers who self-report early face lower sanctions than those who self-report later.[1] For instance, in 2014, the Securities and Exchange Commission (SEC) brought insider trading charges against Christopher Saridakis, a top executive at GSI Commerce, and several co-conspirators for providing tips to family and friends in advance of eBay's acquisition of GSI. Saridakis paid a penalty equal to twice the amount of his tippees' profits,[2] and was imprisoned after pleading guilty to criminal charges. One of Saridakis' co-conspirators was forced to disgorge his own profits and paid a penalty equal to three times his own profits and all of the profits of his own tippees. In contrast, a co-conspirator who aided the prosecution paid a reduced penalty equal to one half of his profits, while another co-conspirator who cooperated early paid no penalty at all (Ceresney, 2015).[3]

This paper studies the design of enforcement policies to detect and deter illegal short-term activities committed by groups of injurers.[4] We focus on a class of mechanisms where the degree of leniency granted to an injurer depends on his or her position in a self-reporting queue. The earlier an injurer reports the act, the higher his or her position in the self-reporting

[1] An example of such a program is the Securities and Exchange Commission's Cooperation Program.

[2] In insider trading cases, the term "tipper" refers to a person who has broken his fiduciary duty by revealing inside information. The term "tippee" refers to a person who knowingly uses inside information to make a trade.

[3] See also *SEC v. Saridakis and Gardner,* Civil Action No. 14 2397 (U.S. District Court Eastern District of Pennsylvania 2014). For another interesting insider-trading case involving leniency for early cooperation, see *SEC v. Wrangell* (2012), https://www.sec.gov/litigation/complaints/2012/comp-pr2012-193-wrangell.pdf.

[4] Illegal *short-term* activities do not involve an ongoing relationship among group members. They are sometimes referred to as illegal "occasional" activities. See Buccirossi and Spagnolo (2006). In game-theoretic terms, they correspond to one-shot strategic environments. Leniency programs have been also applied to illegal *long-term* activities such as cartels. For a recent survey of this literature, see Spagnolo and Marvão (2016).

queue. We call these mechanisms "ordered-leniency policies." Our analysis demonstrates that the optimal ordered-leniency policy involves a cascade of reduced sanctions for injurers who self-report and generates a so-called "race to the courthouse" where all injurers self-report immediately.[5] By inducing self-reporting, enforcement policies with ordered leniency increase the likelihood of detection of harmful acts without raising the enforcement costs. As a result of the higher likelihood of detection, the expected fines increase, deterrence is strengthened and social welfare is improved. Although our paper is motivated by insider trading and securities fraud, our analysis applies to any kind of harmful short-term activity committed by a group of wrongdoers.[6] To the best of our knowledge, there are no previous formal studies of ordered-leniency policies for short-term group activities.[7]

We begin our analysis with a benchmark model involving an enforcement agency and two injurers. First, the enforcement agency publicly commits to an enforcement policy involving investigation efforts, a sanction, and possibly reduced sanctions for self-reporting that depend on the injurers' positions in the self-reporting queue (ordered leniency). Next, given the enforcement policy, the potential injurers decide whether to participate in a harmful group act. If the act is committed, then the injurers decide *whether* and *when* to report themselves to the authorities. The decision of an injurer to self-report hinges on the likelihood of detection if he remains silent, which itself depends on both the enforcement efforts of the agency and the self-reporting decision of the other injurer. There are negative externalities in the self-reporting stage: The likelihood that an injurer will be detected and sanctioned is higher when the other injurer reports the act.

We show that the optimal degree of leniency granted to injurers who self-report depends critically on the refinement criterion for equilibrium selection when multiple equilibria arise. When small discounts are granted to injurers who self-report (mild leniency), the self-reporting stage resembles a coordination game with two (pure-strategy) Nash equilibria: One where all injurers self-report, the risk-dominant equilibrium (Harsanyi and Selten, 1988); and, the other where no injurer self-reports, the Pareto-dominant equilibrium. When the risk-dominance

[5]The expression "race to the courthouse" typically refers to the first-to-file legal rule that provides superior rights to the first action filed in civil litigation cases. In our environment, earlier reporting raises the chances of being the first in the self-reporting queue.

[6]See Section 6 for a discussion of applications to other relevant contexts.

[7]See Landeo and Spier (2018) for a recent experimental study on law enforcement policies with ordered leniency for short-term group activities.

refinement is applied, mild leniency is the optimal leniency policy. When the Pareto-dominance refinement is applied instead, mild leniency is ineffective. In that case, the optimal leniency policy involves larger discounts to injurers who self-report (strong leniency). With strong leniency, the self-reporting stage resembles a prisoners' dilemma game with a unique (pure-strategy) Nash equilibrium where all injurers self-report.

We demonstrate that the optimal ordered-leniency policy imposes the highest possible sanction on injurers who fail to self-report but are caught nonetheless, and grants a reduced sanction for the first injurer to self-report. Depending on the strength of inculpatory evidence provided by the first injurer to self-report,[8] the second injurer to self-report may receive lenient treatment as well (albeit to a lesser degree). Granting leniency to the second injurer who reports the act is particularly valuable when the inculpatory evidence provided by the first injurer to report the act is insufficient to convict the second injurer with certainty. Granting leniency to the first injurer to self-report, or to both the first and second injurer, creates a race-to-the-courthouse where, in equilibrium, both injurers self-report immediately. As a result, the likelihood of detection increases, expected sanctions rise, and fewer harmful acts are committed.

Our social welfare analysis indicates that optimal ordered-leniency policies are welfare improving whenever the injurers have limited wealth or there is an upper bound on the fines that can be imposed. Without leniency for self-reporting, the enforcement agency's efforts must be strictly positive and there will be underdeterrence of harmful activities relative to the first-best level. Holding the fine and the costs of enforcement fixed, the optimal ordered-leniency policy will increase the expected fine, thus raising level of deterrence and increasing social welfare. We show that the socially-optimal level of deterrence can be obtained at zero cost to the enforcement agency when the externalities associated with the harmful activities (harms inflicted on others) are not too high.

We then extend our benchmark framework to groups of injurers with more than two members. Attention is restricted to coalition-proof Nash equilibria (Bernheim et al., 1987). Our analysis demonstrates that the key insights of the benchmark model apply to this setting. In particular, we show that enforcement policies with ordered leniency for self-reporting outperform enforcement policies without leniency for self-reporting. The highest level of deterrence

[8]The higher the strength of inculpatory evidence provided by an injurer who self-reports, the higher the detection probability of the injurer who does not self-report.

is achieved when all injurers who commit the act later self-report immediately, and receive successive discounts for self-reporting based on their positions in the self-reporting queue. In general, the leniency for the first injurer to report may not be full, and the leniency for the last injurer to report may not be zero. We show that the race-to-the-courthouse effect is robust to the number of members in the group of injurers. New insights are derived as well. Our analysis suggests that, by creating diseconomies of scale with respect to group size, ordered-leniency policies discourage larger-scale harmful group activities in favor of smaller-scale activities.

Finally, we discuss several additional extensions to our benchmark model such as stochastic detection rates, asymmetric benefits from committing a harmful act across injurers, and endogenous decisions about whether to commit a group or an individual harmful act. Although these environments raise some new and interesting issues, the main lessons derived from our benchmark model remain relevant.

Our paper contributes to the literature on the control of harmful externalities by presenting the first formal analysis of optimal enforcement policies with ordered leniency for harmful short-term activities conducted by a group of wrongdoers.[9] Our work is related to several strands of literature. The closest to our work are the studies on enforcement and self-reporting. Kaplow and Shavell (1994) study a probabilistic enforcement model where harmful activities are committed by individuals, not by groups. They demonstrate that leniency for self-reporting can directly reduce enforcement costs without significantly compromising deterrence. In their model, injurers who self-report pay a sanction slightly less than the expected sanction they would face if they did not report the act. Given that enforcement efforts do not need to be allocated to identify the injurers who self-report, the enforcement agency can economize on its investigatory efforts.[10] In contrast, we focus on harmful activities committed by groups of injurers. We show that granting leniency to the first injurer to report, and possibly to the subsequent injurers, increases the likelihood of detection without raising investigatory costs, raises the expected sanctions, and strengthens deterrence. In our environment, the

[9] In seminal work, Becker (1968) demonstrates that a very small probability of detection coupled with a very high sanction can deter crime at essentially zero cost. Polinsky and Shavell (1984) show that when injurers have limited assets and sanctions are bounded above, then the optimal enforcement policy involves investigation costs and deterrence falls short of the first-best level.

[10] Self-reporting is also socially valuable because early detection of harmful activities might minimize further social costs (Malik, 1993; Innes, 1999). Self-reporting has also been studied in the context of pollution (Livernois and McKenna, 1999) and tax evasion (Andreoni, 1991; and Malik and Schwab, 1991).

optimal enforcement policy with ordered leniency exploits the negative externalities between the injurers at the self-reporting stage. Our results complement the findings of Kaplow and Shavell (1994).

Feess and Walzl (2004) study enforcement with self-reporting for illegal activities committed by two-member criminal teams, focusing on the consequences of injurers' cooperation in the self-reporting stage on enforcement. In their proposed leniency policy, the *number* of injurers who self-report determines the degree of leniency for self-reporting. In equilibrium, the enforcement agency grants immunity for self-reporting (a fine equal to zero) when *exactly one* injurer self-reports, but grants (almost) no leniency when *both* injurers self-report.[11] The ordered-leniency policy studied in our paper is fundamentally different. In our framework, the first injurer to report the act receives leniency *whether or not the other injurers also report.* More specifically, with ordered leniency, the degree of leniency for an injurer who self-reports depends only on his or her position in the self-reporting queue. Importantly, an inherent feature of ordered leniency is the *time to self-report.* With ordered leniency, there is a race to the courthouse where injurers jockey for the first position in the self-reporting queue. In Feess and Walzl's (2004) environment, there is no advantage to being the first to report.[12] Our mechanism is arguably more closely aligned with how leniency policies are designed and implemented in the real world.[13]

Another strand of literature related to our paper is that on plea bargaining, where an individual has the option to plead guilty in exchange for a reduced sentence. In models with a single defendant, Landes (1971) demonstrates that plea bargaining agreements reduce prosecutorial costs and Grossman and Katz (1983) find that plea bargaining might produce insurance

[11] Through a proverbial prisoners' dilemma, maximal deterrence may be obtained at virtually no cost to the enforcement agency when the injurers do not cooperate or the probability of cooperation is exogenous.

[12] There are many other important differences between our model and theirs. Feess and Walzl (2004) assume that the Pareto-dominance refinement applies in case of multiplicity of equilibria, and hence, leniency policies under the risk-dominance refinement are not investigated. In addition, their social welfare analysis focuses on minimizing harm to victims and does not include the injurers' private benefits, and environments with more than two injurers are not investigated.

[13] Buccirossi and Spagnolo (2006) investigate the effects of leniency policies on sequential bilateral short-term illegal activities. They find that moderate leniency policies (i.e., policies involving a reduction in the sanction but not a reward for self-reporting) might have the perverse effect of providing an effective governance mechanism for illegal short-term activities that otherwise will not be implemented due to a hold-up problem. Optimal enforcement policies are not studied.

and screening effects.[14] Kobayashi (1992) studies plea bargaining using a model with two defendants where the acceptance of a plea agreement by one defendant raises the probability of conviction of the other, the probability of conviction of the more culpable defendant is higher than the probability of conviction of the less culpable defendant, and the identities of the defendants are known by the prosecutor. He finds that the plea bargaining policy that maximizes deterrence involves a lower negotiated penalty for the most culpable defendant. None of these papers consider ordered-leniency policies.[15] Our findings suggest that ordered-leniency policies would be highly effective in plea-bargaining environments too. In particular, our analysis demonstrates that maximal cooperation might be achieved by implementing coordination games through mild reductions in sanctions when the wrongdoers are sufficiently distrustful of each other after committing the unlawful act.

Our paper is also related to the literature on enforcement of competition policy and leniency programs for illegal long-term activities committed by criminal groups. Motta and Polo (2003) find that, when the enforcement authority has limited resources and, hence, is unable to prevent collusion ex-ante, leniency policies enhance welfare by increasing the likelihood of cartel cessation and shortening investigation. Spagnolo (2005) demonstrates that leniency policies undermine internal trust by increasing individual incentives to defect.[16] As a result, these policies destabilize cartels. Optimal leniency policies reward the first party to report with the fines paid by all other parties. When fines and rewards are sufficiently high, the first best is obtained at a zero cost.[17] Bigoni et al. (2012) provide experimental evidence of the effects of leniency and rewards on enforcement and the stability of collusion in repeated-game environments. They find that leniency enhances deterrence but contributes to the stabilization of surviving cartels. Prices fall to the competitive levels when rewards are provided to whistle-blowers.[18] Our framework can be an appropriate component of a repeated-game analysis of

[14]Negative effects might occur if innocent defendants are more risk-averse than guilty defendants, and innocent defendants might be induced to plead guilty. See also Reinganum (1988).

[15]See also Kraakman (1986) and Arlen and Kraakman (1997) for seminal work on third-party enforcement.

[16]See also Aubert et al. (2006).

[17]Chen and Rey (2013) extend Spagnolo (2005) by considering not only pre-investigation leniency but also post-investigation leniency. Harrington (2013) investigates the incentives to apply for leniency when each cartel member has private information about the likelihood of conviction without self-reporting and leniency is granted only to the first cartel member to self-report. Feess and Walzl (2010) study leniency policies when one cartel member might provide stronger evidence than the other member.

[18]See Apestegui et. (2007), Hinloopen and Soetevent (2008), Bigoni et al. (2015) and Feltovich and Hamaguchi (2018) for additional experimental work on leniency and cartels.

enforcement policies with self-reporting for illegal long-term group activities.

Our work shares some features with studies on contract design in the presence of externalities among contract recipients. In the context of exclusionary vertical restraints, Rasmusen et al. (1991) and Segal and Whinston (2000) demonstrate that, when there are economies of scale in production, incumbent monopolists can design profitable exclusive-dealing contracts by exploiting the negative externalities among the buyers. Landeo and Spier (2009, 2012) provide experimental evidence of the exclusionary power of these types of contracts.[19]

The rest of the paper is organized as follows. Section 2 introduces the model setup. Section 3 presents the equilibrium analysis of the injurers' decisions about committing the act, self-reporting, and the time to report. Section 4 constructs the optimal enforcement policies with and without leniency for self-reporting. We show that the optimal ordered-leniency policy always creates superior incentives, and identify necessary and sufficient conditions for achieving the first-best outcome. Section 5 extends our benchmark model to groups of injurers with more than two members, and demonstrates that the main insights derived from our benchmark model and their implications for the design of optimal enforcement policies are robust. Section 6 presents additional relevant extensions. Section 7 discusses applications to other relevant environments and concludes. Formal proofs are presented in the Appendix.

2 Model Setup

Our strategic environment consists of a game of complete information. Our benchmark framework involves three risk-neutral players: Two identical representative potential injurers and an enforcement agency. (Section 5 studies an environment involving groups of injurers with more than two members.) We assume that the potential injurers seek to maximize their private net benefits from committing a harmful act. The enforcement agency seeks to maximize social welfare. Social welfare includes the aggregation of the benefits to the injurers. It also includes the social costs: The harm inflicted on others (externalities associated with the harmful activities) and the cost of enforcement. We assume that the enforcement agency cannot costlessly identify the parties responsible for committing the harmful act. Without loss of generality, we abstract from time discounting.

[19] See Landeo and Spier (2015) and Che and Yoo (2001) for applications to incentive contracts for teams, and Kornhauser and Revesz (1994) and Spier (1994) for applications to civil litigation under joint and several liability.

The timing of the game is as follows. First, the enforcement agency publicly commits to an enforcement policy with ordered leniency to detect and prevent harmful short-term activities committed by groups of injurers. The enforcement policy components are (f, r_1, r_2, e). (1) $f \in (0, \bar{f}]$ denotes a fine or monetary sanction (measured per injurer).[20] The maximal fine, $\bar{f}$, can be greater than, lower than, or equal to the harm inflicted on others (measured per injurer), $h > 0$. (2) $r_1, r_2 \in [0, 1]$ denote the leniency multipliers that correspond to the first and second positions in the self-reporting queue, respectively, where $r_1 < r_2$, $r_1 > r_2$ or $r_1 = r_2$. The discount for position i in the reporting queue is then $1 - r_i$, $i = 1, 2$.[21] Thus, we study *ordered-leniency policies* where the first injurer to report pays $r_1 f$, regardless of whether a second injurer reports, and the second injurer to report pays $r_2 f$.[22] (3) $e \in [0, 1)$ denotes the enforcement agency's effort (investigation effort), which, as we will describe below, determines the probability that harmful acts are detected. We let $c(e)$ be the cost of enforcement or investigation (measured per injurer), and assume that $c(0) = 0$, $c'(0) = 0$, $c'(e) \geq 0$, $c''(e) > 0$, and $lim_{e \to 1} c'(e) = \infty$.[23]

Second, after observing the enforcement policy, the potential injurers play a two-stage game. In Stage 1, the potential injurers simultaneously and independently decide whether to participate in a harmful activity. The act is committed if and only if both injurers decide to participate. The benefit for each injurer is $b \in [0, \infty)$. It is distributed according to probability density function $g(b)$ and cumulative distribution function $G(b)$. Both injurers receive the same benefit. The realization of b is revealed to both potential injurers before they make their decisions regarding committing the act.[24] If the act is committed, Stage 2 starts; otherwise, the game ends. In Stage 2, the injurers simultaneously and independently decide *whether* and *when* to report the harmful act to the enforcement agency. Specifically, each injurer can choose to report the act at time $t \in [0, 1]$ where $t = 0$ represents immediate reporting and $t > 0$ represents delayed reporting.

Third, the injurers (parties responsible for causing harm), if detected, are accurately identified by the enforcement agency and sanctioned. The probabilities of detection and the sanctions are as follows. Absent any self-reporting by the injurers, harmful acts are detected

[20] $\bar{f}$ can be interpreted as the potential injurer's wealth. When the fine is above $\bar{f}$, the injurer is judgment-proof.

[21] Multipliers $(r_1, r_2) = (1, 1)$ imply that the enforcement policy does not grant leniency for self-reporting.

[22] Later, we demonstrate that the optimal ordered-leniency policy involves $r_1 < r_2$.

[23] These assumptions ensure an interior solution for the social welfare maximization problem.

[24] Note that committing the act is socially desirable if and only if the benefits, b, exceed the social harm, h.

with probability p_0 and each injurer pays a fine f. If one injurer reports the act, then the injurer who reports pays fine $r_1 f$ and the silent accomplice is accurately detected and fully sanctioned (i.e., pays a fine f) with probability p_1. If both injurers report the act, then the first to report pays fine $r_1 f$ and the second to report pays fine $r_2 f$. If the two injurers report at exactly the same time, then an equally-weighted coin flip determines who obtains the first and second positions in the self-reporting queue. Finally, we assume that p_0 and p_1 depend on the enforcement agency's effort, $e \in [0, 1)$, and p_1 also depends on the exogenous strength of inculpatory evidence, $\pi \in (0, 1)$. Specifically, $p_0(e) = e$ and $p_1(e, \pi) = e + (1 - e)\pi$.[25] Then, $0 \leq p_0(e) < p_1(e, \pi) < 1$.

The equilibrium concept is subgame-perfect Nash equilibrium. Our focus is on pure-strategy equilibria that survive the elimination of weakly-dominated strategies. When multiple pure-strategy equilibria arise, we present separate equilibrium analyses for the Pareto-dominance and risk-dominance refinements (Harsanyi and Selten, 1988).

The first-best outcome is used as a benchmark in the welfare analysis of ordered-leniency policies. The first best is defined as the social welfare outcome of an environment in which the enforcement agency can costlessly identify the parties responsible for committing the harmful act (and their private benefits) and decide which acts to prohibit. Then, in the first-best outcome, the cost of effort is zero and acts are committed if and only if $b > h$.[26]

We proceed backwards and begin our analysis with the injurers' decisions. We then analyze the optimal enforcement policy with ordered leniency and conduct social welfare analysis.

3 Injurers' Decisions: Equilibrium Characterization

We begin by characterizing the equilibrium behavior of the injurers in Stage 2, the self-reporting stage. Next, we study the potential injurers' decisions regarding participating in the act in Stage 1.

[25] This specification may be derived from first principles. Suppose that absent self-reporting by either injurer, detection is the outcome of a single Bernoulli trial with success probability $p_0 = e$. When one injurer self reports and another does not, there is a second independent Bernoulli trial that succeeds in detecting the non-reporting injurer with probability π. Then $p_1 = e + (1 - e)\pi$ is the probability that the silent injurer is detected.

[26] In practice, of course, the enforcement agency cannot costlessly identify the injurers. Hence, to detect and deter harmful acts, the enforcement agency needs to spend resources on detection and implement leniency programs for self-reporting.

3.1 Decision to Report the Act and Time to Report

Recall that when both potential injurers decide to participate in the harmful act in Stage 1, the act is committed and Stage 2 occurs. In Stage 2, the injurers simultaneously and independently decide *whether* and *when* to report the harmful act to the enforcement authority. Specifically, an injurer who decides to report the act also needs to choose the time of his or her report, $t \in [0, 1]$.

We first analyze the length of time taken by the injurers to report the harmful act. The analysis presented here is general in the sense that it allows r_1 to be greater than, equal to, or lower than r_2. In later sections, we verify that optimal enforcement policies with ordered leniency require $r_1 < r_2$. Lemma 1 characterizes the equilibrium report time.

Lemma 1: *If $r_1 < r_2$, then an injurer who reports the act will do so immediately, $t = 0$. If $r_1 > r_2$, then an injurer who reports the act will delay reporting, $t = 1$. If $r_1 = r_2$, then an injurer who reports the act may do so at any time, $t \in [0, 1]$.*

Lemma 1 follows from the elimination of weakly-dominated strategies. Suppose that $r_1 < r_2$, so the first injurer to report the act receives a larger penalty reduction than the second injurer to report. Intuitively, $r_1 < r_2$ generates an incentive to minimize the time to report in order to secure the first position in the self-reporting queue, i.e., "a race to the courthouse." If injurer k ($k = 1, 2$) believes that injurer j ($j = 1, 2, j \neq k$) will not report at all, then injurer k is just as well off reporting immediately as delaying. However, if injurer k believes that there is a non-zero chance that injurer j will report at time $t = 0$, then injurer k is strictly better off reporting immediately as well. In other words, late reporting is a weakly-dominated strategy. If instead $r_1 > r_2$, then the second injurer to report receives a larger penalty reduction than the first injurer to report. In this case, early reporting is a weakly-dominated strategy.[27] Importantly, Lemma 1 implies that if both injurers report the harmful act, and if $r_1 \neq r_2$, then both injurers are equally likely to get the first position or the second position in the self-reporting queue.[28]

[27] If an injurer believes that there is a non-zero chance that the other injurer will report the act at $t = 1$, then the injurer strictly prefers to wait until $t = 1$ to report as well. If $r_1 = r_2$, then there is no advantage to being first or second, and the injurers are indifferent about the reporting time.

[28] When $r_1 < r_2$, self-reporting occurs immediately at $t = 0$, and when $r_1 > r_2$ self-reporting occurs at $t = 1$. By assumption, when the two injurers report at exactly the same time, an equally-weighted coin flip determines who obtains the first position in the self-reporting queue.

Figure 1: Strategic-Form Representation of the Self-Reporting Subgame (Expected Payoffs)

	No Report (NR)	Report (R)
No Report (NR)	$b - p_0 f$, $b - p_0 f$	$b - p_1 f$, $b - r_1 f$
Report (R)	$b - r_1 f$, $b - p_1 f$	$b - \left(\frac{r_1+r_2}{2}\right) f$, $b - \left(\frac{r_1+r_2}{2}\right) f$

Second, we study the injurers' decisions about whether to report the act. The strategic-form representation of the self-reporting subgame is presented in Figure 1. If neither injurer self-reports, then the act is detected with probability p_0 and each injurer receives a payoff of $b - p_0 f$. If one injurer self-reports but the other does not, then the injurer who self-reports pays $r_1 f$ with certainty and the silent accomplice pays $p_1 f$ in expectation giving payoffs $b - r_1 f$ and $b - p_1 f$, respectively. Finally, if both injurers self-report, then they are equally likely to get the first and second positions in the self-reporting queue. So, each injurer receives an expected payoff of $b - \left(\frac{r_1+r_2}{2}\right) f$.[29] Lemma 2 characterizes the pure-strategy Nash equilibria of the self-reporting subgame.

Lemma 2. *Take the benefit b, the fine f, and the detection probabilities, p_0 and p_1, as fixed. The pure-strategy Nash equilibria of the self-reporting subgame are as follows.*

1. *$r_1 \leq p_0$ and $\frac{r_1+r_2}{2} \leq p_1$: There is a unique pure-strategy Nash equilibrium where both injurers self-report, (R, R).*

2. *$r_1 \leq p_0$ and $\frac{r_1+r_2}{2} > p_1$: There are two pure-strategy Nash equilibria where exactly one injurer self-reports, (R, NR) and (NR, R).*

3. *$r_1 > p_0$ and $\frac{r_1+r_2}{2} \leq p_1$: There are two pure-strategy Nash equilibria, one where both injurers self-report and one where neither injurer self-reports. (R, R) Pareto dominates (NR, NR) if and only if $\frac{r_1+r_2}{2} \leq p_0$. (R, R) risk dominates (NR, NR) if and only if $\frac{3r_1+r_2}{4} \leq \frac{p_0+p_1}{2}$.*

4. *$r_1 > p_0$ and $\frac{r_1+r_2}{2} > p_1$: There is a unique pure-strategy Nash equilibrium where neither injurer self-reports, (NR, NR).*

In Case 1 of Lemma 2, self-reporting is a weakly-dominant strategy for both injurers. So, (R, R) is the unique Nash equilibrium that survives the elimination of weakly-dominated

[29] If $r_1 = r_2$, different reporting times would lead to the same expected payoffs.

strategies.[30] When the expected sanction for self-reporting is not too small, $\left(\frac{r_1+r_2}{2}\right)f > p_0 f$, then the injurers are jointly worse off self-reporting than they are remaining silent and the self-reporting subgame resembles a prisoners' dilemma environment.[31] In Case 2, there are two pure-strategy Nash equilibria, (R, NR) and (NR, R), where one injurer reports the act and the other does not.[32] In Case 3, both (NR, NR) and (R, R) are Nash equilibria. If one injurer believes that the other will remain silent then he will remain silent as well, since the expected penalty associated with remaining silent, $p_0 f$, is smaller than the penalty from being the only injurer to report, $r_1 f$. But if he believes that the other injurer will report, then he is better off reporting too since paying $\left(\frac{r_1+r_2}{2}\right)f$ on average is better than paying $p_1 f$. Thus, the self-reporting subgame in Case 3 is a coordination game. Finally, in Case 4, no-reporting is a strictly-dominant strategy for both injurers. So, (NR, NR) is the unique Nash equilibrium.

The set of Nash equilibria associated with Case 2, (R, NR) and (NR, R), cannot be narrowed with either the Pareto-dominance or the risk-dominance refinements (Harsanyi and Selten, 1988). In contrast, the two pure-strategy Nash equilibria that arise in Case 3, (R, R) and (NR, NR), may be ranked using standard equilibrium refinements. When $\frac{r_1+r_2}{2} \leq p_0$, the expected sanction is lower when both injurers report committing the act. So, (R, R) is the Pareto-dominant Nash equilibrium if and only if $\frac{r_1+r_2}{2} \leq p_0$. When $\frac{3r_1+r_2}{4} \leq \frac{p_0+p_1}{2}$, an injurer would prefer to self-report when there is a fifty-percent chance that the other injurer will also report. Thus, (R, R) is the risk-dominant Nash equilibrium if and only if $\frac{3r_1+r_2}{4} \leq \frac{p_0+p_1}{2}$.

3.2 Decision to Participate in the Act

In Stage 1, the potential injurers simultaneously and independently decide whether to participate in the harmful activity.[33] If *both* potential injurers decide to participate in the activity, then the act is committed. The payoff for each injurer is equal to the payoff that corresponds to the Nash equilibrium of the self-reporting subgame shown in Figure 1. If one or both po-

[30] More specifically, the second Nash equilibrium where both injurers decide not to report, (NR, NR) does not survive the elimination of weakly-dominated strategies.

[31] If $\left(\frac{r_1+r_2}{2}\right)f < p_0 f$, self-reporting is jointly efficient for the injurers and the game is not a prisoners' dilemma.

[32] Without loss of generality, we assume that, when indifferent, the injurers decide to self-report. This assumption allows us to eliminate the potential Nash equilibrium where both injurers decide not to report, (NR, NR).

[33] Our findings also hold in environments in which the injurers jointly decide whether to commit an act, but binding agreements between the injurers regarding their reporting choices in Stage 2 are not allowed.

tential injurers decide not to participate, then the act is not committed. The game ends and the payoff for each potential injurer is zero.

A potential injurer's decision about whether to participate in the harmful activity in Stage 1 depends on his private benefit from committing the act and the expected fine (which is determined in the Stage 2 continuation game). The b-value that equals the expected fine represents the "deterrence threshold" and is denoted by $\hat{b}$. When the individual benefit of committing the act, b, is greater than the deterrence threshold, $\hat{b}$, then participating in the activity is a weakly-dominant strategy. In particular, if injurer k ($k = 1, 2$) believes that injurer j ($j = 1, 2,\ j \neq k$) will participate in the activity with non-zero probability, then injurer k strictly prefers to participate in the act.[34] Conversely, when b is smaller than the deterrence threshold, $\hat{b}$, the injurer will choose not to participate in the activity.[35] Finally, when b is exactly equal to the deterrence threshold, $\hat{b}$, then the injurer is indifferent between participating and not participating in the act and, without loss of generality, we assume that the injurer does not participate in the act. The deterrence thresholds are constructed using Lemma 2 above. Lemma 3 characterizes the equilibrium decisions in Stage 1. Cases 1–4 correspond to Cases 1–4 included in Lemma 2.

Lemma 3. *Take the fine f, and the detection probabilities, p_0 and p_1, as fixed. Each potential injurer will decide to participate in the activity under the following conditions.*

1. *$r_1 \leq p_0$ and $\frac{r_1+r_2}{2} \leq p_1$: The injurer decides to participate if and only if $b > \hat{b} = \left(\frac{r_1+r_2}{2}\right)f$.*
2. *$r_1 \leq p_0$ and $\frac{r_1+r_2}{2} > p_1$: The injurer decides to participate if and only if $b > \hat{b} = \left(\frac{r_1+p_1}{2}\right)f$.*
3. *$r_1 > p_0$ and $\frac{r_1+r_2}{2} \leq p_1$: If $\frac{r_1+r_2}{2} \leq p_0$ (Pareto Dominance) or $\frac{3r_1+r_2}{4} \leq \frac{p_0+p_1}{2}$ (Risk Dominance), the injurer decides to participate if and only if $b > \hat{b} = \left(\frac{r_1+r_2}{2}\right)f$. If $\frac{r_1+r_2}{2} > p_0$ (Pareto Dominance) or $\frac{3r_1+r_2}{4} > \frac{p_0+p_1}{2}$ (Risk Dominance), the injurer decides to participate if and only if $b > \hat{b} = p_0 f$.*
4. *$r_1 > p_0$ and $\frac{r_1+r_2}{2} > p_1$: The injurer decides to participate if and only if $b > \hat{b} = p_0 f$.*

In Case 1, since both injurers self-report in the unique Nash equilibrium, the deterrence threshold is $\hat{b} = \left(\frac{r_1+r_2}{2}\right)f$. Then, each potential injurer will participate in the harmful act

[34] When b is greater than the deterrence threshold, then not participating is a weakly-dominated strategy.

[35] Participating is a weakly-dominated strategy in this scenario. If injurer k believes that there is a non-zero chance that injurer j will participate in the act, then injurer k strictly prefers not to participate.

when $b > \hat{b} = \left(\frac{r_1+r_2}{2}\right)f$. In Case 2, where multiple equilibria arise, (R, NR) and (NR, R), our refinements do not eliminate either one and we assume that the deterrence threshold is the expected fine, $\hat{b} = \left(\frac{r_1+p_1}{2}\right)f$.[36] Then, each potential injurer will participate in the harmful activity when $b > \hat{b} = \left(\frac{r_1+p_1}{2}\right)f$. In Case 3, the equilibrium refinement will determine which of the two outcomes is obtained, (R, R) and (NR, NR), and so the deterrence threshold is either $\hat{b} = \left(\frac{r_1+r_2}{2}\right)f$ or $\hat{b} = p_0 f$. Then, each potential injurer will participate when $b > \hat{b} = \left(\frac{r_1+r_2}{2}\right)f$ or $b > \hat{b} = p_0 f$, depending on the equilibrium. Finally, in Case 4, since neither injurer self-reports in equilibrium, the deterrence threshold is $\hat{b} = p_0 f$. Then, each injurer will participate when $b > \hat{b} = p_0 f$.

Our results suggest that ordered-leniency policies have the potential to create significant social-welfare benefits. Without any opportunities to self-report, the likelihood of detection of an injurer is p_0 and the expected fine for each injurer is capped at $p_0\bar{f}$. Through a leniency program that grants a reduced fine to the first injurer to report the harmful act, $r_1 = p_0 - \varepsilon$ ($\varepsilon > 0$) for example, the enforcement agency can induce at least one of the two injurers to come forward and report the act, and hence increase the likelihood of detection without raising investigatory costs. In particular, when one injurer self-reports, the likelihood of detection of the silent accomplice rises from p_0 to p_1. When both injurers self-report, socially-harmful acts are detected with certainty. With a well-designed enforcement policy with ordered leniency, the enforcement agency can exploit negative externalities between the injurers in the self-reporting subgame to deter a broader range of harmful acts.

4 Optimal Enforcement Policies

This section characterizes the optimal enforcement policies with and without leniency. First, we identify the optimal enforcement policy in the absence of leniency for self-reporting and

[36] Given that neither the Pareto-dominance nor risk-dominance refinements reduce the set of equilibrium outcomes, it is reasonable to assume that neither the enforcement agency nor the players themselves can predict which outcome will occur, and hence, they assign an equal weight to each outcome. This assumption is intuitive and empirically relevant but much stronger than necessary. All that is required for the results that follow is that the deterrence threshold in Case 2 is strictly smaller than $p_1 f$. This would be true if the players, at the time that they are committing the act, put a non-zero chance on both (R, NR) and (NR, R). We will see that the enforcement agency can implement a deterrence threshold of $p_1 f$ in a setting where self-reporting is a dominant strategy for both players as in Case 1. Thus, for several reasons, the enforcement agency would eschew enforcement policies associated with Case 2.

show that it involves positive enforcement costs, maximal fines, and underdeterrence relative to the first-best level. Second, we consider enforcement policies with ordered leniency. We prove that policies that offer leniency for self-reporting are superior to the optimal enforcement policy without leniency. Holding the enforcement costs fixed, deterrence can be improved with ordered leniency for self-reporting. We then highlight several key features of optimal ordered-leniency policies. Finally, we demonstrate that the first-best outcome can be achieved with an ordered-leniency policy when the externality from the harmful activities, h, is not too high.

4.1 Optimal Enforcement Policy without Leniency

Consider an environment where leniency for self-reporting is not granted, so the leniency multipliers are $(r_1, r_2) = (1, 1)$.[37] According to Lemma 2 (Case 4), there is a unique pure-strategy Nash equilibrium where neither injurer self-reports.[38] The probability that the injurers are detected and fined is $p_0 = e$. Then, each injurer faces an expected fine ef, and so each will commit the act if and only if $b > \hat{b} = ef$ (Lemma 3, Case 4). Social welfare is the aggregation of the benefits to the individuals who commit the act minus the social costs associated with the act (the harm inflicted on others, h, and the cost of enforcement $c(e)$).[39] Normalizing the size of the population of injurers to unity, the social welfare function can be written as:

$$W = \int_{ef}^{\infty} (b - h)\, g(b) db - c(e). \tag{1}$$

Next, we identify the optimal fine, f, and the optimal detection probability (optimal enforcement effort), e,[40] that maximize social welfare in the no-leniency environment. Consider first the optimal fine f. It is easy to show that the optimal fine will be maximal, $f = \bar{f}$. To see why, suppose that the optimal $e > 0$ and that the optimal fine is less than maximal, $f < \bar{f}$. By raising the fine slightly while at the same time lowering the probability of detection so as to keep the product ef constant, the same level of deterrence can achieved but at a lower cost than $c(e)$.

[37] The environment without leniency is a special case of enforcement with ordered leniency for self-reporting.

[38] If an injurer reports, he pays f (irrespective of the decision of his accomplice); if an injurer remains silent, he pays $p_0 f$ (if his accomplice does not report the act) or $p_1 f$ (if his accomplice reports the act). Then, self-reporting is a strictly-dominated strategy.

[39] The fines are simply transfers from the injurers to the enforcement agency, and therefore are not included in the social welfare function.

[40] By assumption, the detection probability when no injurer self-reports is $p_0 = e$.

Consider now the optimal detection probability (optimal enforcement effort), e. Substitute $\bar{f}$ into the social welfare function and differentiate it with respect to e. The first-order condition is given by:

$$(h - e\bar{f})\bar{f}g(e\bar{f}) - c'(e) = 0. \tag{2}$$

The first term is the incremental social benefit of increased deterrence. When the probability of detection is raised, the acts that were previously exactly on the margin between committing and not committing the act (those with private benefits $b = e\bar{f}$) are now deterred. The social benefit of deterring these marginal acts is $h - e\bar{f}$.[41] The volume of additional cases that are deterred when e is raised is $\bar{f}g(e\bar{f})$, which depends upon the height of the probability density function when evaluated at $e\bar{f}$. The second term, $c'(e)$, is the incremental social cost associated with the higher detection probability. It is easy to verify that the optimal e will be always positive. Taking the fine $\bar{f}$ as fixed and starting with $e = 0$, the incremental social value of raising the probability is positive (since harmful acts with very small benefits will no longer be committed) while the incremental social cost is negligible since, by assumption, $c'(0) = 0$. Hence, the enforcement cost, $c(e)$, will be also positive.

Using equation (2) and rearranging terms, we find that under an enforcement policy with no-leniency, the optimal deterrence threshold (optimal expected fine), $\hat{b}$, satisfies:

$$\hat{b} = e\bar{f} = h - \frac{c'(e)}{\bar{f}g(e\bar{f})}. \tag{3}$$

There may be multiple solutions to this equation. However, under our assumptions on $c(e)$, all of the solutions involve $e > 0$. Then, the optimal enforcement policy has a deterrence threshold $\hat{b} \in (0, h)$.[42] It is interesting to compare the optimal enforcement policy without leniency to a social-welfare benchmark. Without leniency for self-reporting, the first-best outcome is not achievable. Since $\hat{b} < h$, the optimal enforcement policy without leniency has positive enforcement costs and a deterrence threshold that is strictly smaller than the first-best level. Proposition 1 outlines our findings.

Proposition 1. *An enforcement policy without leniency for self-reporting cannot implement the first-best outcome. The optimal enforcement policy has a maximal fine, a positive enforcement cost, and underdeterrence.*

[41] If $h - e\bar{f} < 0$, there will be a destruction of social value when deterring the marginal act.

[42] Our results regarding optimal enforcement without leniency policies for groups of injurers are aligned with Kaplow and Shavell's (1994) finding on enforcement without self-reporting in single-injurer environments.

As has been emphasized in the literature on control of harmful externalities (Polinsky and Shavell, 1984), the failure to implement the first-best outcome with an enforcement policy without leniency for self-reporting is a consequence of having a maximal fine, $\bar{f}$.[43] If the fine was not bounded, the enforcement agency could get arbitrarily close to the first-best outcome with an extremely high fine coupled with an arbitrarily small probability of detection (Becker, 1968).

4.2 Optimal Enforcement Policy with Ordered Leniency

This section characterizes the optimal enforcement policy with ordered leniency. First, we show that for any given fine, f, there exists an enforcement policy with ordered leniency that is strictly superior the optimal enforcement policy without leniency described in the previous section. Second, we take the agency's enforcement effort, e, and the corresponding probabilities of detection, p_0 and p_1, as fixed and identify the fine, f, and the leniency multipliers, r_1 and r_2, that generate maximal deterrence (i.e., the highest expected fine). Third, we demonstrate that the first-best outcome may be achieved with ordered-leniency policies at a zero cost when the externalities associated with the harmful activities are not too high.

4.2.1 Superiority of Ordered Leniency

We will show that enforcement policies with ordered leniency for self-reporting always outperform enforcement policies without leniency for self-reporting. As demonstrated in the previous section, without leniency, the optimal enforcement policy has strictly positive enforcement costs, maximal fines, and underdeterrence of harmful activities relative to the first-best level. With ordered leniency, and holding enforcement efforts fixed, the enforcement agency can raise the expected fines and achieve a higher level of deterrence.

Proposition 2. *There exists an enforcement policy with ordered leniency for self-reporting that is strictly superior to the optimal enforcement policy without leniency for self-reporting.*

The proof of Proposition 2, which is omitted, is straightforward. Intuitive explanation follows. When there is no leniency for self-reporting, $(r_1, r_2) = (1, 1)$, the injurers do not

[43] Intuitively, having a maximal fine implies that increasing deterrence is expensive. When the benefit to the injurer, b, is very close to social harm, h, then the social benefit of increasing the expected fine is very small (because $b - h$ is negative but small). Since $c'(e) > 0$, increasing the fine leads to a first-order increase in costs.

self-report and the deterrence threshold is $\hat{b} = p_0\bar{f} < h$. There is underdeterrence relative to the first-best level, and too many harmful acts are committed. Consider now an ordered-leniency policy $(r_1, r_2) = (p_0 - v, p_0 + 2v)$, where $0 < v < p_0$ is a small positive number. With these leniency multipliers, self-reporting is a strictly dominant strategy for both injurers (Case 1 of Lemma 3). Moreover, the expected fine with ordered leniency is higher than the optimal expected fine without leniency, $p_0\bar{f} < (p_0 + v/2)\bar{f} < h$. Holding the level of enforcement effort and the probabilities of detection fixed, the ordered leniency policy $(r_1, r_2) = (p_0 - v, p_0 + 2v)$ raises the deterrence threshold closer to the first-best level and increases social welfare. More generally, given any optimal enforcement policy without leniency, one can always construct an enforcement policy with ordered leniency that is strictly superior: By exploiting the negative externalities between the injurers at the self-reporting subgame, ordered-leniency mechanisms always achieve higher levels of deterrence. Our findings regarding the superiority of enforcement policies with ordered leniency for groups of injurers complement Kaplow and Shavell's (1994) results for single-injurer environments.

4.2.2 Maximal Deterrence with Ordered Leniency

Taking the enforcement effort, e, and the corresponding probabilities of detection, p_0 and p_1, as fixed, we now characterize the fine, f, and leniency multipliers, (r_1, r_2), that create the highest possible deterrence (i.e., highest expected fine). We will demonstrate that the fine should be set at the maximal level, $\bar{f}$, and that the ordered-leniency policies that implement maximal deterrence give greater leniency to the first injurer to report and induce immediate self-reporting by both injurers. Importantly, we will show that the optimal leniency multipliers will be different for the Pareto-dominance and risk-dominance refinements. Leniency will be stronger (smaller multipliers) under the Pareto-dominance refinement, and leniency will be milder (larger multipliers) under the risk-dominance refinement.

Denote (r_1^S, r_2^S) and (r_1^M, r_2^M) as the leniency multipliers for the Pareto- and risk-dominance refinements, respectively, and $\hat{b}^S$ and $\hat{b}^M$ as the corresponding deterrence thresholds (expected fines). The superscript S refers to "Strong Leniency" and the superscript M refers to "Mild Leniency." Proposition 3 characterizes the fine and leniency multipliers that create maximal deterrence for groups of potential injurers.

Proposition 3. *Take the enforcement effort e as fixed. Maximal deterrence is obtained with*

a maximal fine, $f = \bar{f}$, and the following leniency multipliers:[44]

1. *If $p_1 \leq \frac{1+p_0}{2}$, then $(r_1^S, r_2^S) = (r_1^M, r_2^M) = (p_1 - \Delta, p_1 + \Delta)$ where $\Delta \in [p_1 - p_0, \min\{p_1, 1 - p_1\}]$. The injurers commit the act and self-report at time $t = 0$ if $b > \hat{b}^S = \hat{b}^M = p_1 \bar{f}$, and do not commit the act otherwise.*

2. *If $p_1 > \frac{1+p_0}{2}$, then $(r_1^S, r_2^S) = (p_0, 1)$ and $(r_1^M, r_2^M) = \left(\frac{2(p_0+p_1)-1}{3}, 1\right)$. The injurers commit the act and self-report at time $t = 0$ if $b > \hat{b}^S = \left(\frac{1+p_0}{2}\right) \bar{f}$ (Pareto Dominance) and $b > \hat{b}^M = \left(\frac{1+p_0+p_1}{3}\right) \bar{f}$ (Risk Dominance), where $\hat{b}^S < \hat{b}^M$, and do not commit the act otherwise.*

Proposition 3 provides fundamental implications for the optimal design of enforcement policies with ordered leniency. The formal analysis is presented in the Appendix. An intuitive discussion of the main insights follows.

Remark 1. *The Fine Is Maximal.*

The highest deterrence is obtained by imposing the maximal fine, $f = \bar{f}$. This follows from the fact that the equilibria of the self-reporting subgame described in Lemmas 2 and 3 do not depend on the level of the fine, f.

Remark 2. *Both Injurers Self-Report.*

Maximal deterrence is achieved when both injurers self-report. It is obvious that a leniency policy where at least one injurer self-reports creates stronger deterrence than a policy where no injurer self-reports. By offering $(r_1, r_2) = (p_0, 1)$, at least one injurer self-reports and the expected fine rises above $p_0 \bar{f}$ (the expected fine if neither reports). More specifically, if $p_1 \geq \frac{1+p_0}{2}$, then we are in Case 1 of Lemma 2 where both injurers self-report, and the expected fine is $\left(\frac{1+p_0}{2}\right) \bar{f} > p_0 \bar{f}$. On the other hand, if $p_1 < \frac{1+p_0}{2}$, then we are in Case 2 of Lemmas 2 and 3 where exactly one injurer self-reports and the expected fine is $\left(\frac{p_0+p_1}{2}\right) \bar{f} > p_0 \bar{f}$. In this latter case, where only one injurer self-reports, deterrence will be even stronger if leniency is granted to the second injurer as well. When $(r_1, r_2) = (p_0, 2p_1 - p_0)$, there is a race to the courthouse where both injurers self-report, and the expected fine rises to $p_1 \bar{f}$.[45]

[44] When $p_1 \leq \frac{1+p_0}{2}$, the leniency multipliers are not unique.

[45] According to Proposition 3 Case 2, this is an optimal policy ($\Delta = p_1 - p_0$).

Remark 3. *The First Injurer to Self-Report Always Receives More Lenient Treatment.*

Suppose that $p_1 \geq \frac{1+p_0}{2}$ and $(r_1, r_2) = (p_0, 1)$. We are in Case 1 of Lemma 2, where both injurers self-report. Rewarding the first injurer creates a proverbial race to the courthouse between the two injurers, and the expected fine is $\left(\frac{1+p_0}{2}\right)\bar{f} > p_0\bar{f}$.[46]. If the multipliers were reversed, so $(r_1, r_2) = (1, p_0)$ (i.e., the second to report gets the more lenient treatment), then neither injurer would self-report and the expected fine would be $p_0\bar{f}$, the same as in the absence of a leniency policy.[47] Giving more leniency to the first injurer to report the act increases deterrence.

Remark 4. *The Second Injurer to Self-Report May Also Receive Leniency.*

When the strength of the inculpatory evidence is weak then the second injurer to report the act receives leniency, too. To see why, suppose that $p_1 < \frac{1+p_0}{2}$. If leniency is granted only to the first injurer, $(r_1, r_2) = (p_0, 1)$, we are in Case 2 of Lemmas 2 and 3 where only one injurer reports the act and the other remains silent, and the deterrence threshold is $\left(\frac{p_0+p_1}{2}\right)\bar{f}$. Now suppose instead that the agency gives partial leniency to the second injurer too, $(r_1, r_2) = (p_0, 2p_1 - p_0)$. With these leniency multipliers, there is a race to the courthouse, both injurers self-report, and the deterrence threshold rises to $p_1\bar{f}$.[48] Deterrence is stronger when the second injurer also receives leniency.[49]

Remark 5. *Stronger Deterrence Is Obtained with the Risk-Dominance Refinement.*

[46] If $p_1 \geq \frac{1+p_0}{2}$ then only one injurer would self-report, and the expected fine is still strictly higher than $p_0\bar{f}$

[47] More generally, given an ordered-leniency policy with $r_1 > r_2$, there exists an ordered-leniency policy with $r_1' < r_2'$ that creates stronger deterrence.

[48] When $p_1 > 1/2$, maximal deterrence can be achieved by granting leniency to just the first injurer to report, $(r_1, r_2) = (2p_1 - 1, 1)$. With these multipliers, both injurers self-report and the expected fine is $p_1\bar{f}$. When $p_1 < 1/2$, however, $2p_1 - 1$ is a negative number. Some degree of leniency must be granted to the second injurer, too.

[49] Note that, when viewed from an ex post perspective, the second injurer is worse off when he self-reports. Since $r_2^i > p_1$ for $i = S, M$, the second injurer would be better off remaining silent and paying $p_1\bar{f}$ in expectation than self-reporting and paying $r_2^i\bar{f}$. The reason why the second injurer is willing to self-report is because when the injurer is making the important decision about whether or not to self-report, the injurer does not know whether he will obtain the first position or the second position in the self-reporting queue. Hence, a race-to-the-courthouse effect will be always observed in equilibrium when ordered-leniency policies are implemented.

Proposition 3 implies that the deterrence threshold is never lower, and may be higher, when the risk-dominance refinement is applied in the self-reporting subgame.[50] In the first part of Proposition 3, when $p_1 \leq \frac{1+p_0}{2}$ (weak inculpatory evidence), leniency multipliers are the same under the Pareto-dominance and risk-dominance refinements, and so the two equilibrium refinements lead to the same deterrence threshold, $\hat{b}^S = \hat{b}^M = p_1\bar{f}$. In the second part of Proposition 3, when $p_1 > \frac{1+p_0}{2}$ (strong inculpatory evidence), the optimal leniency multipliers under the two equilibrium refinements diverge. Suppose that the enforcement agency chooses the mild leniency policy, $(r_1^M, r_2^M) = \left(\frac{2(p_0+p_1)-1}{3}, 1\right)$.[51] Notice that $r_1^M > p_0$, so neither self-reporting nor no-reporting are dominant strategies. When the risk-dominance refinement is applied in the self-reporting subgame, both injurers self-report and the deterrence threshold is $\hat{b}^M = \left(\frac{1+p_0+p_1}{3}\right)\bar{f} > p_0\bar{f}$. When the Pareto-dominance refinement is applied in the self-reporting subgame, neither injurer self-reports and the deterrence threshold is $p_0\bar{f}$. Then, when the Pareto-dominance refinement is applied in the self-reporting subgame, the enforcement agency must lower the multipliers to $(r_1^S, r_2^S) = (p_0, 1)$ to transform the self-reporting subgame into a prisoner's dilemma.[52] The resulting deterrence threshold is $\hat{b}^S = \left(\frac{1+p_0}{2}\right)\bar{f} < \hat{b}^M$. Hence, when Pareto dominance is applied in the self-reporting subgame, the deterrence threshold is smaller and the incentives to engage in the harmful activity rise.

Next, we provide a numerical example to illustrate the main insights regarding the design of ordered-leniency policies that generate maximal deterrence.

Example 1. Suppose that the maximal fine is $\bar{f} = 1$ and that $p_0 = .2$. Without leniency for self-reporting, neither injurer self-reports and the expected fine is $\hat{b} = p_0\bar{f} = .2$.

According to Proposition 3, the design of the ordered-leniency policy depends on the value of p_1, the probability of catching and sanctioning a silent injurer if the other injurer has self-reported. Suppose that $p_1 = .4 < \frac{1+p_0}{2}$, so the likelihood of catching a silent conspirator

[50] As demonstrated in the Appendix (proof of Proposition 3), the leniency multipliers under Pareto dominance, (r_1^S, r_2^S), satisfy the conditions stated in Case 1 of Lemma 2. When $p_1 \leq \frac{1+p_0}{2}$, the leniency multipliers under risk dominance, (r_1^M, r_2^M), satisfy either the conditions stated in Case 1 of Lemma 2 or the conditions stated in Case 3 of Lemma 2 (both provide the same level of deterrence); when $p_1 > \frac{1+p_0}{2}$, the leniency multiplier under risk dominance, (r_1^M, r_2^M), satisfy the conditions stated in Case 3 of Lemma 2.

[51] Under these leniency multipliers, the environment corresponds to Case 3 of Lemma 2, where the self-reporting subgame is a coordination game with two Nash equilibria (R, R) and (NR, NR). When risk-dominance is applied, maximal deterrence is achieved.

[52] This new strategic environment corresponds to Case 1 of Lemma 2, where (R, R) is the unique Nash equilibrium.

is relatively low. Granting leniency to the second injurer who reports the act is necessary. Proposition 3 implies that deterrence is maximal when the enforcement agency grants leniency $r_1^S = r_1^M = p_0 = .2$ and $r_2^S = r_2^M = \frac{2p_1 - p_0}{2} = .6$ to the first and second injurer to report.[53] The injurers race to be the first in line and the expected sanction rises to $\hat{b}^S = \hat{b}^M = p_1 \bar{f} = .4$.

Suppose instead that $p_1 = .75 > \frac{1+p_0}{2}$, so the chance of catching a silent conspirator is relatively high. Now, granting leniency to the second injurer who reports the act is unnecessary. When Pareto dominance is applied in the self-reporting subgame, the enforcement agency grants leniency $r_1^S = p_0 = .2$ to the first injurer who self-reports but holds the second injurer fully accountable, $r_2^S = 1$. Leniency for the first injurer alone creates a race between the two injurers to secure the first position in the self-reporting queue. Self-reporting is a dominant strategy for both injurers, and the self-reporting stage involves a prisoner's dilemma game. Both injurers self-report immediately and the expected fine is $\hat{b}^S = \left(\frac{1+p_0}{2}\right)\bar{f} = .6$.

When $p_1 = .75$ and risk dominance is applied in the self-reporting subgame, deterrence can be made even higher by raising the leniency multiplier for the first injurer to $r_1^M = \frac{2(p_0+p_1)-1}{3} = .3$. Self-reporting is clearly not a dominant strategy in this case, and the self-reporting stage involves a coordination game. Nevertheless, with the risk-dominance refinement, both injurers self-report immediately and the expected fine rises to $\hat{b}^M = \left(\frac{1+p_0+p_1}{3}\right)\bar{f} = .65$.[54]

4.2.3 Optimal Enforcement Effort with Ordered Leniency

This section characterizes the optimal enforcement effort e when ordered-leniency policies that generate maximal deterrence are implemented. Remember that Proposition 3 identifies the leniency multipliers and fine that create maximal deterrence (i.e., the highest expected fine), and that superscripts S and M denote the leniency policies under the Pareto- and risk-dominance refinements, respectively.

The next lemma, which follows from Proposition 3, will be used in the analysis of the optimal enforcement effort e when ordered-leniency policies are implemented. Recall that $p_0 = e$ and $p_1 = e + (1-e)\pi$, where $\pi \in (0,1)$ represents the exogenous strength of inculpatory evidence. Then, $p_1 \leq \frac{1+p_0}{2}$ holds if and only if $\pi \leq \frac{1}{2}$, and $p_1 > \frac{1+p_0}{2}$ holds if and only if $\pi > \frac{1}{2}$.

[53] When $p_1 \leq \frac{1+p_0}{2}$, the leniency multipliers that create maximal deterrence are not unique but are similarly defined under the Pareto- and risk-dominance refinements.

[54] Although no-reporting by both injurers is the Pareto-dominant Nash equilibrium, self-reporting by both injurers is the risk-dominant Nash equilibrium.

In other words, Cases 1 and 2 of Lemma 4 correspond to Cases 1 and 2 of Proposition 3.[55]

Lemma 4. *The ordered-leniency multipliers (r_1^S, r_2^S) and (r_1^M, r_2^M), characterized in Proposition 3, yield corresponding expected fines $\hat{b}^S(e,\pi)$ and $\hat{b}^M(e,\pi)$ for the injurers. These functions, which are continuous and piecewise differentiable, satisfy:*

1. *If $\pi \leq \frac{1}{2}$, then $\hat{b}^S(e,\pi) = \hat{b}^M(e,\pi) = [\pi + (1-\pi)e]\,\bar{f}$ and $0 < \frac{\partial \hat{b}^i(e,\pi)}{\partial e} < \bar{f}$ for $i = S, M$.*
2. *If $\pi > \frac{1}{2}$, then $\hat{b}^S(e,\pi) = \left(\frac{1+e}{2}\right)\bar{f}$ and $\hat{b}^M(e,\pi) = \left[\frac{(1+\pi)+(2-\pi)e}{3}\right]\bar{f}$. Furthermore, $\hat{b}^S(e,\pi) < \hat{b}^M(e,\pi)$ and $0 < \frac{\partial \hat{b}^M(e,\pi)}{\partial e} < \frac{\partial \hat{b}^S(e,\pi)}{\partial e} < \bar{f}$.*

We now describe the circumstances under which optimal ordered-leniency policies can achieve the first-best outcome. Recall that, in the first-best outcome, the injurers commit the act if and only if the benefit exceeds the social harm, $b > h$ and no effort is spent on enforcement, $e = 0$. In this benchmark, $p_0 = 0$ and $p_1 = \pi$.

When $\pi \leq \frac{1}{2}$ (weak inculpatory evidence), we are in Case 1 of Lemma 4. With no enforcement effort, $e = 0$, the maximal deterrence is obtained with a maximal fine $\bar{f}$ and leniency multipliers $(r_1^S, r_2^S) = (r_1^M, r_2^M) = (0, 2\pi)$. With these multipliers, the injurers are deterred from committing the act when $b \leq \hat{b}^S = \hat{b}^M = \pi\bar{f}$. Note that if the level of harm is less than the deterrence threshold, $h < \pi\bar{f}$, then there would be overdeterrence relative to the first-best level. However, this may be easily solved by reducing the fine below its maximal level, granting additional leniency to the injurers, or both. When the expected fine is exactly equal to the social harm, h, then the injurers will commit the act if and only if $b > h$, as desired. When the level of harm exceeds the deterrence threshold, $h > \pi\bar{f}$, then there is underdeterrence relative to the first-best level. In this case, deterrence can be improved by spending resources on enforcement. Taken together, when $\pi \leq \frac{1}{2}$, the first-best outcome is achieved at zero cost if and only if the harm is not too high, $h \leq \pi\bar{f}$.

When $\pi > \frac{1}{2}$ (strong inculpatory evidence), we are in Case 2 of Lemma 4. Suppose the enforcement efforts are zero, $e = 0$. If the Pareto-dominance refinement is applied to the self-reporting subgame, then the multipliers that create maximal deterrence are $(r_1^S, r_2^S) = (0, 1)$ and the associated deterrence threshold is $\hat{b}^S = \left(\frac{1}{2}\right)\bar{f}$. If the level of harm is below this threshold, $h < \left(\frac{1}{2}\right)\bar{f}$, then the first-best outcome may be obtained by lowering the fine, lowering the

[55] Consider Case 1 of Proposition 3, where $p_1 \leq \frac{1+p_0}{2}$. Substituting $p_0 = e$ and $p_1 = e + (1-e)\pi$, we find that $p_1 \leq \frac{1+p_0}{2}$ holds if and only if $\pi \leq \frac{1}{2}$. Similarly logic applies to Case 2 of Proposition 3.

leniency multiplier for the second injurer, or both. If the risk-dominance refinement applies, then the leniency multipliers that create the maximal deterrence are $(r_1^M, r_2^M) = \left(\frac{2\pi-1}{3}, 1\right)$ and the associated deterrence threshold is $\hat{b}^M = \left(\frac{1+\pi}{3}\right)\bar{f}$. Applying the same logic as before, when $h < \left(\frac{1+\pi}{3}\right)\bar{f}$, the first-best outcome can be obtained by lowering the fine, lowering the leniency multipliers, or both. Hence, when $\pi > \frac{1}{2}$, the first-best outcome is achieved at zero cost if and only if the harm is not too high, $h \leq \left(\frac{1}{2}\right)\bar{f}$ (Pareto Dominance) and $h \leq \left(\frac{1+\pi}{3}\right)\bar{f}$ (Risk Dominance).

Proposition 4 establishes the necessary and sufficient conditions under which the enforcement agency can implement the first-best outcome with an ordered-leniency policy at a zero cost, and describes the second-best enforcement policy when the first-best outcome cannot be achieved.

Proposition 4. *An optimal enforcement policy with ordered leniency for self-reporting can implement the first-best outcome at zero cost if and only if* $h \leq \hat{b}^S(0,\pi) = \min\{\pi, \frac{1}{2}\}\bar{f}$ *under the Pareto-dominance refinement, and* $h \leq \hat{b}^M(0,\pi) = \min\{\pi, \frac{1+\pi}{3}\}\bar{f}$ *under the risk-dominance refinement. When* $h > \hat{b}^i(0,\pi), i = S, M$*, the second-best enforcement policy involves a maximal fine, positive enforcement costs, and underdeterrence relative to the first best.*

Intuitively, when the externalities associated with the harmful activities, h, are not too high, the optimal ordered-leniency policy allows the enforcement agency to achieve the first-best outcome without spending resources on enforcement, $c(e) = 0$. When the externalities associated with the harmful activities, h, are relatively high, the first-best outcome cannot be not obtained. The enforcement agency must spend resources to detect the harmful activities and too many harmful activities will be committed.

Taken together, our previous findings provide a social welfare rationale for the current use of ordered-leniency policies in the real-world. First, holding the enforcement costs fixed, we proved that an enforcement policy with ordered leniency is strictly superior to the optimal enforcement policy without leniency (Proposition 2). Second, we showed that ordered-leniency policies that generate maximal deterrence give successively larger discounts to injurers who secure higher positions in the self-reporting queue, creating a so-called "race to the courthouse" where all injurers report the act immediately (Proposition 3). Third, we demonstrated that socially-optimal level of deterrence can be obtained at zero cost when the externalities associated with the harmful activities are not too high (Proposition 4).

5 Groups with More than Two Members

This section extends our benchmark framework by studying an environment where the groups of injurers may have more than two members. Our analysis demonstrates that the key insights of the benchmark model extend to this setting and offers new important insights.

The strategic environment now consists of a game of complete information with the following risk-neutral players: $n \geq 2$ identical representative potential injurers and an enforcement agency. The potential injurers seek to maximize their private net benefits from committing a harmful act. The enforcement agency seeks to maximize social welfare, which includes the aggregation of the benefits to the injurers and the social costs (the harm inflicted on others and the cost of enforcement).

First, the enforcement agency publicly commits to an enforcement policy $(f, \boldsymbol{r}, e)$. (1) $f \in (0, \bar{f}]$ denotes the fine. As before, $\bar{f}$ can be greater than, lower than, or equal to the harm inflicted on others (measured per injurer), $h > 0$. (2) $\boldsymbol{r} = \{r_i\}_{i=1}^n$ denotes the vector of leniency multipliers that assigns leniency multiplier $r_i \in [0, 1]$ to position i in the self-reporting queue. The sequence $\{r_i\}_{i=1}^n$ may be either weakly increasing or weakly decreasing in i.[56] (3) $e \in [0, 1)$ is the enforcement agency's effort, and $c(e)$ is the cost of enforcement (measured per injurer).[57]

Second, after observing the enforcement policy, the potential injurers play a two-stage game. In Stage 1, the potential injurers simultaneously and independently decide whether to participate in the act. The benefit for each injurer, $b \in [0, \infty)$,[58] is revealed to the injurers before they make their decisions regarding committing the act. The act is committed if and only if all n potential injurers agree to participate. If the act is committed, Stage 2 starts; otherwise the game ends. In Stage 2, the injurers simultaneously and independently decide whether to self-report and the time of reporting, $t \in [0, 1]$. If injurers report at exactly the same time, then they are randomly assigned to the highest available positions in the self-reporting queue.

Third, the injurers, if detected, are sanctioned. We let p_i for $i = 0, 1, ..., n$ be the probability that a silent injurer will be detected and sanctioned when exactly i injurers self-report. We

[56] We later show that the optimal ordered-leniency policy involves a weakly-increasing sequence of leniency multipliers: Injurers who self-report early receive lighter sanctions than those who report late.

[57] The previously mentioned assumptions about $c(e)$ apply.

[58] As before, $g(b)$ and $G(b)$ denote the probability density function and cumulative distribution function.

assume that $0 \leq p_0 < p_1 < .... < p_{n-1} < 1$, so self-reporting by an injurer raises the probability that the silent injurers will be apprehended, and that the sequence $\{ip_{i-1}\}_{i=1}^{n}$ is convex in i.[59] These probabilities may depend on the agency's effort, $e \in [0, 1)$, and on the exogenous strength of the inculpatory evidence provided by the injurers who self-report, $\pi \in (0, 1)$. Specifically, we let $p_0(e) = e$ and $p_i(e, \pi) = e + (1 - e)[1 - (1 - \pi)^i]$ for $i = 1, ..., n - 1$.[60]

The equilibrium concept is subgame-perfect Nash equilibrium. As in our benchmark model, multiple equilibria may arise in the self-reporting subgame. We restrict attention to coalition-proof Nash equilibria – CPNE (Bernheim et al., 1987).[61]

The first-best outcome is used as a benchmark in the welfare analysis of ordered-leniency policies. In the first best, the cost of effort is zero and acts are committed if and only if $b > h$.

We apply backward induction and begin with the analysis of the injurers' decisions. We then study the optimal enforcement policy with ordered leniency and conduct social welfare analysis.

5.1 Injurers' Decisions

We begin by characterizing the equilibrium behavior of the injurers in Stage 2, the self-reporting stage. Next, we study the potential injurers' decisions regarding participating in the harmful act in Stage 1.

We first analyze the length of time taken by the injurers to report the act. Lemma 5

[59] This is equivalent to assuming that $ip_{i-1} - (i-1)p_{i-2}$ is increasing in i, and holds so long as the sequence $\{p_i\}_{i=0}^{n-1}$ is not too concave. It is satisfied when the sequence of probabilities is linear in i, and also when $p_i = \frac{i}{i+1}$. Convexity simplifies the characterization of the optimal ordered-leniency policy in Proposition 5.

[60] As in the benchmark model, this specification may be derived from first principles. Absent self-reporting by any injurer, detection is the outcome of a single Bernoulli trial with success probability $p_0 = e$. When i injurers self-report, there are i independent Bernoulli trials each of which uncovers incriminating evidence with probability π. So $[1 - (1 - \pi)^i]$ is the probability at least one of the i Bernoulli trials uncovers the evidence. One can verify that the sequence $\{ip_{i-1}\}_{i=1}^{n}$ is convex so long as n is not too large.

[61] An outcome is self-enforcing if and only if no proper subset (coalition) of players can deviate in a way that makes all of its members better off. The CPNE refinement captures the concept of efficient self-enforcing outcomes for environments with more than two players: An outcome is a CPNE if and only if it is Pareto efficient within the class of self-enforcing outcomes. Finally note that the application of the Pareto- or risk-dominance refinements in two-player games with no communication implicitly assumes that the players agree on the refinement. The application of the CPNE refinement here follows a similar approach, and hence, communication is not required.

characterizes the equilibrium report time.

Lemma 5. *If $\{r_i\}_{i=1}^n$ is weakly increasing in i with $r_i < r_{i+1}$ for some i, then an injurer who reports the act will do so immediately, $t = 0$. If $\{r_i\}_{i=1}^n$ is weakly decreasing in i with $r_i > r_{i+1}$ for some i, then an injurer who reports the act will delay reporting, $t = 1$. If $\{r_i\}_{i=1}^n$ is constant, then an injurer who reports the act may do so at any time, $t \in [0, 1]$.*

The proof of Lemma 5, which follows from the elimination of weakly-dominated strategies, is analogous to the proof of Lemma 1 and is omitted.[62] Lemma 5 implies that, except for the knife-edged case where $\{r_i\}_{i=1}^n$ is constant for all i, the injurers who report the act will either all self-report immediately or will all delay reporting. So, if $m \leq n$ injurers report the act in equilibrium, they report at the same time, are randomly assigned to the top m positions in the self-reporting queue, and pay an expected fine of $\frac{1}{m}\sum_{i=1}^m r_i f$.

Next, we study the injurers' decisions about whether to report the harmful act. Lemma 6 presents a sufficient condition for a unique CPNE with self-reporting by all injurers.[63]

Lemma 6. *Take the fine f and the detection probabilities $\{p_i\}_{i=0}^{n-1}$ as fixed. If*

$$\frac{1}{m}\sum_{i=1}^{m} r_i \leq p_{m-1} \quad \forall m = 1, ..., n, \tag{4}$$

then there is a unique CPNE where all n injurers self-report.

If condition (4) holds, then all injurers who commit the act will later self-report. No *individual injurer* ($m = 1$) would want to deviate and remain silent since the expected fine from self-reporting, $\frac{1}{n}\sum_{i=1}^n r_i f$, is smaller than expected fine from remaining silent, $p_{n-1}f$. A *coalition of two injurers* ($m = 2$) would not deviate either. If one of the coalition members expected the other coalition member to remain silent, that coalition member would prefer to join the $n-2$ self-reporters since $\frac{1}{n-1}\sum_{i=1}^{n-1} r_i f \leq p_{n-2}f$ according to condition (4). Following the same logic, no coalition of any size m can deviate in a way that is mutually self-enforcing.

[62] Intuitively, if $\{r_i\}_{i=1}^n$ is increasing in i, then waiting to report the act is a weakly-dominated strategy. So, in equilibrium, any injurer who chooses to report the act will do so immediately. Similarly, if the sequence $\{r_i\}_{i=1}^n$ is decreasing in i, then reporting early is weakly dominated. Hence, in equilibrium, an injurer who chooses to report will delay reporting.

[63] We assume that, when indifferent, the player reports the act. The proof of Proposition 5 verifies that condition (4) is necessary as well as sufficient for self-reporting by all n injurers to be a CPNE when $\{r_i\}_{i=1}^n$ is weakly increasing in i, and that the optimal ordered-leniency mechanism induces all injurers to self-report.

Finally, consider the potential injurers' decisions about whether to participate in the harmful act. Lemma 7 describes the injurers' equilibrium decisions in Stage 1. The proof, which follows the same logic as the proof of Lemma 3, is omitted.

Lemma 7. *Take the fine f and the detection probabilities $\{p_i\}_{i=0}^{n-1}$ as fixed, and suppose that condition (4) holds. Each potential injurer decides to participate in the activity if and only if $b > \hat{b} = \frac{1}{n}\sum_{i=1}^{n} r_i f$.*

As in our benchmark model, a potential injurer will decide to participate in the harmful act if and only if his or her private benefit from committing the act, b, is greater than the deterrence threshold $\hat{b}$ (the expected fine).

5.2 Optimal Enforcement Policies

This section first characterizes the optimal enforcement policies without ordered leniency for self-reporting. We then identify the fine and leniency multipliers that generate maximal deterrence. Finally, we establish the necessary and sufficient conditions under which the first-best outcome may be obtained with law enforcement policies with ordered leniency.

First, consider an environment where leniency for self-reporting is not granted. The analysis of the optimal enforcement policy without leniency is very similar to the analysis in the benchmark model. Without leniency for self-reporting, no injurer self-reports, and the probability that an injurer will be detected and fined is $p_0 = e$. Notably, without leniency for self-reporting, the first-best outcome is not achievable. If the enforcement agency takes no effort to detect illegal activities, $e = 0$, then illegal activities go undetected, $p_0 = 0$, and harmful acts with $b > 0$ are committed. As in Section 4.1, when no leniency is granted to injurers who self-report, the second-best enforcement policy has strictly positive enforcement efforts, $e > 0$, a maximal fine, $f = \bar{f}$, and underdeterrence relative to the first-best level, $\hat{b} = p_0\bar{f} < h$.

Next, consider the optimal enforcement policy with ordered-leniency for self-reporting. We begin by taking the enforcement effort, e, and the corresponding probabilities of detection $\{p_i\}_{i=0}^{n-1}$ as fixed, and characterize the fine and the leniency multipliers that create the highest possible deterrence (i.e., highest expected fine). Formally, the enforcement agency seeks to maximize the expected fine, $\frac{1}{n}\sum_{i=1}^{n} r_i$, subject to condition (4) and $r_m \in [0, 1]$ $(\forall m = 1, 2, ..., n)$.

Proposition 5 characterizes the fine and leniency multipliers that create maximal deterrence.

Proposition 5. *Take the enforcement effort e as fixed. Maximal deterrence is obtained with a maximal fine, $f = \bar{f}$, and the leniency multipliers $r_1 = p_0$ and $r_m = min\{mp_{m-1} - (m-1)p_{m-2}, 1\}$ for $m = 2, ...n$ where $r_1 < r_2 \leq ... \leq r_n$. The injurers commit the act and self-report at time $t = 0$ if $b > \hat{b} = \frac{1}{n}\sum_{i=1}^{n} r_i = \left[\frac{\bar{m}}{n}p_{\bar{m}-1} + \frac{n-\bar{m}}{n}\right]\bar{f}$ where $\bar{m} = \sup\{m \in \{1, ..., n\} | r_m < 1\}$, and do not commit the act otherwise.*

Proposition 5 offers several fundamental insights. The formal analysis is presented in the Appendix. An intuitive discussion follows. First, the highest deterrence is achieved when the fine is set at the maximal level, $f = \bar{f}$. Second, the highest level of deterrence is achieved by implementing a weakly increasing ordered-leniency policy that gives injurers successive discounts for self-reporting based on their positions in the self-reporting queue. Leniency for the first injurer to report will not be full ($r_1 < 1$) and the fine for the last to report may not be maximal (r_n may be smaller than 1).[64] Third, all injurers self-report immediately in the CPNE. In other words, the optimal policy generates a race to the courthouse among the injurers. These findings are aligned with our results in the two-injurer environment.

New insights are derived as well. With an ordered-leniency policy, larger groups of injurers face higher expected fines than smaller groups of injurers. Holding the vector of detection probabilities fixed, when the group size grows, more conspirators will report the act and so the likelihood of detection and the expected fine will rise. More formally, suppose $x_n < 1$, i.e., leniency is also granted to the last position in the self-reporting queue. Since $\bar{m} = n$, the expected fine for a member of a group of size n is $\hat{b} = p_{n-1}\bar{f}$. As the size of the group, n, increases, the probability p_{n-1} increases, and so the expected fine increases too.[65] Corollary 1 summarizes this result.

Corollary 1. *The expected fine faced by an injurer, $\hat{b}$, is strictly increasing in the size of the group, n.*

[64] If $p_i = \frac{i}{i+1}$, then one can easily show that $r_m = \frac{m(m+1)-1}{m(m+1)} < 1$ for all m. Hence, regardless of group size, partial leniency is granted for every position in the self-reporting queue.

[65] Suppose instead that $x_n > 1$. So $\bar{m} < n$. In this case, $\bar{m}$ does not change as the group size n grows. Since $p_{\bar{m}-1} < 1$, the expected fine $\hat{b}$ increases as n increases.

Intuitively, by creating diseconomies of scale with respect to group size, ordered-leniency policies discourage large-scale harmful group activities in favor of small-scale activities.

Next, we present a numerical example to illustrate our main findings.

Example 2. Suppose that the maximal fine is $\bar{f} = 1$. Suppose also the group includes three members, $n = 3$, and $(p_0, p_1, p_2) = (.2, .4, .55)$. Consider an ordered-leniency policy that grants leniency only to the first injurer to self-report, $(r_1, r_2, r_3) = (.2, 1, 1)$. In equilibrium, only one injurer self-reports and the expected fine is .33.[66] The enforcement agency can increase deterrence by also giving leniency to the second and third injurers to report the act. In fact, the leniency multipliers that generate maximal deterrence are $(r_1, r_2, r_3) = (.2, .6, .85)$.[67] In equilibrium, there is a race to the courthouse where the three injurers self-report immediately. The expected fine is .55.[68] Suppose instead that the group includes four injurers, $n = 4$, and $(p_0, p_1, p_2, p_3) = (.2, .4, .55, .6625)$. The leniency multipliers that generate maximal deterrence are $(r_1, r_2, r_3, r_4) = (.2, .6, .85, 1)$.[69] In equilibrium, there is a race to the courthouse where the four injurers self-report immediately. The expected fine is .6625.[70]

We now characterize the optimal enforcement effort e when ordered-leniency policies that generate maximal deterrence are implemented. The next lemma, which follows from Proposition 5, will be used in the analysis of the optimal enforcement effort e. Recall that $p_0(e) = e$ is the probability of detection when no injurer self-reports, $\pi \in (0, 1)$ is the strength of inculpatory evidence, and $p_i(e, \pi) = e + (1 - e)\left[1 - (1 - \pi)^i\right]$ is the probability of detection if $i \in \{1, ..., n - 1\}$ injurers self report.[71]

Lemma 8. *Take the enforcement effort e as fixed. The ordered-leniency multipliers are weakly increasing in i and given by $r_1 = e$ and $r_m = min\{1-(1-e)(1-m\pi)(1-\pi)^{m-2}, 1\}$ for $m = 2, ...n$. The expected fine is $\hat{b}(e, \pi) = \left[1 - \frac{\bar{m}}{n}(1 - e)(1 - \pi)^{\bar{m}-1}\right] \bar{f}$ where $\bar{m} = \sup\{m \in \{1, ..., n\} | m < 1/\pi\}$. The expected fine is continuous, piecewise differentiable, and satisfies $0 < \frac{\partial \hat{b}(e,\pi)}{\partial e} < \bar{f}$ and $\frac{\partial \hat{b}(e,\pi)}{\partial \pi} > 0$.*

[66] The likelihood of detection of silent injurers is $p_1 = .4$. Then, the expected fine is $(.2 + .4 + .4)/3 = 1/3$.

[67] $\bar{m} = 3 = n$.

[68] The expected fine is $(.2 + .6 + .85)/3 = .55 = p_2$.

[69] $\bar{m} = 3 < n$.

[70] The expected fine is $(.2 + .6 + .85 + 1)/4 = ..6625 = p_3$.

[71] $(1 - \pi)^m$ is the chance that a silent conspirator will evade detection if m conspirators self-report. Then, the chance that the silent conspirator is detected and sanctioned is $1 - (1 - \pi)^m$.

Lemma 8 has important implications. First, the expected fine increases when the enforcement agency puts greater effort into detecting illegal activities. Since $\frac{dp_0(e)}{de} = 1 > 0$ and $\frac{\partial p_i(e,\pi)}{\partial e} = (1-\pi)^i > 0$, the entire schedule of detection probabilities rises when the enforcement effort is higher. Second, the expected fine increases when the the inculpatory evidence is stronger. When π rises, p_0 remains fixed but the other detection probabilities rise, $\frac{\partial p_i(e,\pi)}{\partial \pi} = i(1-e)(1-\pi)^{i-1} > 0$ for $i = 1, ..., n-1$. Intuitively, when π is higher, the negative externalities among the injurers are stronger and so the leniency multipliers can be raised, leading to a higher expected fine.

Proposition 6 establishes the necessary and sufficient conditions under which the enforcement agency can implement the first-best outcome with an ordered-leniency policy at zero cost, and describes the second-best enforcement policy when the first-best outcome cannot be achieved.

Proposition 6. *An optimal enforcement policy with ordered leniency for self-reporting can implement the first-best outcome at zero cost if and only if* $h \leq \hat{b}(0,\pi) = \left[1 - \frac{\bar{m}}{n}(1-\pi)^{\bar{m}-1}\right]\bar{f}$ *where* $\bar{m} = \sup\{m \in \{1, ..., n\} | m < 1/\pi\}$. *When* $h > \hat{b}(0,\pi)$ *the second-best enforcement policy involves a maximal fine, positive enforcement costs, and underdeterrence relative to the first best.*

Proposition 6 generalizes Proposition 4 to groups of injurers with more than two members.[72] The proof, which follows immediately after substituting $e = 0$ into Lemma 7, is omitted. Intuitively, when the externalities associated with the harmful activities, h, are not too high, the optimal ordered-leniency policy allows the enforcement agency to achieve the first-best outcome without spending resources on enforcement.

6 Extensions

This section discusses several additional extensions of the benchmark model. The first setting relaxes the assumption of deterministic probabilities of detection by considering an environment where the detection rates depend on the enforcement effort in a stochastic way. The second environment allows for asymmetric benefits across injurers. The third extension allows for endogenous decisions about whether to commit a harmful act alone or in a group.

[72] Suppose $n = 2$. If $\pi \leq \frac{1}{2}$ then $\bar{m} = 2$ and so $\hat{b}(0,\pi) = \frac{1}{2}\pi\bar{f}$. If $\pi > \frac{1}{2}$ then $\bar{m} = 1$ and $\hat{b}(0,\pi) = \frac{1}{2}\bar{f}$. Taken together, we may write $\hat{b}(0,\pi) = \min\{\pi, \frac{1}{2}\}\bar{f}$ as in Proposition 4.

6.1 Stochastic Detection Rate

In our benchmark model, we assumed that the social planner perfectly controls the probabilities of detection, p_0 and p_1, via its enforcement effort e. Injurers, when deciding whether to commit the harmful act, know exactly what these probabilities are, and therefore can accurately forecast their future self-reporting decisions. In the second-best enforcement mechanism, injurers who decide to commit the act in Stage 1 later decide to self-report in Stage 2 (see Proposition 3). Thus, in our baseline model, self-reporting of harmful acts is ubiquitous. Our framework can be extended to allow for a stochastic detection rate.

Consider first our benchmark environment. Suppose that the inculpatory evidence is strong enough to convict a silent injurer with almost certainty. Then, $p_1 = 1 - \varepsilon$, where $\varepsilon > 0$ is an an arbitrarily small number.[73] Suppose also that all the other assumptions of our benchmark model hold. Recall that the probability of detection in the absence of self-reporting is $p_0 = e$. Following our main analysis, the maximal deterrence will be obtained with multipliers $(r_1^S, r_2^S) = (e, 1)$ and $(r_1^M, r_2^M) = \left(\frac{1+2e}{3}, 1\right)$, for the Pareto- and risk-dominance refinements, respectively. Then, the act will be deterred if $b \leq \hat{b}^S(e) = \left(\frac{1+e}{2}\right) \bar{f}$ and $b \leq \hat{b}^M(e) = \left(\frac{2+e}{3}\right) \bar{f}$, for the Pareto- and risk-dominance refinements, respectively (see Proposition 3).

Now suppose that the detection rate p_0 is stochastic. Specifically, after the injurers commit the act, p_0 is drawn from a commonly-known density $q(p_0; e)$ on the unit interval where the median value is e (the enforcement effort of the agency). The realization of p_0 is observed by the injurers. Holding the leniency multipliers, $(r_1^i, r_2^i), i = S, M$, fixed as described above, if $p_0 < e$ (i.e., if detection is relatively unlikely), then the injurers will both remain silent in Stage 2 and not report the act, and will pay a sanction $p_0 \bar{f}$. If instead $p_0 > e$ (i.e., if detection is relatively likely), then the injurers will choose to self-report in Stage 2 and will pay an expected sanction $\hat{b}^i(e), i = S, M$.

In Stage 1, before learning the realization of the random variable p_0, the injurers must decide whether to commit the act. They are deterred from committing the act when

$$b < \int_0^e p_0 \bar{f} q(p_0; e) dp_0 + \int_e^1 \hat{b}^i(e) q(p_0; e) dp_0. \tag{5}$$

Note that the deterrence threshold in this stochastic environment (right-hand side of the inequality) is smaller than $\hat{b}^i(e)$, the deterrence threshold with a certain detection rate. Then, having an uncertain detection rate compromises deterrence in Stage 1. Intuitively, when p_0

[73] For simplicity, and without loss of generality, we abstract from ε for the rest of the analysis.

is stochastic with a median value of e rather than a deterministic value of e, the potential injures benefit from the option of not reporting the act when the probability of detection is small ($p_0 < e$) but do not experience any loss when the probability of detection is large ($p_0 > e$). As a result, the deterrence threshold is lower, and hence, harmful acts are committed more frequently in stochastic environments. As demonstrated earlier, deterrence is at its socially-optimal level when the harm is not too high (see Proposition 3). Hence, social welfare will be unambiguously lower in environments with stochastic detection rates.[74]

6.2 Asymmetric Benefits to Group Members

In our benchmark framework, we assume that the two injurers derive the same private benefit from committing the harmful act. Injurers might not be always symmetric.[75] Our model can be extended to allow for asymmetric benefits to group members.

Suppose that b_1 and b_2 are drawn from a joint density $\phi(b_1, b_2)$ and so the benefit to the first injurer, b_1, could be larger than or smaller than the benefit to the second injurer, b_2. Suppose also that all the other assumptions of our benchmark model hold. The act is socially desirable if the sum of the private benefits of the act exceed the total harm, $b_1 + b_2 > 2h$ or equivalently $(b_1 + b_2)/2 > h$, and socially undesirable otherwise. When transfer payments between the two injurers are impossible, the act will be committed when both potential injurers are willing to participate: $\min(b_1, b_2) \geq \hat{b}$, where $\hat{b}$ is the expected sanction. Interestingly, this new environment may feature overdeterrence of certain socially beneficial acts. To see why, consider an act where the private benefit to the first injurer is very high, $b_1 > 2h$, and the benefit to the second injurer is zero, $b_2 = 0$. This act is socially desirable, since $b_1 + b_2 \geq 2h$, but the act will not be committed for any positive expected sanction $\hat{b} > 0$. The second potential injurer will simply refuse to participate.

With side payments, the potential injurers will commit the act when their joint benefit exceeds the joint expected sanction, $b_1 + b_2 \geq 2\hat{b}$.[76] Hence, our earlier results will carry over to this enriched setting with bargaining at Stage 1. To illustrate this point, consider the

[74] As in our benchmark model, the optimal enforcement effort and the leniency multipliers that maximize deterrence will depend on a variety of factors including the characteristics of the densities $q(p_0; e)$ and $g(b)$.

[75] For instance, asymmetries might arise in environments where one injurer is the mastermind who conceives and plans the harmful act, and recruits others to help him commit the act. Then, the mastermind is the residual claimant of the benefits of the act, while the accomplices are just hired hands.

[76] Note that this environment does not allow side payments to depend on future self-reporting.

environment presented in the previous paragraph. The first potential injurer who anticipates receiving $b_1 > 2h$ can pay the second potential injurer with $b_2 = 0$ to participate in the act as well. Finally note that, since $(b_1 + b_2)/2 \geq \min(b_1, b_2)$, the potential injurers will commit the act for a broader range of values when bargaining is possible.

6.3 Group Acts versus Individual Acts

Our benchmark framework assumes that harmful acts require the participation of two injurers. In practice, however, there are socially harmful activities that can be committed by injurers acting alone rather than in concert with others. In these environments, potential injurers may decide to pursue harmful activities individually instead of in groups. When this possibility is taken into account, the social benefit of implementing an ordered-leniency policy might be smaller than suggested by our previous analysis.

Suppose that two individuals can choose between either committing a harmful act together as a team or committing a (possibly different) act alone. Suppose that the the law enforcement policy, (f, r_1, r_2, e) is a general policy. If nobody reports the act, the probability of detection is p_0 wether the act was pursued by an individual or by a team.[77] With no leniency for self-reporting, $r_1 = r_2 = 1$, the expected fine for an injurer is $p_0 f$ whether he or she commits the act alone or as part of a team. Now suppose instead that the enforcement agency offers leniency for the first position in the self-reporting queue, $r_1 = p_0 - \varepsilon$ where ε is a small positive number, but no leniency for the second position, $r_2 = 1$. The expected fine for an injurer acting alone is $(p_0 - \varepsilon)f$. When acting as part of a team, however, at least one injurer self reports (as suggested by Lemma 2) and the expected fine is $\left(\frac{1+p_0-\varepsilon}{2}\right)f$ which is strictly higher than $p_0 f$. Because of the negative externalities in the self-reporting subgame, an ordered-leniency policy will raise the expected sanction for acts committed by teams but will not affect the expected sanction for individually-committed acts. As a result, ordered-leniency policies might induce injurers to substitute away from harmful group activities and towards harmful individual activities.[78]

[77] In practice, the p_0 for the group act may be higher. Group activities may create more evidence – including hard information, tips, and clues – by virtue of their scale.

[78] Formally, suppose that an injurer derives a private benefit αb where $\alpha \in (0, 1)$ for committing an act alone, but obtains a private benefit b if acting with an accomplice. Suppose further that the harm associated with the act committed by an injurer alone is lower as well, βh where $\beta \in (0, 1]$. The injurers would choose to act individually if and only if $\alpha b - p_0 f \geq \max\{b - \frac{1+p_0}{2} f, 0\}$. That is, acting individually must give the injurers a higher net benefit than committing the act as a team or not committing the act at all.

Law enforcement policies with ordered leniency might have less social value in settings where the alternative to group misbehavior is individual misbehavior (rather than not engaging in any act at all). Although the movement away from group misbehavior towards individual misbehavior is socially desirable if the harm from the individual acts is smaller than the harm from the group act (measured per injurer), the social value created with ordered leniency is smaller than previously described.

Although environments involving stochastic detection rates, asymmetric benefits to group members, and the endogenous decision about whether to participate in group acts or individual acts obviously raise some new and interesting issues, the main insights derived from our benchmark model and the implications for the design of optimal enforcement policies with ordered leniency remain relevant.

7 Discussion and Conclusions

This paper studies the design of enforcement schemes with ordered leniency for detecting and preventing harmful short-term activities conducted by groups of two or more injurers. We demonstrate that ordered-leniency policies that generate maximal deterrence give successively larger discounts to injurers who secure higher positions in the reporting queue, creating a so-called "race to the courthouse" among the members of the group of injurers. As a result, detection of harmful acts occurs with certainty in equilibrium. Our analysis shows that the socially-optimal level of deterrence can be obtained at zero cost with an enforcement policy with ordered leniency when the externalities associated with the harmful activities are not too high. In contrast, enforcement policies that do not grant leniency for self-reporting cannot implement the first-best outcome when there is an upper bound on the fines that can be imposed. More generally, enforcement policies with ordered leniency are superior to enforcement policies that do not grant leniency for self-reporting. Thus, we provide a social welfare rationale for the current use of ordered-leniency policies in the real world.

Our findings regarding the superiority of enforcement policies with ordered leniency for groups of injurers complement Kaplow and Shavell's (1994) results for single-injurer environments. Kaplow and Shavell (1994) show that leniency for self-reporting reduces the enforcement agency's cost without compromising deterrence. In our model, ordered-leniency policies are socially desirable because these policies create stronger detection and deterrence

and reduce the number of socially harmful activities. Ordered-leniency policies may have other significant benefits as well. Since they create a race to the courthouse, they allow the enforcement agency to detect crimes faster. Faster detection may have independent value, insofar as it allows the agency to prevent the crime from continuing, helps mitigate the harm, and economizes on future detection efforts (as in Kaplow and Shavell, 1994).

Several relevant extensions are discussed. We consider an environment where the detection rate depends on the enforcement effort in a stochastic way. In this setting, injurers who commit the act may refrain from self-reporting if the probabilities of detection are sufficiently low. As a result, deterrence might be compromised. We also discuss a setting that allows for asymmetric benefits from committing a harmful act across injurers. In this setting, the equilibrium outcomes heavily depend on the ability of group members to write side contracts with each other and negotiate transfer payments. Our earlier results carry over when monetary transfers are possible. Finally, we explore an environment where the potential injurers can decide whether to commit individual or group harmful acts, and show that under certain conditions, the social value of ordered-leniency policies might be reduced.

In separate recent work (Landeo and Spier, 2018), we investigate the effectiveness of law enforcement policies with ordered leniency in a laboratory setting. We replicate the strategic environment described here. Three leniency conditions are considered. (1) Strong Leniency, where the first to report receives a strong reduction in the penalty. Strong Leniency implements a prisoners' dilemma game. (2) Mild Leniency, where the first to report receives a mild reduction in the penalty. Mild Leniency implements a coordination game. (3) No Leniency, where penalty reductions for self-reporting are not granted. Our experimental results suggest that the injurers' behaviors are aligned with the risk-dominance refinement. In other words, when the wrongdoers are sufficiently distrustful of each other, the enforcement agency can implement an optimal enforcement policy using a coordination game or a prisoners' dilemma game. More specifically, our findings indicate that Mild Leniency, as well as Strong Leniency, create a "race-to-the-courthouse" and, hence, induce both injurers to self-report immediately. As a result, harmful acts are detected with (almost) certainty and the average fines paid by the injurers are higher (compared to No Leniency).

Although our model focuses on short-term criminal activities, the main insights are relevant for ongoing criminal activities as well. Ordered-leniency policies could change the incentives of firms to form and maintain cartels, for example, and could hasten the detection of criminal

price fixing schemes.[79] In the United States, the first firm to report the illegal cartel activity and cooperate with the authorities receives full leniency from prosecution,[80] and the second and subsequent firms to self-report may receive lenient treatment as well.[81] The European Union has a similar policy.[82] Ordered-leniency policies have been successfully used in a number of high-profile antitrust cases, including the 2006 international investigation and prosecution of several air cargo carriers who paid more than $3B in criminal and regulatory fines.[83] Our framework would be a natural (and realistic) component of a repeated-game analysis.[84]

Our paper is motivated by insider trading and securities fraud. We believe, however, that the analysis and insights derived from our work might apply to other contexts as well. For instance, our findings are relevant to *qui tam* (whistleblower) lawsuits brought under the U.S. False Claims Act (FCA). The FCA allows regular citizens to bring lawsuits against federal contractors claiming fraud against the federal government.[85] The *qui tam* provision of the act grants the whistleblower a fraction of ultimate recovery, often on the order of 15 to 25 percent. Under a first-to-file rule, "[w]hen a person brings an action under the False Claims Act, no person other than the Government may intervene or bring a related action based on the facts underlying the pending action."[86] In practice, however, the second and subsequent

[79]Miller (2009) finds empirical support for the notion that the 1993 leniency program in the United States increased the detection of cartels and improved deterrence. Note however that Gartner and Zhou (2012) argue that, in practice, firms apply for leniency long after cartels collapse.

[80]See Antitrust Division, U.S. Department of Justice Corporate Leniency Policy (1993), available at https://www.justice.gov/atr/corporate-leniency-policy, last visited July 6, 2018.

[81]Discounts for the second and subsequent firms are supported by the United States Sentencing Guidelines (U.S.S.G.8C4.1). In the words of former Deputy Assistant Attorney General Scott Hammond (2006, p. 2), "A second-in company's cooperation can vary dramatically from case to case. While a second-in company's cooperation typically will significantly advance an investigation, there are times when the cooperation is either cumulative or no longer needed."

[82]See http://ec.europa.eu/competition/cartels/leniency/leniency.html, last visited July 6, 2018.

[83]See Press Release $IP/10/1487$ (2010). Lufthansa and its subsidiary Swiss Air were the first to assist and received full immunity under the European Union's leniency program. Several other airlines were also granted reductions under the EU program, including Martinair (50%), Japan Airlines (25%), Air France-KLM (20%), Air Canada (15%), and British Airways (10%).

[84]See Motta and Polo (2003), Spagnolo (2005), and Chen and Rey (2013) for seminal theoretical work on law enforcement with leniency for long-term collusive agreements. See Bigoni et al. (2012) for recent experimental work on leniency and cartels in repeated-game environments.

[85]31 U.S.C. §§3729–3733,

[86]31 U.S.C. §3730(b)(5). The rationale for this feature of the policy is "to filter out 'parasitic' qui tam suits that do not offer the government information it does not already have" (Engstrom, 2012, p. 1274).

plaintiffs to file suit may receive compensation as well.[87] Our findings indicate that it might be advisable to expand the scope of *qui tam* privileges to include the second and subsequent whistleblower suits, depending on the strength of the inculpatory evidence and the incremental informational content.

Our work provides important lessons for the design of optimal law enforcement policies involving corporate and individual criminal liability. In the United States, both corporations and individual lawbreakers face criminal liability for corporate crimes committed in the scope of employment. As noted by Arlen (2012), corporate criminal liability might be justified on the grounds that firms can "more cost-effectively ... identify the individuals responsible for crimes. ... and can access information and employees (e.g. foreign based employees) more effectively than can the state" (p. 166). Corporate liability is particularly valuable when the assets of the individual lawbreakers are insufficient to deter the harmful act. Leniency for self-reporting might be granted to corporations and employees.[88] In the context of corporate and individual liability, our results suggest that the implementation of ordered-leniency policies might create a race between the employer and the employee to self-report criminal activities. The sanction reduction granted to the first to report would not generally be full, and the sanctions faced by the second to report may not be maximal.[89]

This paper, together with the empirical evidence in Landeo and Spier (2018), calls into question the sole application of the proverbial prisoners' dilemma in the design of plea-bargaining agreements in the real world. The famous story about two prisoners being held in separate cells was first articulated by a Princeton mathematics professor, Albert William Tucker, while addressing an audience of psychologists in 1950.[90] Since then, the story has been told and retold countless times, and a Google Scholar search for the phrase "prisoners'

[87]For instance, several different whistleblowers received significant rewards in their *qui tam* suits against Pfizer (see Pfizer Settlement Agreement, https://www.justice.gov/usao-ma/file/847081/download, last visited July 6, 2018).

[88]Corporations that implement internal compliance systems might also receive leniency. Note that our findings might also apply to the design of optimal internal compliance systems with self-reporting. See Arlen and Kraakman (1997) and Kraakman (1986) for seminal work on corporations as third-party law enforcers.

[89]Our findings and insights might be also relevant for the design of law enforcement mechanisms associated with environmental policies and standards and tax policies and control of tax evasion.

[90]"In 1950 addressing an audience of psychologists at Stanford University, where he was a visiting professor, Tucker created the Prisoners' Dilemma to illustrate the difficulty of analyzing non-zero-sum games" (https://www.princeton.edu/pr/news/95/q1/0126tucker.html, last visited July 6, 2018).

dilemma" delivers more than two hundred thousand articles in academic fields as diverse as economics, biology, philosophy, sociology, political science, and of course law.[91] Our analysis demonstrates that the proverbial prisoners' dilemma is not the only way to conduct plea bargaining or to detect and punish socially harmful activities. When the prisoners are sufficiently distrustful of each other, the prosecutor could forego the prisoners' dilemma and employ a coordination mechanism instead.

[91]Last searched, July 6, 2018.

References

Andreoni, James. 1991. "The Desirability of a Permanent Tax Amnesty." *Journal of Public Economics*, 45: 143–159.

Apesteguia, Jose, Martin Dufwenberg, and Reinhard Selten. 2007. "Blowing the Whistle." *Economic Theory,* 31: 143–166.

Arlen, Jennifer and Reinier Kraakman. 1997. "Controlling Corporate Misconduct: An Analysis of Corporate Liability Regimes." *New York University Law Review,* 72: 687–779.

Arlen, Jennifer. 2012. "Corporate Criminal Liability: Theory and Evidence." In Alon Harel and Keith Hylton, eds., *Research Handbook on Criminal Law*. Massachusetts: Edward Elgar Publishing.

Aubert, Cécile, Patrick Rey, and William E. Kovacic. 2006. "The Impact of Leniency and Whistle-Blowing Programs on Cartels." *International Journal of Industrial Organization,* 24: 1241–1266.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, 76: 169–217.

Bernheim, Douglas B., Bezalel Peleg, and Michael D. Whinston. 1987. "Coalition Proof Nash Equilibria I: Concepts." *Journal of Economic Theory*, 42: 1–12.

Bigoni, Maria, Sven-Olof Fridolfsson, Chloe Le Coq, and Giancarlo Spagnolo. 2012. "Fines, Leniency, and Rewards in Antitrust." *RAND Journal of Economics*, 43: 368–90.

Bigoni, Maria, Sven-Olof Fridolfsson, Chloe Le Coq, and Giancarlo Spagnolo. 2015. "Trust, Leniency, and Deterrence." *Journal of Law, Economics, and Organization*, 31: 663–689.

Buccirossi, Paolo and Giancarlo Spagnolo. 2006. "Leniency Policies and Illegal Transactions." *Journal of Public Economics*, 90: 1281–1297.

Ceresney, Andrew. 2015. "The SEC's Cooperation Program: Reflections on Five Years of Experience." http://www.sec.gov./news/speech/sec-cooperation-program.html.

Che, Yeon-Koo and Seung-Weon Yoo. 2001. "Optimal Incentives for Teams." *American Economic Review,* 91: 525–541.

Chen, Zhijun and Patrick Rey. 2013. "On the Design of Leniency Programs." *Journal of Law and Economics*, 56: 917–957.

Engstrom, David F. 2012. "Harnessing the Private Attorney General: Evidence from Qui Tam Litigation." *Columbia Law Review*, 112: 1244-1325.

Feltovich, Nick and Yasuyo Hamaguchi. 2018. "The Effect of Leniency Programmes on Anti-Competitive Behaviour: An Experimental Study." *Southern Economic Journal,* 84: 1024–1049

FBI. 2012. "Financial Crimes Report 2010–2011." https://www.fbi.gov/stats-services/publications/financial-crimes-report-2010-2011.

Feess, Eberhardt and Markus Walzl. 2004. "Self-Reporting in Optimal Law Enforcement When There Are Criminal Teams." *Economica*, 71: 333-348.

Feess, Eberhardt and Markus Walzl. 2010. "Evidence Dependence of Fine Reductions in Corporate Leniency Programs." *Journal of Institutional and Theoretical Economics*, 166: 573-590.

Gärtner, Dennis L. and Jun Zhou. 2012. "Delays in Leniency Application: Is there Really a Race to the Enforcer's Door?" Discussion Paper No. 395, University of Bonn.

Grossman, Gene M. and Michael L. Katz. 1983. "Plea Bargaining and Social Welfare." *American Economic Review*, 73: 749–757.

Hammond, Scott D. 2006. "Measuring the Value of Second-In Cooperation in Corporate Plea Negotiations." The 54th Annual American Bar Association Section of Antitrust Law Spring Meeting, March 29, 2006. http://www.justice.gov/atr/public/speeches/215514.htm.

Harrington, Joseph E. 2013. "Corporate Leniency Programs When Firms Have Private Information: The Push of Prosecution and the Pull of Pre-Emption." *Journal of Industrial Economics*, 51: 1–27.

Harsanyi, John C. and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press.

Hinloopen, Jeroen and Adriaan R. Soetevent. 2008. "Laboratory Evidence on the Effectiveness of Corporate Leniency Programs." *RAND Journal of Economics,* 39: 607–616.

Innes, Robert. 1999. "Remediation and Self-reporting in Optimal Law Enforcement." *Journal of Public Economics*, 72: 379-393.

Kaplow, Louis and Steven Shavell. 1994. "Optimal Law Enforcement with Self-Reporting of Behavior." *Journal of Political Economy*, 102: 583–606.

Kobayashi, Bruce. 1992. "Deterrence with Multiple Defendants: An Explanation for 'Unfair' Plea Bargains." *RAND Journal of Economics*, 23: 507–517.

Kornhauser, Lewis A. and Richard L. Revesz. 1994. "Multidefendant Settlements under Joint and Several Liability: The Problem of Insolvency." *Journal of Legal Studies,* 23: 517–542.

Kraakman, Reinier H. 1986. "Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy." *Journal of Law, Economics & Organization*, 2: 53–104.

Landeo, Claudia M. and Kathryn E. Spier. 2009. "Naked Exclusion: An Experimental Study of Contracts with Externalities." *American Economic Review*, 99: 1850–1877.

Landeo, Claudia M. and Kathryn E. Spier. 2012. "Exclusive Dealing and Market Foreclosure: Further Experimental Results." *Journal of Institutional and Theoretical Economics*, 168: 150–170.

Landeo, Claudia M. and Kathryn E. Spier. 2015. "Incentive Contracts for Teams: Experimental Evidence." *Journal of Economic Behavior and Organization*, 119: 496–511.

Landeo, Claudia M. and Kathryn E. Spier. 2018. "Ordered Leniency: An Experimental Study of Law Enforcement with Self-Reporting." Mimeo, University of Alberta and Harvard University.

Landes, William M. 1971. "An Economic Analysis of the Courts." *Journal of Law and Economics*, 14: 61–108.

Livernois, John and C.J. McKenna. 1999. "Truth or Consequences: Enforcing Pollution Standards with Self-Reporting." *Journal of Public Economics*, 71: 415–440.

Malik, Arun S. and Robert M. Schwab. 1991. "The Economics of Tax Amnesties." *Journal of Public Economics,* 46: 29–49.

Malik, Arun S. 1993. "Self-Reporting and the Design of Policies for Regulating Stochastic Pollution." *Journal of Environmental Economics and Management,* 24: 241–257.

Miller, Nathan. 2009. "Strategic Leniency and Cartel Enforcement." *American Economic Review.* 99: 750–768.

Motta, Massimo and Michele Polo. 2003. "Leniency Programs and Cartel Prosecution." *International Journal of Industrial Organization,* 21: 347–379.

Polinsky, A. Mitchell and Steven Shavell. 1984. "The Optimal Use of Fines and Imprisonment." *Journal of Public Economics*, 24: 89–99.

Press Release IP/10/1487. 2010. "Commission fines 11 air cargo carriers €799 million in price fixing cartel." European Commission, November 9, 2010. http://europa.eu/rapid /press-release_IP-10-1487_en.htm .

Reinganum, Jennifer F. 1988. "Plea Bargaining and Prosecutorial Discretion." *American Economic Review,* 78: 713–728.

Spagnolo, Giancarlo. 2005. "Divide et Impera: Optimal Leniency Programs." Mimeo, Stockholm School of Economics.

Spagnolo, Giancarlo and Catarina Marvão. 2016. "Cartels and Leniency: Taking Stock of What We Learnt." In Luis .C. Corchón and Marco A. Marini, eds., *Handbook of Game Theory and Industrial Organization.* Massachusetts: Edward Elgar Publishing.

Spier, Kathryn E. 1994. "A Note on Joint and Several Liability: Insolvency, Settlement, and Incentives." *Journal of Legal Studies,* 23: 559–568.

Appendix

This Appendix presents formal proofs of the lemmas and propositions.

Proof of Lemma 1. Denote the strategy of player j as $\sigma_j = (\rho_j, t_j)$ where $\rho_j \in \{R, NR\}$ is whether to report the act and $t_j \in [0,1]$ is when to report the act. Suppose $r_1 < r_2$. If $\sigma_{-j} = (NR, t_{-j})$, then player j is indifferent about their reporting time, $(R,0) \sim (R,t_j)\ \forall t_j \in (0,1]$. If $\sigma_{-j} = (R, t_{-j})$, then for player j we have $(R,0) \sim (R,t_j)\ \forall t_j < t_{-j}$ and $(R,0) \succ (R,t_j)\ \forall t_j \geq t_{-j}$. Therefore $(R,0)$ weakly dominates $(R,t_j)\ \forall t_j \in (0,1]$ when $r_1 < r_2$. Suppose instead that $r_1 > r_2$. f $\sigma_{-j} = (NR, t_{-j})$, then player j is indifferent, $(R,1) \sim (R,t_j)\ \forall t_j \in [0,1)$. If $\sigma_{-j} = (R, t_{-j})$, then $(R,1) \sim (R,t_j)\ \forall t_j > t_{-j}$ and $(R,1) \succ (R,t_j)\ \forall t_j \leq t_{-j}$. Therefore $(R,1)$ weakly dominates $(R,t_j)\ \forall t_j \in [0,1)$ when $r_1 > r_2$. If $r_1 = r_2$ then there is no advantage to being first or second and so the players are indifferent as to the reporting time. ■

Proof of Lemma 2. In Case 1, $b - r_1 f \geq b - p_0 f$ and $b - \left(\frac{r_1+r_2}{2}\right) f \geq b - p_1 f$. With the tie-breaking assumption, self-reporting is a dominant strategy and (R,R) is the unique Nash equilibrium (NE). In Case 4, $b - r_1 f < b - p_0 f$ and $b - \left(\frac{r_1+r_2}{2}\right) f < b - p_1 f$ so not reporting is a dominant strategy and (NR,NR) is the unique NE. In Case 2, $b - r_1 f < b - p_0 f$ and $b - \left(\frac{r_1+r_2}{2}\right) f \geq b - p_1 f$ so (R, NR) and (NR, R) are both pure-strategy NE. In Case 3 there are two pure-strategy NE, (R, R) and (NR, NR). (R, R) Pareto-dominates (NR, NR) if $b - \left(\frac{r_1+r_2}{2}\right) f \geq b - p_0 f$ or $\frac{r_1+r_2}{2} \leq p_0$. (R, R) risk-dominates (NR, NR) if the former is preferred by player j if player $-j$ is randomizing 50/50 between R and NR, or $\frac{1}{2}(b - r_1 f) + \frac{1}{2}\left(b - \left(\frac{r_1+r_2}{2}\right) f\right) \geq \frac{1}{2}(b - p_0 f) + \frac{1}{2}(b - p_1 f)$, or $\frac{3r_1+r_2}{4} \leq \frac{p_1+p_1}{2}$. ■

Proof Lemma 3. Consider the four cases included in Lemma 2. In Case 1, (R, R) is the unique NE and each injurer receives a payoff of $b - \left(\frac{r_1+r_2}{2}\right) f$. It is therefore a weakly dominant strategy for an injurer to participate in the act if $b > \left(\frac{r_1+r_2}{2}\right) f$. In Case 2, (R, NR) and (NR, R) are both pure-strategy NE with an average payoff of $b - \left(\frac{r_1+p_1}{2}\right) f$. The act is committed when $b > \left(\frac{r_1+p_1}{2}\right) f$. In Case 3 three are two NE, (R, R) and (NR, NR). The act is committed if $b > p_0 f$ or $b > \left(\frac{r_1+r_2}{2}\right) f$, depending on which of the two equilibria is expected to prevail. Finally, in Case 4, (R, NR) is the unique pure-strategy NE and the act is committed if $b > p_0 f$. ■

Proof of Proposition 3. First, we characterize the expected fine for each of the four cases included in Lemma 2, and identify the maximal expected fines.

Case 1. Both injurers self-report in this case. We now characterize the values (r_1, r_2) that maximize the expected fine$\left(\frac{r_1+r_2}{2}\right) f$ subject to the constraints that (i) $\frac{r_1+r_2}{2} \leq p_1$, (ii) $r_1 \in [0, p_0]$, and (iii) $r_2 \in [0,1]$. Two sub-cases are considered.

Case 1.1 The first case refers to $p_1 \leq \frac{1+p_0}{2}$. If $p_1 \leq \frac{1+p_0}{2}$, then constraint (i) must hold with equality, $\frac{r_1+r_2}{2} = p_1$. Suppose not: $\frac{r_1+r_2}{2} < p_1$. This would imply that both $r_1 = p_0$ and $r_2 = 1$, for otherwise the expected fine $\left(\frac{r_1+r_2}{2}\right) f$ could be increased. Then, $\frac{r_1+r_2}{2} = \frac{1+p_0}{2} < p_1$,

a contradiction. Therefore $\frac{r_1+r_2}{2} = p_1$. We can write $(r_1, r_2) = (p_1 - \Delta, p_1 + \Delta)$, where Δ is a constant. Since $r_1 \in [0, p_0]$, it must be that $p_1 - p_0 \leq \Delta \leq p_1$. Since $r_2 \in [0, 1]$, it must be that $-p_1 \leq \Delta \leq 1 - p_1$. Taken together, $\Delta \in [p_1 - p_0, \min\{p_1, 1 - p_1\}]$. $p_1 \leq \frac{1+p_0}{2}$ implies that $p_1 - p_0 \leq \min\{p_1, 1 - p_1\}$, so this range exists. The expected fine is $p_1 f$.

Case 1.2. The second case refers to $p_1 > \frac{1+p_0}{2}$. If $p_1 > \frac{1+p_0}{2}$, then constraint (i) does not bind at the optimum: $\frac{r_1+r_2}{2} < p_1$. Suppose not: $\frac{r_1+r_2}{2} = p_1$, Then, as above we would have $(r_1, r_2) = (p_1 - \Delta, p_1 + \Delta)$, where $\Delta \in [p_1 - p_0, \min\{p_1, 1 - p_1\}]$. But $p_1 > \frac{1+p_0}{2}$ implies $2p_1 > 1 + p_0$, which implies further that $p_1 - p_0 > \min\{p_1, 1 - p_1\}$. So no such value for Δ exists. Therefore $\frac{r_1+r_2}{2} < p_1$. It must also be true that $(r_1, r_2) = (p_0, 1)$. If $r_1 < p_0$ and/or $r_2 < 1$, then the expected fine would be higher (and no constraints violated) if r_1 and/or r_2 were raised. The expected fine is $\left(\frac{1+p_0}{2}\right) f < p_1 f$.

Case 2. Since only one injurer self-reports, the expected fine is $\left(\frac{r_1+p_1}{2}\right) f$. Since r_1 is constrained to be less than or equal to p_0 in this case, the strongest possible deterrence is obtained when $r_1 = p_0$. So the expected fine is less than or equal to $\left(\frac{p_0+p_1}{2}\right) f$. This expected fine is strictly lower than the expected fine in Case 1.

Case 3. There are multiple equilibria in this case.

With *Pareto dominance*, the injurers self-report if and only if $\frac{r_1+r_2}{2} \leq p_0$. The expected fine is less than or equal to $p_0 f$. This expected fine is always strictly lower than the expected fine in Case 1.

With *risk dominance*, the enforcer maximizes $\frac{r_1+r_2}{2}$ subject to the constraints that (i) $\frac{3r_1+r_2}{4} \leq \frac{p_0+p_1}{2}$, (ii) $r_1 \in [p_0, 1]$, and (iii) $r_2 \in [0, 1]$. Holding r_1 fixed, deterrence is increased by raising r_2 to the point where constraint (i) or constraint (iii) binds. Given r_1, we must have $r_2 = \min\{2(p_0 + p_1) - 3r_1, 1\}$. The enforcer's problem can be represented as choosing $r_1 \in [p_0, 1]$ to maximize $\frac{r_1+\min\{2(p_0+p_1)-3r_1,1\}}{2}$. Two sub-cases are considered.

Case 3.1 The first case refers to *risk dominance* and $p_1 \leq \frac{1+p_0}{2}$. If $p_1 \leq \frac{1+p_0}{2}$, then $2p_1 \leq 1 + p_0$. This implies that $2(p_0 + p_1) - 3r_1 \leq 1 - 3(r_1 - p_0) \leq 1$, for all $r_1 \in [p_0, 1]$. So $\min\{2(p_0 + p_1) - 3r_1, 1\} = 2(p_0 + p_1) - 3r_1$, and the expected fine is $(p_0 + p_1 - r_1)f$ for all $r_1 \in [p_0, 1]$. Deterrence is maximized by making r_1 as small as possible, so $r_1 = p_0$ and $r_2 = 2(p_0 + p_1) - 3r_1 = 2p_1 - p_0$, and the expected fine is $p_1 f$. This expected fine is the same as the expected fine in Case 1.

Case 3.2 The second case refers to *risk dominance* and $p_1 > \frac{1+p_0}{2}$. If $p_1 > \frac{1+p_0}{2}$, then r_1 will be strictly greater than p_0, and the expected fine strictly higher than $p_1 \bar{f}$. To see why this is true, suppose $r_1 = p_0 + \varepsilon$ where $\varepsilon > 0$. Since $p_1 > \frac{1+p_0}{2}$ implies $2p_1 > 1 + p_0$, we have $2(p_0+p_1)-3r_1 = 2p_1-p_0-3\varepsilon > 1$ when ε is not too large. Therefore $\min\{2(p_0+p_1)-3r_1, 1\} = 1$ when $r_1 = p_0 + \varepsilon$ for $\varepsilon > 0$ sufficiently small. The expected fine in this case is $\left(\frac{r_1+1}{2}\right) f$. Deterrence would be higher if r_1 were raised above p_0. r_1 will be raised to the point where $2(p_0 + p_1) - 3r_1 = 1$ and so $r_1 = \frac{2(p_0+p_1)-1}{3}$ and $r_2 = 1$. The expected fine is $\left(\frac{1+p_0+p_1}{3}\right) f$. This expected fine is strictly higher than the expected fine in Case 1.

Case 4. Neither injurer self-reports. The expected fine is $p_0 f$. This expected fine is strictly lower than the expected fine in Case 1.

Hence, when *Pareto dominance* is applied in Case 3, the maximal expected fine always corresponds to Case 1. When *risk dominance* is applied in Case 3 and $p_1 \leq \frac{1+p_0}{2}$, the maximal expected fine corresponds to Case 1 or Case 3; when *risk dominance* is applied in Case 3 and $p_1 > \frac{1+p_0}{2}$, the maximal expected fine corresponds to Case 3.

Second, since $r_1^j < r_2^j$ for $j = S, M$, all reporting takes place at $t = 0$, by Lemma 1.

Third, since the equilibria of the self-reporting subgame described in Lemmas 1 and 2 do not depend on the level of the fine, f, the highest deterrence is obtained with the maximal fine, $f = \bar{f}$. ■

Proof of Lemma 4. Proposition 3 implies (1) if $p_1 \leq \frac{1+p_0}{2}$, then $\hat{b}^S = \hat{b}^M = p_1\bar{f}$; and, (2) if $p_1 > \frac{1+p_0}{2}$, then $\hat{b}^S = \left(\frac{1+p_0}{2}\right)\bar{f}$, $\hat{b}^M = \left(\frac{1+p_0+p_1}{3}\right)\bar{f}$, and $\hat{b}^S < \hat{b}^M$. Substituting $p_0 = e$ and $p_1 = e + (1-e)\pi$ gives parts (1) and (2) of the lemma. ■

Proof of Proposition 4. First, the characterization of the first-best outcome follows immediately from the proofs of Proposition 3 and Lemma 4.

Second, the characterization of the fine and leniency multipliers implemented in the second-best outcome follow the proofs of Proposition 3 and Lemma 4.

Third, we demonstrate that the second-best outcome involves positive enforcement efforts. The social welfare function is given by:

$$W = \int_{\hat{b}^i(e,\pi)}^{\infty} (b-h)g(b)db - c(e),$$

where $\hat{b}^i(e,\pi)$, $i = S, M$, correspond to the deterrence thresholds under the Pareto-dominance and risk-dominance refinements, respectively. The enforcement agency chooses e to maximize social welfare. The first-order condition is:

$$(h - \hat{b}^i(e,\pi))\frac{\partial \hat{b}^i(e,\pi)}{\partial e} g(\hat{b}^i(e,\pi)) - c'(e) = 0.$$

As before, the first term represents the incremental benefit from increasing the probability e: $h - \hat{b}^i(e,\pi)$ is the social gain associated with deterring an additional harmful act, and $\frac{\partial \hat{b}(e,\pi)}{\partial e} g(\hat{b}^i(e,\pi))$ is the incremental volume of harmful acts that are deterred when the detection rate e increases. The second term, $c'(e)$, represents the marginal cost of effort. Rearranging terms, we find that the second-best optimal deterrence threshold (optimal expected fine) satisfies:

$$\hat{b}^i(e,\pi) = h - \frac{c'(e)}{\frac{\partial \hat{b}^i(e,\pi)}{\partial e} g(\hat{b}^i(e,\pi))}.$$

We need to show that the second-best outcome involves $e^i > 0$. Suppose not: $e^i = 0$. In this case, $h > \hat{b}^i(0,\pi)$ since by assumption the first-best enforcement policy cannot be obtained;

$\frac{\partial \hat{b}^i(e,\pi)}{\partial e} > 0$ by Lemma 4; and $g(\hat{b}^i(0,\pi)) > 0$ since the density function has full support. Since $c'(0) = 0$, we have that the slope of the social welfare function is strictly positive when $e^i = 0$ and so we conclude that $e^i > 0$. Next, we show that $\hat{b}^i(e^i,\pi) < h$. Suppose instead that $\hat{b}^i(e^i,\pi) \geq h$. Since $\frac{\partial \hat{b}^i(e,\pi)}{\partial e} g(\hat{b}^i(e,\pi)) > 0$, the slope of the welfare function would be strictly negative. Social welfare would be higher if e were reduced. ■

Proof of Proposition 5. Given that the injurers' incentives in the self-reporting subgame are not affected by f, for simplicity and without loss of generality, assume that $f = 1$.

The proof involves several steps. We begin with a critical building block. Let $\mathbf{x}$ be the vector of multipliers for which condition (4) holds with equality. The system of equations is as follows:

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ & & \dots\dots\dots\dots & & & \\ 1 & 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} p_0 \\ 2p_1 \\ \dots \\ (n-1)p_{n-2} \\ np_{n-1} \end{bmatrix}.$$

Multiplying by the inverse of the (lower) triangular matrix, we get:

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ & & \dots\dots\dots\dots & & & \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \begin{bmatrix} p_0 \\ 2p_1 \\ \dots\dots \\ (n-1)p_{n-2} \\ np_{n-1} \end{bmatrix} = \begin{bmatrix} p_0 \\ 2p_1 - p_0 \\ \dots\dots\dots \\ (n-1)p_{m-2} - (n-2)p_{m-3} \\ np_{n-1} - (n-1)p_{n-2} \end{bmatrix}.$$

The vector $\mathbf{x}$ has important properties: $x_1 = p_0 > 0$; $x_1 < x_2 < ....x_n$, by our assumption that the sequence $\{ip_{i-1}\}_{i=1}^n$ is convex in i; and, x_j $(j = 2, ..., n)$ may be less than, equal to, or greater than 1. Let $\bar{m}$ be the position in the self-reporting queue for which $x_{\bar{m}} < 1 \leq x_{\bar{m}+1}$.

Next, we will demonstrate that an optimal ordered-leniency policy has $r_i = \min\{x_i, 1\}$ for all i, that all injurers self-report in the CPNE, and that the sum of the fines is

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{\bar{m}} r_i + \sum_{i=\bar{m}+1}^{n} 1 = \bar{m}p_{\bar{m}-1} + (n - \bar{m}).$$

Four claims and their respective proofs follow.

Claim 1. *Suppose $\{r_i\}_{i=1}^n$ is weakly decreasing in i. In any CPNE, the expected fine is less than or equal to p_0.*

Proof of Claim 1. First, suppose $\{r_i\}_{i=1}^n$ is constant in i, so $r_1 = ... = r_n$. If $r_1 < p_0$, then there is a unique CPNE where all injurers self-report the act and the fine is less than p_0. If $r_1 > p_0$, then there is a unique CPNE where no injurer self-reports the act and the expected fine is p_0.

Next, suppose $\{r_i\}_{i=1}^n$ is weakly decreasing in i with at least one strict inequality. We will now verify that in any CPNE, either all n injurers self-report or all n injurers do not self-report. We proceed by contradiction. Suppose there is a CPNE where $m < n$ injurers self-report and the remaining $n - m + 1$ injurers do not self-report. It must be true that $\frac{1}{m}\sum_{i=1}^{m} r_i \leq p_{m-1}$. If this was not true, then an individual who self-reports (somebody in the group of m) would strictly prefer to deviate, not report, and pay fine p_{m-1}. It also must be true that $p_m \leq \frac{1}{m+1}\sum_{i=1}^{m+1} r_i$, since otherwise a silent individual (in the group of $n - m + 1$) would strictly prefer to self-report. Combining expressions, and using the premise that $\{r_i\}_{i=1}^n$ is weakly decreasing in i, we have:

$$p_m \leq \frac{1}{m+1}\sum_{i=1}^{m+1} r_i \leq \frac{1}{m}\sum_{i=1}^{m} r_i \leq p_{m-1}.$$

This is a contradiction, since by assumption $p_m > p_{m-1}$ for all m. This completes the proof that, in any CPNE, either all n injurers self-report or all n injurers do not self-report.

We now construct the unique CPNE of the game. There are two cases to consider.

(i) Suppose $\frac{1}{n}\sum_{i=1}^{n} r_i > p_0$. There is a unique CPNE where no injurer self-reports and the expected fine is p_0. Since $\{r_i\}_{i=1}^n$ is weakly decreasing, we have $r_i > p_0$ for all i and $\frac{1}{m}\sum_{i=1}^{m} r_i > p_0$ for all m. No individual or group of m injurers would deviate and self-report. Since nobody self-reports the expected fine is p_0.

(ii) Suppose instead that $\frac{1}{n}\sum_{i=1}^{n} r_i < p_0$. There is a unique CPNE where all n injurers self-report. No individual would prefer to unilaterally deviate and not report, since the expected fine from the unilateral deviation is $p_{n-1} > p_0 > \frac{1}{n}\sum_{i=1}^{n} r_i$. More generally, no coalition of size m would deviate and self-report, because $p_{n-m} > p_0 > \frac{1}{n}\sum_{i=1}^{n} r_i$. Since everyone self-reports, then the expected fine is smaller than p_0. □

Claim 2. *Suppose $\{r_i\}_{i=1}^n$ is weakly increasing in i. Condition (4), which states that $\frac{1}{m}\sum_{i=1}^{m} r_i \leq p_{m-1}$ for all $m = 1, 2, ..., n$, is both necessary and sufficient for self-reporting by all n injurers to be a CPNE.*

Proof of Claim 2. The proof that condition (4) is sufficient is in the main text of the paper. We now prove that condition (4) is necessary.

Suppose self-reporting by all n injurers is a CPNE. It must be true that no *individual* injurer is better off deviating and not reporting, so $\frac{1}{n}\sum_{i=1}^{n} r_i \leq p_{n-1}$. Suppose that a coalition of *two or more* injurers deviates from the equilibrium and does not report. Let $m < n$ denote the number of injurers who are not part of the deviating coalition.[92] The injurers in the

[92]So the coalition has $n - m \geq 2$ members who deviate and do not self-report.

deviating coalition would pay an expected fine of p_m each, since the m injurers who are not part of the deviating coalition continue to self-report.

We will now verify that in any CPNE, $\frac{1}{m+1}\sum_{i=1}^{m+1} r_i \leq p_m$ for all $m = 1, ..., n-1$. There are two cases to consider.

(i) Suppose $\frac{1}{n}\sum_{i=1}^{n} r_i > p_m$, so the members of the deviating coalition pay a lower fine p_m by deviating. Since self-reporting by all n injurers is a CPNE, it must be the case that this is not self-enforcing. Thus, we require that an individual would prefer to abandon the coalition and join the group of m injurers who self-report: $\frac{1}{m+1}\sum_{i=1}^{m+1} r_i \leq p_m$. This is condition (4).

(ii) Suppose $\frac{1}{n}\sum_{i=1}^{n} r_i \leq p_m$, so the members of the deviating coalition would pay a weakly higher fine. Since $\{r_i\}_{i=1}^{n}$ is weakly increasing (by assumption), it must also be true that $\frac{1}{m+1}\sum_{i=1}^{m+1} r_i \leq p_m$. Again, this is condition (4). □

Claim 3. *Consider the set of ordered-leniency policies where $\{r_i\}_{i=1}^{n}$ is weakly increasing in i and satisfies condition (4), so self-reporting by all n injurers is a CPNE. An ordered-leniency policy within this set that leads to the highest expected fine is $\{r_1, r_2, ..., r_{\bar{m}}, r_{\bar{m}+1}, ..., r_n\} = \{x_1, x_2, ..., x_{\bar{m}}, 1, ..., 1\}$ where $\mathbf{x}$ and $\bar{m}$ are defined above.*

Proof of Claim 3. Suppose the ordered-leniency policy, $\mathbf{r}$, maximizes the sum of the leniency multipliers subject to (4) that $\sum_{i=1}^{m} r_i \leq m p_{m-1}$ and $r_m \in [0, 1]$ for all $m = 1, 2, ..., n$. This linear program may be written as follows.

$$\max_{\mathbf{r}} \sum_{i=1}^{n} r_i$$

subject to:

$$r_m \leq \min\left\{ m p_{m-1} - \sum_{i=1}^{m-1} r_i, 1 \right\}, \text{ for all } m = 1, 2, ...n.$$

We start by demonstrating that if $\mathbf{r}$ is a solution to this program, then there is another (possibly different) solution $\mathbf{r}'$ with the property that $r'_m = 1$ if and only if $m > \bar{m}$ for some value $\bar{m}$. Suppose that the vector $\mathbf{r}$ is a solution to the program, and suppose that $r_{m-1} = 1$ and $r_m < 1$ for some value m. Now consider a new vector $\mathbf{r}'$ that is identical to $\mathbf{r}$ except that two values are swapped: $r'_{m-1} = r_m < 1$ and $r'_m = r_{m-1} = 1$. Notice that expected fine associated with $\mathbf{r}'$ is the same as $\mathbf{r}$. We will now show that vector $\mathbf{r}'$ satisfies the system of equations. The only constraints we need to check are $m-1$ and m. First, consider constraint $m-1$: $r'_{m-1} \leq \min\left\{(m-1)p_{m-2} - \sum_{i=1}^{m-2} r_i, 1\right\}$. The right-hand side is the same with $\mathbf{r}'$ as with $\mathbf{r}$. Since $r'_{m-1} < r_{m-1}$, constraint $m-1$ is satisfied by the new vector $\mathbf{r}'$ too. Next, consider constraint m: $r'_m \leq \min\left\{m p_{m-1} - \sum_{i=1}^{m-2} r_i - r'_{m-1}, 1\right\}$. The right-hand side is different with $\mathbf{r}'$ than with $\mathbf{r}$, since $r'_{m-1} < r_{m-1}$. We have $r'_{m-1} < 1$ by assumption (since $r'_{m-1} = r_m < 1$). We also have $r'_m \leq m p_{m-1} - \sum_{i=1}^{m-2} r_i - r'_{m-1}$ since $r'_{m-1} + r'_m = r_{m-1} + r_m$.

Given the previous result, we may restrict attention to ordered leniency policies $\mathbf{r}$ where $r_m \leq m p_{m-1} - \sum_{i=1}^{m-1} r_i$ if $m \leq \widetilde{m}$ and $r_i = 1$ if $m > \widetilde{m}$, for some value $\widetilde{m}$. Importantly, constraint $\widetilde{m}$ must bind, since otherwise $r_{\widetilde{m}}$ could be raised without violating any constraint.

So, in the solution to the program, $r_{\widetilde{m}} = \widetilde{m}p_{\widetilde{m}-1} - \sum_{i=1}^{\widetilde{m}-1} r_i$ or equivalently $\sum_{i=1}^{\widetilde{m}} r_i = \widetilde{m}p_{\widetilde{m}-1}$. (Constraints $i = 1, ..., \widetilde{m} - 1$ need not bind and there are generally a continuum of solutions to the linear program, just as there are a continuum of solutions in Proposition 3 case 1.) The solution to the program will therefore have:

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{\widetilde{m}} r_i + \sum_{i=\widetilde{m}+1}^{n} 1 = \widetilde{m}p_{\widetilde{m}-1} + (n - \widetilde{m}).$$

We now make use of the definitions of the vector $\mathbf{x}$ and $\bar{m}$ above. Suppose that $r_i = x_i$ for all $i \leq \bar{m}$ and $r_i = 1$ for $i > \bar{m}$. This ordered-leniency policy satisfies all of the program's constraints, and has a higher total fine, $\bar{m}p_{\bar{m}-1} + (n - \bar{m})$. □

Claim 4. *Suppose $\{r_i\}_{i=1}^n$ is weakly increasing in i. Consider the set of ordered-leniency policies for which self-reporting by $n' < n$ injurers is a CPNE. The expected fine is smaller than the expected fine where all n injurers self-report.*

Proof of Claim 4. Consider an ordered-leniency policy where exactly $n' < n$ injurers self-report. A necessary condition for this to be a CPNE is that no *individual* injurer in the group that self reports is better off deviating: $\frac{1}{n'} \sum_{i=1}^{n'} r_i \leq p_{n'-1}$. More generally, there cannot be a self-enforcing deviation of a coalition of size $m' < n'$. Following the proof in Claim 2, a necessary condition for n' injurers to self report is $\frac{1}{m} \sum_{i=1}^{m} r_i \leq p_{m-1}$ for all $m = 1, 2, ..., n'$. Following the logic in Claim 3, the leniency multipliers for the n' injurers who self-report are $r_1 \leq p_0$ and $r_m \leq \min\{mp_{m-1} - (m-1)p_{m-2}, 1\}$ for $m = 2, ...n'$, and the fines for the injurers who remain silent are $p_{n'}$.

With the ordered-leniency policy where all n injurers self-report, the expected fines are weakly higher for all n injurers. Consider first the n' injurers who self report. Since $r_1 = p_0$ and $r_m = \min\{mp_{m-1} - (m-1)p_{m-2}, 1\}$ for all $m = 2, ...n$, the first n' injurers face weakly higher fines. Next, consider the $n-n'$ injurers who do not self report. With the ordered leniency policy where all n injurers self-report, the injurer in the $n' + 1$ position in the self-reporting queue pays $r_{n'+1} = \min\{(n'+1)p_{n'} - (n')p_{n'-1}, 1\}$. The first term in the brackets is equal to $p_{n'} + n'(p_{n'} - p_{n'-1})$ which is greater than $p_{n'}$. Since $mp_{m-1} - (m-1)p_{m-2}$ is an increasing function of m (by assumption), the fines paid by all of the injurers $i = n'+1, n'+2, ..., n$ are higher than $p_{n'}$ (the expected fine if exactly n' injurers self-report).

We conclude that any ordered-leniency policy where $n' < n$ injurers self-report has a lower expected fine than the ordered-leniency policy where all n injurers self-report.□

Since the optimal ordered-leniency policy involves a weakly-increasing sequence of leniency multipliers with at least one strict inequality, by Lemma 5, all n injurers report the act immediately, $t = 0$. Finally, since the equilibria of the self-reporting subgame do not depend on the level of the fine, f, the highest deterrence is obtained with the maximal fine, $f = \bar{f}$. Taken together, Claims 1–4 and the last result concerning the maximal fine have proved Proposition 5.■

Proof of Lemma 7. Taking the enforcement effort e as fixed, we have $p_0 = e$ and $p_m = e + (1-e)(1-(1-\pi)^m)$ for $m \in \{2, ..., n-1\}$. Using the expressions included in Proposition 5,

$$x_m = mp_{m-1} - (m-1)p_{m-2} =$$
$$= m[e + (1-e)(1-(1-\pi)^{m-1})] - (m-1)[e + (1-e)(1-(1-\pi)^{m-2})] =$$
$$= e + (1-e)[m - m(1-\pi)^{m-1} - (m-1) + (m-1)(1-\pi)^{m-2}] =$$
$$= e + (1-e)[1 - m(1-\pi)^{m-1} + (m-1)(1-\pi)^{m-2}] =$$
$$= e + (1-e)[1 - m(1-\pi)(1-\pi)^{m-2} + (m-1)(1-\pi)^{m-2}] =$$
$$= e + (1-e)[1 + [m - 1 - m(1-\pi)](1-\pi)^{m-2}] =$$
$$= e + (1-e)[1 - (1-m\pi)(1-\pi)^{m-2}].$$

So, we have $x_1 = e$ and

$$x_m = 1 - (1-e)(1-m\pi)(1-\pi)^{m-2}, \tag{6}$$

for all $m = 2, ..., n$. Notice that if $1 - m\pi > 0$ then $x_m < 1$, and if $1 - m\pi < 0$ then $x_m > 1$. Therefore, we have

$$\bar{m} = \sup\{m \in \{1, 2, ..., n\} | m < 1/\pi\}. \tag{7}$$

So, if $n \leq 1/\pi$, then some degree of leniency is given to all injurers who self-report, including the last injurer in the self-reporting queue. Taking the derivative of x_m with respect to m,

$$\frac{dx_m}{dm} = (1-e)\pi(1-\pi)^{m-2} - (1-e)(1-m\pi)\ln(1-\pi)(1-\pi)^{m-2}, \tag{8}$$

which has the same sign as $\pi - (1-m\pi)\ln(1-\pi)$. Since $\ln(1-\pi) < 0$, we have that $\frac{\partial x_m}{\partial m} > 0$ when $m \leq \bar{m}$ and $\frac{\partial x_m}{\partial m} < 0$ when $m > \bar{m}$. So, our convexity assumption that ip_{i-1} is increasing holds in the relevant range (for all $n \leq \bar{m}$).

The expression for $\hat{b}(e, \pi)$ in the lemma follows from substituting $p_{m-1} = e + (1-e)(1-(1-\pi)^{m-1})$ into the expression in Proposition 5. Taking the derivative with respect to e gives: $\frac{\partial \hat{b}(e,\pi)}{\partial e} = \frac{\bar{m}}{n}(1-\pi)^{\bar{m}-1}\bar{f} \in (0, \bar{f})$. ■