Optimal Law Enforcement with Ordered Leniency

Claudia M. Landeo *University of Alberta*
Kathryn E. Spier *Harvard University*

Abstract

This paper studies the design of optimal enforcement policies with ordered leniency to detect and deter harmful short-term activities committed by groups of injurers. With ordered leniency, the degree of leniency granted to an injurer who self-reports depends on his or her position in the self-reporting queue. We show that the ordered-leniency policy that induces maximal deterrence gives successively larger discounts to injurers who secure higher positions in the reporting queue. This creates a so-called race to the courthouse in which all injurers self-report promptly and, as a result, social harm is reduced. We show that the expected fine increases with the size of the group, which thus discourages the formation of large illegal enterprises. The first-best outcome is obtained with ordered leniency when the externalities associated with the harmful activities are not too great. Our findings complement Kaplow and Shavell's results for single-injurer environments.

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 Federal Bureau of Investigation 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 (Federal Bureau of Investigation 2012, pp. 6, 11). More generally,

Spier is also a research associate at the National Bureau of Economic Research. We acknowledge financial support from the National Science Foundation (grant SES-1155761). We thank the editor and the referees for insightful comments and suggestions. We also thank Tom Brennan, Dan Coquillette, Nick Feltovich, Nuno Garoupa, Mike Gilbert, John Goldberg, Andrew Hayashi, Christine Jolls, Louis Kaplow, Mike Meurer, Max Nikitin, Jack Ochs, Steve Shavell, Abe Wickelgren, and Kathy Zeiler for helpful discussions and comments. We are grateful for suggestions from conference and seminar participants at the 2019 annual meeting of the American Law and Economics Association, the 2018 NBER Summer Institute in Law and Economics, Harvard University, the University of Virginia, and Boston University. We thank Susan Norton for administrative assistance.

[*Journal of Law and Economics,* vol. 63 (February 2020)]

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 coconspirators 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's coconspirators was forced to disgorge his own profits and paid a penalty equal to three times his profits and all of the profits of his tippees. In contrast, a coconspirator who aided the prosecution paid a reduced penalty equal to one-half of his profits, while another coconspirator who cooperated early paid no penalty at all (Ceresney 2015; see also *Securities and Exchange Comm'n v. Saridakis,* No. 14-2397 [E.D. Pa. 2014]).

This paper adopts a normative economics approach (Caplin and Schotter 2010) and studies the design of optimal enforcement policies with ordered leniency to detect and deter illegal short-term activities committed by groups of injurers.[3] With an ordered-leniency policy, 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 queue. Our analysis demonstrates that the ordered-leniency policy that induces maximal deterrence gives successively larger discounts to injurers who secure higher positions in the self-reporting queue, creating a so-called race to the courthouse in which all injurers self-report promptly.[4] Prompt self-reporting also allows the enforcement agency to detect illegal activities sooner and, consequently, to mitigate the harms inflicted on others. We also show that the expected fine increases with the size of the group, which thus discourages the formation of large illegal enterprises.[5] The first-best outcome is obtained with ordered leniency when the externalities associated with the harmful activities are not too great. Although our paper is motivated by insider trading and securities fraud, our analysis applies to

[1] An example 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 or her fiduciary duty by revealing inside information. The term "tippee" refers to a person who knowingly uses inside information to make a trade.

[3] 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. See Motta and Polo (2003), Spagnolo (2005), Aubert, Rey, and Kovacic (2006), Harrington (2013), and Chen and Rey (2013). For a recent survey of this literature, see Marvão and Spagnolo (2018).

[4] 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.

[5] As discussed in Section 4, when the social harm inflicted on others increases with the size of the group, ordered leniency further reduces social harm by discouraging large groups.

any kind of harmful short-term activity committed by a group of wrongdoers.[6] To the best of our knowledge, there are no previous theoretical studies of optimal enforcement policies with ordered leniency 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 ordered leniency.[8] 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 or she 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 in which all injurers self-report, the risk-dominant equilibrium (Harsanyi and Selten 1988), and the other in which no injurer self-reports, the Pareto-dominant equilibrium. When the risk-dominance 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 in which all injurers self-report.

We demonstrate that the optimal enforcement policy with ordered leniency 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,[9] 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 can be socially valuable when the inculpatory evidence provided by the first injurer to report the act is insufficient to convict the second injurer with certainty. The optimal enforcement policy with ordered leniency creates a race to the courthouse in which, in equilibrium, both injurers self-report

[6] See Section 5 for a discussion of applications to other relevant contexts.

[7] See Landeo and Spier (2018b) for an experimental study on ordered-leniency mechanisms for short-term group activities.

[8] Our framework allows for sanction reductions that are increasing or decreasing in the injurer's position in the self-reporting queue. Later we demonstrate that the ordered-leniency policy that maximizes deterrence always involves more lenient treatment for the first injurer to self-report.

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

promptly. As a result, the likelihood of detection increases, expected sanctions rise, fewer harmful acts are committed, and the harm associated with the acts is reduced. Importantly, our findings suggest that the optimal enforcement policy with ordered leniency will achieve the first-best outcome when the externalities associated with the harmful activities (harms inflicted on others) are not too great.

We then study a general model involving groups of injurers with more than two members. Attention is restricted to coalition-proof Nash equilibria (Bernheim, Peleg, and Whinston 1987). The key insights of the benchmark model extend to this more general setting. The highest level of deterrence is achieved when the injurers 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 will not be full, and the leniency for the last injurer to report may not be 0. We show that the race-to-the-courthouse effect is robust to the number of members in the group of injurers. The injurers self-report promptly, and thus the social harm associated with the illegal activity is mitigated. New insights are derived as well. Our analysis demonstrates that the expected fine increases with the size of the group, and hence ordered-leniency policies discourage large illegal enterprises.

Finally, we investigate several relevant extensions of our benchmark model: the erroneous conviction of innocent parties, public information regarding the positions in the self-reporting queue, endogenous group size, stochastic detection rates, and rewards for self-reporting. Although these extensions raise new and interesting issues, the main lessons derived from our benchmark model remain relevant.

Our paper contributes to the theoretical 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.[10] The closest to our work are the studies on enforcement and self-reporting. Kaplow and Shavell (1994) study a probabilistic enforcement model in which 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 that is 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 (see also Malik 1993; Innes 1999; Livernois and McKenna 1999; Andreoni 1991; Malik and Schwab 1991). In contrast, we focus on harmful activities committed by groups of injurers. We show

[10] In seminal work, Becker (1968) demonstrates that a very small probability of detection coupled with a very large sanction can deter crime at essentially no cost. Polinsky and Shavell (1984) show that when injurers have limited assets and sanctions are bounded above, the optimal enforcement policy involves investigation costs, and deterrence falls short of the first-best level. See Garoupa (1997) for a survey of early theoretical work on law enforcement.

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, strengthens deterrence, and reduces social harm. In our environment, the 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 criminal teams with just two members. In their environment, self-reporting by an injurer provides enough evidence to convict the silent partner with certainty, the Pareto-dominance refinement applies in case of multiplicity of equilibria, and the injurers are not judgment proof. The degree of leniency granted depends on the number of injurers who self-report. Their optimal leniency policy grants immunity for self-reporting (a fine equal to 0) when exactly one injurer self-reports but grants (almost) no leniency when both injurers self-report.[11] The leniency mechanism studied in our paper is fundamentally different from that in Feess and Walzl (2004). In our framework, the first injurer to report the act receives leniency whether or not the other injurers also report. In particular, 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. With ordered leniency, there is a race to the courthouse in which injurers jockey for the first position in the self-reporting queue. In the environment in Feess and Walzl (2004), there is no advantage to being the first to report, and hence prompt self-reporting is not elicited and the harm inflicted on others is not always minimized. In our environment, the injurers report promptly, and the harm inflicted on others is reduced, thus increasing social welfare. In addition, our mechanism is arguably more closely aligned with how leniency policies for groups of wrongdoers are designed and implemented in the real world.[12]

Another strand of literature related to our paper is that on plea bargaining, in which 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 and screening effects.[13] Kobayashi (1992) studies plea bargaining using a model with two defendants in which 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 nego-

[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 in the self-reporting stage or the probability of cooperation is exogenous.

[12] See also Garoupa (2000), Buccirossi and Spagnolo (2006), Cooter and Garoupa (2014), and Piccolo and Immordino (2017) for theoretical work on organized crime and leniency policies.

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

tiated penalty for the most culpable defendant. More recently, Silva (2019) studies truth-telling mechanisms for groups of suspects when only one is guilty and finds that the optimal mechanism involves leniency for confession before investigation.[14] None of these papers consider ordered-leniency policies. 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 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, Ramseyer, and Wiley (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. Elsewhere (Landeo and Spier 2009, 2012) we provide experimental evidence of the exclusionary power of these types of contracts.[15]

The rest of the paper is organized as follows. Section 2 introduces the benchmark model, presents the equilibrium analysis, and identifies conditions under which the first-best outcome is achieved when enforcement policies with ordered leniency are implemented. Section 3 studies a general model that allows for groups of injurers with more than two members, demonstrates that the main insights derived from our benchmark model are robust, and provides additional important insights regarding the effect of group size on the expected fine. Section 4 presents relevant extensions. Section 5 discusses applications to other environments and concludes. Formal proofs are presented in the Appendix.

2. Benchmark Model

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 3 examines 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 commit-

[14] See also Siegel and Strulovici (2018) for a study of optimal deterrence with direct-revelation mechanisms for harmful acts committed by single injurers.

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

ting the harmful act. Without loss of generality, we abstract from time discounting.

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). First, $f \in (0, \overline{f}]$ denotes a fine or monetary sanction (measured per injurer).[16] The maximal fine, $\overline{f}$, can be greater than, lower than, or equal to the harm inflicted on others (measured per injurer), $h \in [\underline{h}, \overline{h}]$. Second, $r_1, r_2 \in [0, 1]$ 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 $r_1 = r_2$.[17] The discount for position i in the reporting queue is then $1 - r_i$ ($i = 1, 2$).[18] Thus, we study ordered-leniency policies in which the first injurer to report pays $r_1 f$, regardless of whether a second injurer reports, and the second injurer to report pays $r_2 f$. Third, $e \in [0, 1]$ denotes the enforcement agency's investigation effort, which, as we 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 $\lim_{e \to 1} c'(e) = \infty$.[19]

Second, after observing the enforcement policy, the potential injurers play a two-stage game. In stage 1, they decide jointly whether to commit a socially harmful act.[20] The private benefit from committing the act, measured per injurer, is $b \in [0, \infty)$ distributed according to probability density function $g(b)$ and cumulative distribution function $G(b)$. The realization of b is revealed to the potential injurers before they decide whether to commit the act.[21] With Coasean bargaining, the injurers will commit the act if their joint benefit, $2b$, exceeds the joint private cost (that is, expected fines that will be determined in equilibrium).[22] If they commit the act, stage 2 starts; otherwise, the game ends, and the payoff for each potential injurer is 0.

In stage 2, the injurers simultaneously and independently decide whether and when to report the harmful act to the enforcement agency. Each injurer can

[16] The term $\overline{f}$ can be interpreted as the potential injurer's wealth. When the fine is above $\overline{f}$, the injurer is judgment proof.

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

[18] If $(r_1, r_2) = (1, 1)$, then the enforcement policy does not grant leniency for self-reporting.

[19] These assumptions ensure an interior solution for the social welfare maximization problem and are standard in the literature on enforcement.

[20] Alternatively, one could assume that the injurers decide noncooperatively whether to participate in the activity and that the act is committed only if both injurers choose to participate.

[21] Committing the act is socially desirable if and only if the benefits, b, exceed the social harm, h.

[22] Bargaining may be modeled as a noncooperative game such as a Rubinstein bargaining protocol with alternating offers or a random-offeror protocol, among others. Our results are robust to the specification of the noncooperative game and to the allocation of bargaining surplus between the injurers (which may or may not be shared equally). Our framework accommodates asymmetric direct benefits from committing the act, b_1 and b_2. When side payments are possible, the injurers will commit the act if and only if $b_1 + b_2$ is greater than the sum of the expected fines. We do not allow the injurers to write forward contracts based on their future self-reporting decisions. Such contracts would not be enforceable in a court of law.

choose to report the act at time $t \in [0, 1]$, where $t = 0$ represents prompt reporting and $t > 0$ represents delayed reporting. We assume that $h = \overline{h}$ when both injurers decide not to self-report. We let $h = h(t_i)$ $(i = 1, 2)$ when only one injurer decides to self-report and $h = h(\min\{t_1, t_2\})$ when both injurers decide to self-report. Finally, we assume that $h(0) = \underline{h} > 0$, $h(1) = \overline{h}$, and $h'(t) > 0$. Intuitively, prompt self-reporting, and hence faster detection, allows the enforcer to mitigate the harm inflicted on others.[23]

Third, the injurers are detected 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 with probability p_0, and each injurer pays a fine f. If one injurer reports the act, then that injurer pays a fine $r_1 f$ and the silent accomplice is accurately detected and fully sanctioned (that is, pays a fine f) with probability p_1. If both injurers report the act, then the first to report pays $r_1 f$ and the second to report pays $r_2 f$. 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, $\pi \in (0, 1)$. In particular, $p_0(e) = e$ and $p_1(e, \pi) = e + (1 - e)\pi$.[24] It follows that $0 \leq p_0(e) < p_1(e, \pi) < 1$.

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).[25]

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 and promptly identify the parties responsible for committing the harmful act (and their private benefits) and decide which acts to prohibit.[26] Hence, in the first-best outcome, the harm inflicted on others is minimized, $h = \underline{h}$, the enforcement effort is $e = 0$ and acts are committed if and only if the benefit associated with the act is greater than the harm inflicted on others, $b > \underline{h}$.

[23] Our main results will hold even if we abstract from the relationship between h and t. However, we decided to include h in our framework to underscore the rationale behind the enforcement agency's goal of inducing prompt self-reporting.

[24] 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 a probability of success of $p_0 = e$. When one injurer self-reports and another does not, there is a second independent Bernoulli trial that succeeds in detecting the nonreporting injurer with probability π. Then $p_1 = e + (1 - e)\pi$ is the probability that the silent injurer is detected. Our main results regarding the characterization of the fine and leniency multipliers that create maximal deterrence for groups of potential injurers will hold for more general specifications.

[25] Previous literature on the design of institutions in complex strategic environments uses a similar approach. See, for instance, Che and Yoo (2001) and Feess and Walzl (2004).

[26] In practice, of course, the enforcement agency cannot costlessly and promptly 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.

We apply backward induction and begin our analysis with the injurers' decisions. We then analyze the optimal enforcement policy with ordered leniency.

2.1. *Injurers' Decisions*

We first characterize the equilibrium behavior of the injurers in stage 2, the self-reporting stage. Second, we study the potential injurers' joint decision to commit the act in stage 1.

2.1.1. Decision to Report the Act and Time to Report

If the act is committed in stage 1, then stage 2 occurs. In stage 2, the injurers simultaneously and independently decide whether and when to report the harmful act to the enforcement authority. That is, 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. Below we verify that optimal enforcement policies with ordered leniency require $r_1 < r_2$. Lemma 1 characterizes the equilibrium reporting time.

Lemma 1. If $r_1 < r_2$, then an injurer who reports the act will do so promptly, $t = 0$. 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. Intuitively, a race to the courthouse in which all injurers who self-report the act will do so promptly occurs only when the first injurer to report is granted a larger penalty reduction than the second injurer to self-report ($r_1 < r_2$).[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]

Second, we study the injurers' decisions about whether to report the act. The strategic-form representation of the self-reporting subgame is presented in Table 1. If neither injurer self-reports, then the act is detected with probability p_0, and each injurer pays an expected fine of $p_0 f$. 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

[27] If injurer j believes that injurer $-j$ will not report at all, then injurer j is just as well off reporting promptly as delaying. However, if injurer j believes that there is a nonzero chance that injurer $-j$ will report at time $t = 0$, then injurer j is strictly better off reporting promptly as well. In other words, late reporting is a weakly dominated strategy. If instead $r_1 > r_2$, then early reporting is a weakly dominated strategy. If injurer j believes that there is a nonzero chance that injurer $-j$ will report the act at $t = 1$, then injurer j 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 to self-report, and the injurers are indifferent about the reporting time.

[28] When $r_1 < r_2$, self-reporting occurs promptly 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.

Table 1
Strategic-Form Representation of the Self-Reporting Subgame: Continuation Payoffs

	No Report	Report
No report	$-p_0 f, -p_0 f$	$-p_1 f, -r_1 f$
Report	$-r_1 f, -p_1 f$	$-[(r_1 + r_2)/2]f, -[(r_1 + r_2)/2]f$

accomplice pays $p_1 f$ in expectation. 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 pays an expected fine of $[(r_1 + r_2)/2]f$.[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 given in the following cases:

1) If $r_1 \leq p_0$ and $(r_1 + r_2)/2 \leq p_1$, then there is a unique pure-strategy Nash equilibrium in which both injurers self-report, (R, R).

2) If $r_1 \leq p_0$ and $(r_1 + r_2)/2 > p_1$, then there are two pure-strategy Nash equilibria in which one injurer self-reports, (R, NR) and (NR, R).

3) If $r_1 > p_0$ and $(r_1 + r_2)/2 \leq p_1$, then there are two pure-strategy Nash equilibria, one in which both injurers self-report and one in which neither injurer self-reports. Nash equilibrium (R, R) Pareto dominates (NR, NR) if and only if $(r_1 + r_2)/2 \leq p_0$, and (R, R) risk dominates (NR, NR) if and only if $(3r_1 + r_2)/4 \leq (p_0 + p_1)/2$.

4) If $r_1 > p_0$ and $(r_1 + r_2)/2 > p_1$, then there is a unique pure-strategy Nash equilibrium in which 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 strategies.[30] When the expected sanction for self-reporting is not too small, $[(r_1 + r_2)/2]f > p_0 f$, 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), in which 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 fine associated with remaining silent, $p_0 f$, is smaller than

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

[30] The second Nash equilibrium in which both injurers decide not to report, (NR, NR), does not survive the elimination of weakly dominated strategies.

[31] If $[(r_1 + r_2)/2]f > p_0 f$, 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 in which both injurers decide not to report, (NR, NR).

the fine 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 $[(r_1 + r_2)/2]$ 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, the no-reporting strategy 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): both equilibria satisfy Pareto and risk dominance. In contrast, the two pure-strategy Nash equilibria that arise in case 3, (R, R) and (NR, NR), may be ranked using the Pareto- and risk-dominance refinements. When $(r_1 + r_2)/2 \leq p_0$, the expected fine is lower when both injurers report committing the act. So (R, R) is the Pareto-dominant Nash equilibrium if and only if $(r_1 + r_2)/2 \leq p_0$. When $(3r_1 + r_2)/4 \leq (p_0 + p_1)/2$, an injurer would prefer to self-report when there is a 50 percent chance that the other injurer will also report. Thus, (R, R) is the risk-dominant Nash equilibrium if and only if $(3r_1 + r_2)/4 \leq (p_0 + p_1)/2$.

2.1.2. Decision to Commit the Act

In stage 1, the potential injurers decide whether to commit the socially harmful act. They commit the act when their joint private benefit from doing so is greater than the sum of the expected fines (which are determined in the stage 2 continuation game).[33] Let $\hat{b}$ denote the expected fine or deterrence threshold, measured per injurer, in the stage 2 continuation game. When the benefit of committing the act, b, is greater than the deterrence threshold, $\hat{b}$, then the injurer will choose to commit the act. Conversely, when b is smaller than or equal to the deterrence threshold, $\hat{b}$, the injurers will choose not to commit the act.[34] The deterrence thresholds may be constructed using lemma 2. Lemma 3 characterizes the equilibrium decision to commit the act in stage 1. Cases 1–4 correspond to cases 1–4 in lemma 2.

Lemma 3. Take the fine f and the detection probabilities, p_0 and p_1, as fixed. The potential injurers will commit the act in the following cases:

1) If $r_1 \leq p_0$ and $(r_1 + r_1)/2 \leq p_1$, then the injurers commit the act if and only if $b > \hat{b} = [(r_1 + r_2)/2]f$.

2) If $r_1 \leq p_0$ and $(r_1 + r_2)/2 > p_1$, then the injurers commit the act if and only if $b > \hat{b} = [(r_1 + p_1)/2]f$.

3) If $r_1 > p_0$ and $(r_1 + r_2)/2 \leq p_1$, if $(r_1 + r_2)/2 \leq p_0$ (Pareto dominance) or $(3r_1 + r_2)/4 \leq (p_0 + p_1)/2$ (risk dominance), then the injurers commit the act if and only if $b > \hat{b} = [(r_1 + r_2)/2]f$. If $(r_1 + r_2)/2 > p_0$ (Pareto dominance) or $(3r_1 +$

[33] If the joint benefit exceeds the sum of expected fines, then the potential injurers will negotiate a division of the joint surplus and commit the act. Through Coasean bargaining, both injurers are willing to participate in the harmful activity, and joint value is maximized.

[34] When $b = \hat{b}$, the injurers are indifferent, and we assume, without loss of generality, that that they do not commit the act.

$r_2)/4 > (p_0 + p_1)/2$ (risk dominance), then the injurers commit the act if and only if $b > \hat{b} = p_0 f$.

4) If $r_1 > p_0$ and $(r_1 + r_2)/2 > p_1$, then the injurers commit the act if and only if $b > \hat{b} = p_0 f$.

In case 1 of lemma 3, since both injurers self-report in the unique Nash equilibrium, the deterrence threshold is $\hat{b} = [(r_1 + r_2)/2]f$. Thus, the potential injurers commit the harmful act when $b > \hat{b} = [(r_1 + r_2)/2]f$. In case 2, there are two Nash equilibria, (R, NR) and (NR, R). In both equilibria, the sum of the fines is $(r_1 + p_1)f$, so the deterrence threshold is $\hat{b} = [(r_1 + p_1)/2]f$. Hence, the potential injurers commit the act in case 2 if and only if $b > \hat{b} = [(r_1 + p_1)/2]f$. In case 3, where multiple equilibria also arise, the Pareto- or risk-dominance refinements will determine which of the two outcomes is obtained, (R, R) or (NR, NR), and so the deterrence threshold is either $\hat{b} = [(r_1 + r_2)/2]f$ or $\hat{b} = p_0 f$. Hence, the injurers will commit the act when $b > \hat{b} = [(r_1 + r_2)/2]f$ or $b > \hat{b} = p_0 f$, depending on the equilibrium refinement. Finally, in case 4, since neither injurer self-reports in equilibrium, the deterrence threshold is $\hat{b} = p_0 f$, and the injurers commit the act 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 $p_0 \overline{f}$. 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.

2.2. Optimal Enforcement with Ordered Leniency

In characterizing the optimal enforcement policy with ordered leniency, we first 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 (that is, the highest expected fine). We also show that the harm inflicted on others is minimized when ordered leniency is implemented. Second, we demonstrate that the first-best deterrence outcome may be achieved with ordered-leniency policies at an arbitrarily low enforcement cost when the externalities associated with the harmful activities are not too great.

2.2.1. Maximal Deterrence

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 (that is, highest expected fine). We will demonstrate that the fine should be set at the maximal level, $\overline{f}$, and that the ordered-leniency policies that implement maximal deterrence give greater leniency to the first injurer to report and induce prompt 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 it will be milder (larger multipliers) under the risk-dominance refinement.[35]

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 1 characterizes the fine and leniency multipliers that create maximal deterrence for groups of potential injurers.

Proposition 1. Take the enforcement effort e as fixed. Maximal deterrence is obtained with a maximal fine, $f = \overline{f}$, and the following leniency multipliers:[36]

1) If $p_1 \leq (1 + p_0)/2$, 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 $t = 0$ if $b > \hat{b}^S = \hat{b}^M = p_1\overline{f}$ and do not commit the act otherwise.

2) If $p_1 > (1 + p_0)/2$, then $(r_1^S, r_2^S) = (p_0, 1)$ and $(r_1^M, r_2^M) = \{[2(p_0 + p_1) - 1]/3, 1\}$. The injurers commit the act and self-report at $t = 0$ if $b > \hat{b}^S = [(1 + p_0)/2]\overline{f}$ (Pareto dominance) and $b > \hat{b}^M = [(1 + p_0 + p_1)/3]\overline{f}$ (risk dominance), where $\hat{b}^S < \hat{b}^M$, and do not commit the act otherwise.

Proposition 1 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 greatest deterrence is obtained by imposing the maximal fine, $f = \overline{f}$. 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.

[35] It is simple to show that enforcement policies with ordered leniency for self-reporting always outperform enforcement policies without leniency for self-reporting. Without leniency, the optimal enforcement policy does not incentivize the injurers to self-report. With ordered leniency, and holding enforcement efforts fixed, the enforcement agency can raise the expected fines by inducing both injurers to self-report and hence achieve a higher level of deterrence. See Landeo and Spier (2018a) for formal analysis.

[36] When $p_1 \leq (1 + p_0)/2$, the leniency multipliers are not unique.

Maximal deterrence is achieved when both injurers self-report. It is obvious that a leniency policy in which at least one injurer self-reports creates stronger deterrence than a policy in which 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). In particular, if $p_1 \geq (r_1 + r_2)/2 = (1 + p_0)/2$, then case 1 of lemmas 2 and 3 applies, as both injurers self-report, and the expected fine is $[(1 + p_0)/2]\bar{f} > p_0\bar{f}$. On the other hand, if $p_1 < (r_1 + r_2)/2 = (1 + p_0)/2$, then case 2 of lemmas 2 and 3 apply, as exactly one injurer self-reports, and the expected fine is $[(p_0 + p_1)/2]\bar{f} > p_0\bar{f}$. In this case, in which 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 in which both injurers self-report, and the expected fine rises to $p_1\bar{f}$.[37]

Remark 3. The first injurer to self-report always receives more lenient treatment.

Suppose that $p_1 \geq (1 + p_0)/2$ and $(r_1, r_2) = (p_0, 1)$. Case 1 of lemma 2 applies, as both injurers self-report. Rewarding the first injurer creates a proverbial race to the courthouse between the two injurers, and the expected fine is $[(1 + p_0)/2]\bar{f} > p_0\bar{f}$.[38] If the multipliers were reversed, so $(r_1, r_2) = (1, p_0)$ (that is, 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.[39] 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 inculpatory evidence is weak, then the second injurer to report the act receives leniency too. To see why, suppose that $p_1 < (1 + p_0)/2$. If leniency is granted only to the first injurer, $(r_1, r_2) = (p_0, 1)$, then case 2 of lemmas 2 and 3 applies, as only one injurer reports the act and the other remains silent, and the deterrence threshold is $[(p_0 + p_1)/2]\bar{f}$. 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 $p_1\bar{f}$.[40] Deterrence is stronger when the second injurer also receives leniency.

Remark 5. Stronger deterrence is obtained with the risk-dominance refinement.

[37] According to case 1 of proposition 1, this policy maximizes deterrence ($\Delta = p_1 - p_0$).

[38] If $p_1 \geq (1 + p_0)/2$, then only one injurer would self-report, and the expected fine would still be strictly higher than $p_0\bar{f}$.

[39] 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.

[40] When $p_1 > \frac{1}{2}$, 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 $p_1 < \frac{1}{2}$, however, $2p_1 - 1$ is a negative number. Some degree of leniency must be granted to the second injurer too.

Proposition 1 implies that the deterrence threshold is never lower, and may be higher, when the risk-dominance refinement is applied in the self-reporting subgame.[41] In case 1 of proposition 1, when $p_1 \leq (1 + p_0)/2$ (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 case 2 of proposition 1, when $p_1 > (1 + p_0)/2$ (strong inculpatory evidence), the optimal leniency multipliers under the two equilibrium refinements diverge. Suppose that the enforcement agency chooses the mild leniency policy, $(r_1^M, r_2^M) = \{[2(p_0 + p_1) - 1]/3, 1\}$.[42] 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 $\hat{b}^M = [(1 + p_0 + p_1)/3]\bar{f} > p_0\bar{f}$. When the Pareto-dominance refinement is applied in the self-reporting subgame, neither injurer self-reports, and the deterrence threshold is $p_0\bar{f}$. So 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.[43] The resulting deterrence threshold is $\hat{b}^S = [(1 + p_0)/2]\bar{f} < \hat{b}^M$. 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 increase.

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 1, 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 $p_1 = .4 < (1 + p_0)/2$, so the likelihood of catching a silent conspirator is relatively low. Granting leniency to the second injurer who reports the act is necessary. Proposition 1 implies that deterrence is maximal when the enforcement agency grants leniency $r_1^S = r_1^M = p_0 = .2$ and $r_2^S = r_2^M = (2p_1 - p_0)/2 = .6$ to the first and second injurer to report.[44] The injurers race to be the first in line, and the expected sanction increases to $\hat{b}^S = \hat{b}^M = p_1\bar{f} = .4$.

[41] As demonstrated in the Appendix (proof of proposition 1), the leniency multipliers under Pareto dominance, (r_1^S, r_2^S), satisfy the conditions stated in case 1 of lemma 2. When $p_1 \leq (1 + p_0)/2$, the leniency multipliers under risk dominance, (r_1^M, r_2^M), satisfy either the conditions stated in case 1 of lemma 2 or those in case 3 of lemma 2 (both provide the same level of deterrence); when $p_1 > (1 + p_0)/2$, the leniency multiplier under risk dominance, (r_1^M, r_2^M), satisfies the conditions stated in case 3 of lemma 2.

[42] 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.

[43] This new strategic environment corresponds to case 1 of lemma 2, in which (R, R) is the unique Nash equilibrium.

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

Suppose instead that $p_1 = .75 > (1 + p_0)/2$, 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 promptly, and the expected fine is $\hat{b}^S = [(1 + p_0)/2]\overline{f} = .6$.

When $p_1 = .75$ and risk dominance is applied in the self-reporting subgame, deterrence can be made even greater by increasing the leniency multiplier for the first injurer to $r_1^M = [2(p_0 + p_1) - 1]/3 = .3$. 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 promptly, and the expected fine rises to $\hat{b}^M = [(1 + p_0 + p_1)/3]\overline{f} = .65$.[45]

Finally, corollary 1 summarizes an important result regarding the harm inflicted on others when ordered-leniency policies are implemented:

Corollary 1. Ordered-leniency policies that generate maximal deterrence also minimize the harm inflicted on others conditional on acts being committed.

As stated in proposition 1, ordered-leniency policies that generate maximal deterrence induce both injurers to self-report promptly ($t = 0$). As a result, the social harm is reduced: $h = \underline{h}$.

2.2.2. Optimal Enforcement Effort

We now characterize the optimal enforcement effort, e. Remember that proposition 1 identifies the leniency multipliers and fine that create maximal deterrence (that is, the highest expected fine), and that S and M denote the leniency policies under the Pareto- and risk-dominance refinements, respectively.

Lemma 4, which follows from proposition 1, characterizes the expected functions of fines 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. Note that $p_1 \leq (1 + p_0)/2$ holds if and only if $\pi \leq \frac{1}{2}$, and $p_1 > (1 + p_0)/2$ holds if and only if $\pi > \frac{1}{2}$. In other words, cases 1 and 2 of lemma 4 correspond to cases 1 and 2 of proposition 1.[46]

Lemma 4. The ordered-leniency multipliers (r_1^S, r_2^S) and (r_1^M, r_2^M), characterized in proposition 1, yield corresponding expected fines $\hat{b}^S(e, \pi)$ and

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

[46] Consider case 1 of proposition 1, in which $p_1 \leq (1 + p_0)/2$. Substituting $p_0 = e$ and $p_1 = e + (1 - e)\pi$, we find that $p_1 \leq (1 + p_0)/2$ holds if and only if $\pi \leq \frac{1}{2}$. Similar logic applies to case 2 of proposition 1.

$\hat{b}^{M}(e, \pi)$ for the injurers. These functions, which are continuous and piecewise differentiable, satisfy the following cases:

1) If $\pi \leq \frac{1}{2}$, then $\hat{b}^{S}(e, \pi) = \hat{b}^{M}(e, \pi) = [\pi + (1-\pi)e]\overline{f}$ and $0 < \partial \hat{b}^{i}(e, \pi)/\partial e < \overline{f}$ for $i = S, M$.

2) If $\pi > \frac{1}{2}$, then $\hat{b}^{S}(e, \pi) = [(1+e)/2]\overline{f}$ and $\hat{b}^{M}(e, \pi) = \{[(1+\pi) + (2-\pi)e]/3\}\overline{f}$. Furthermore, $\hat{b}^{S}(e, \pi) < \hat{b}^{M}(e, \pi)$ and $0 < \partial \hat{b}^{M}(e, \pi)/\partial e < \partial \hat{b}^{S}(e, \pi)/\partial e < \overline{f}$.

Next we analyze the optimal enforcement effort e. Recall that, in the first-best outcome, the injurers commit the act if and only if the benefit exceeds the social harm, $b > \underline{h}$, and no effort is spent on enforcement, $e = 0$. Proposition 2 establishes necessary and sufficient conditions under which the enforcement agency can implement the first-best deterrence outcome with an ordered-leniency policy at (almost) no cost[47] and describes the enforcement policy that implements the second-best outcome when the first-best outcome cannot be achieved. It also underscores that the socially optimal level of harm is always achieved with an ordered-leniency policy.

Proposition 2. An optimal enforcement policy with ordered leniency for self-reporting can implement the first-best deterrence outcome at (almost) no cost if and only if $\underline{h} \leq \hat{b}^{S}(0, \pi) = \min\{\pi, \frac{1}{2}\}\overline{f}$ under the Pareto-dominance refinement and $\underline{h} \leq \hat{b}^{M}(0, \pi) = \min\{\pi, (1+\pi)/3\}\overline{f}$ under the risk-dominance refinement. When $\underline{h} > \hat{b}^{i}(0, \pi)$, $i = S, M$, the second-best deterrence outcome involves higher enforcement costs and underdeterrence relative to the first best. The socially optimal level of harm $\underline{h}$ is always implemented.

When $\pi \leq \frac{1}{2}$ (weak inculpatory evidence), case 1 of lemma 4 applies. With (almost) no enforcement effort, the maximal deterrence and minimal harm are obtained with a maximal fine $\overline{f}$ 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 $b \leq \hat{b}^{S} = \hat{b}^{M} = \pi\overline{f}$. Note that if the level of harm is less than the deterrence threshold, $\underline{h} < \pi\overline{f}$, 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. If the level of harm is exactly equal to the deterrence threshold, $\underline{h} = \pi\overline{f}$, then the injurers will commit the act if and only if $b > \underline{h}$, as desired. If the level of harm exceeds the deterrence threshold, $\underline{h} > \pi\overline{f}$, then there is underdeterrence relative to the first-best level. In this case, deterrence can be improved by spending more resources on enforcement. Hence, when $\pi \leq \frac{1}{2}$, the first-best outcome is achieved at (almost) no cost if and only if the social harm is not too high, $\underline{h} \leq \pi\overline{f}$.

When $\pi > \frac{1}{2}$ (strong inculpatory evidence), case 2 of lemma 4 applies. Suppose there is no enforcement effort, $e = 0$. When the Pareto-dominance refinement is applied to the self-reporting subgame, the multipliers that create maximal deter-

[47] Almost no effort refers to $e = 0 + \varepsilon$, and almost no cost refers to $c(0 + \varepsilon)$, where $\varepsilon > 0$ is an arbitrarily small number. For simplicity, and without loss of generality, we abstract from ε for the rest of the paper.

rence are $(r_1^S, r_2^S) = (0, 1)$, and the associated deterrence threshold is $\hat{b}^S = \frac{1}{2}\overline{f}$. If the level of harm is below this threshold, $\underline{h} < \frac{1}{2}\overline{f}$, then the first-best outcome may be obtained by lowering the fine, lowering the leniency multiplier for the second injurer, or both. When the risk-dominance refinement applies, the leniency multipliers that create the maximal deterrence are $(r_1^M, r_2^M) = [(2\pi - 1)/3, 1]$, and the associated deterrence threshold is $\hat{b}^M = [(1 + \pi) / 3]\overline{f}$. Applying the same logic as before, if $\underline{h} < [(1 + \pi)/3]\overline{f}$, then 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 (almost) no cost if and only if the harm is not too great, $\underline{h} \leq \frac{1}{2}\overline{f}$ (Pareto dominance) and $\underline{h} \leq [(1 + \pi)/3]\overline{f}$ (risk dominance).

Taken together, our findings provide a social welfare rationale for the use of ordered-leniency policies. First, 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 race to the courthouse in which all injurers report the act promptly (proposition 1). As a result, the harm inflicted on others is minimized (corollary 1). Second, we demonstrated that the socially optimal level of deterrence can be obtained at an arbitrarily low cost when the externalities associated with the harmful activities are not too great and that the socially optimal level of harm is always implemented (proposition 2). Our findings regarding enforcement policies with ordered leniency for groups of injurers complement the results for single-injurer environments in Kaplow and Shavell (1994).

3. General Model

This section generalizes our benchmark framework by allowing for groups of injurers with more than two members. Our analysis demonstrates that the key insights of the benchmark model extend to this setting. We also derive important new insights regarding the effect of group size on the expected fine.

The strategic environment now consists of a game of complete information with the following risk-neutral players: two or more 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). The term $f \in (0, \overline{f}]$ denotes the fine. As before, $\overline{f}$ can be greater than, lower than, or equal to the harm inflicted on others (measured per injurer), $h \in [\underline{h}, \overline{h}]$. The term $\boldsymbol{r} = \{r_i\}_{i=1}^n$ 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.[48] Finally, $e \in [0, 1)$

[48] We show below 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.

is the enforcement agency's effort, and $c(e)$ is the cost of enforcement (measured per injurer).[49]

Second, after observing the enforcement policy, the potential injurers play a two-stage game. In stage 1, they decide whether to commit the act. The private benefit, measured per injurer, $b \in [0, \infty)$,[50] is revealed to the injurers before they make their decision regarding committing the act. As in our benchmark model, Coasean bargaining assures that they will commit the act when their joint benefit exceeds the sum of the expected fines.[51] 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. We assume that the social harm, measured per injurer, is $h = \overline{h}$ when all injurers decide not to self-report. We let $h = h(t_i)$ $(i = 1, \ldots, n)$ when only one injurer decides to self-report and $h = h(\min\{t_1, \ldots, t_j\})$ when j injurers $(1 < j \leq n)$ decide to self-report. Finally, we assume that $h(0) = \underline{h}$, $h(1) = \overline{h}$, and $h'(t) > 0$.

Third, the injurers are detected by the enforcement agency and sanctioned. We let p_i for $i = 1, \ldots, n$ be the probability that a silent injurer will be detected and sanctioned when exactly i injurers self-report. We assume that $0 \leq p_0 < p_1 < \ldots < p_{n-1} < 1$, 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.[52] 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, $\pi \in (0, 1)$. In particular, we let $p_0(e) = e$ and $p_i(e, \pi) = e + (1 - e)[1 - (1 - \pi)^i]$ for $i = 1, \ldots, n - 1$.[53]

The equilibrium concept is subgame-perfect Nash equilibrium. As in our benchmark model, multiple equilibria may arise in the self-reporting subgame.

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

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

[51] Bargaining may be modeled as a noncooperative divide-the-dollar game with alternating offers, a random-offeror game, or another bargaining protocol. Forward contracts that are contingent on future self-reporting are not allowed (and indeed, would not be enforceable in a court of law).

[52] 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 when $p_i = i/(i + 1)$. Convexity simplifies the characterization of the optimal ordered-leniency policy in proposition 3.

[53] The chance that a silent conspirator will evade detection if i conspirators self-report is $(1 - \pi)^i$. Thus, the chance that the silent conspirator is detected and sanctioned is $1 - (1 - \pi)^i$. 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 a probability of success $p_0 = e$. 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 that 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.

We restrict attention to coalition-proof Nash equilibria (CPNE) (Bernheim, Peleg, and Whinston 1987).[54]

The first-best outcome is used as a benchmark in the welfare analysis of ordered-leniency policies. In the first-best outcome, the cost of effort is 0, and acts are committed if and only if $b > \underline{h}$.

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

3.1. *Injurers' Decisions*

We first analyze the length of time taken by the injurers to report the act. Lemma 5 characterizes the equilibrium reporting 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 promptly, at $t = 0$. 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.[55] Lemma 5 implies that, except for the knife-edged case in which $\{r_i\}_{i=1}^n$ is constant for all i, the injurers who report the act will either all self-report promptly or 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 $(1/m)\sum_{i=1}^m r_i f$.

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.[56] We assume that, when indifferent, the player reports the act.

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

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

[54] 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 coalition-proof Nash equilibria (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.

[55] 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 promptly. 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.

[56] As we show in the proof of proposition 3, condition (1) is also a necessary and sufficient condition for self-reporting by all n injurers to be a CPNE when $\{r_i\}_{i=1}^n$ is weakly increasing in i. See the Appendix.

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

If condition (1) 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, $(1/n)\sum_{i=1}^{n} r_i f$, is smaller than the 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 $[1/(n-1)]\sum_{i=1}^{n-1} r_i f \leq p_{n-2}f$ according to condition (1). Following the same logic, no coalition of any size m can deviate in a way that is mutually self-enforcing.

Finally, we analyze the potential injurers' decision to commit the harmful act. Lemma 7 describes the injurers' equilibrium decision 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 (1) holds. The potential injurers commit the act if and only if $b > \hat{b} = (1/n)\sum_{i=1}^{n} r_i f$.

As in our benchmark model, the potential injurers will commit the harmful act if and only if the private benefit from committing the act (measured per injurer), b, is greater than the deterrence threshold $\hat{b}$ (the expected fine).

3.2. Optimal Enforcement with Ordered Leniency

We begin by taking the enforcement effort e and the corresponding probabilities of detection $\{p_i\}_{i=0}^{n-1}$ as fixed and identify the fine and the leniency multipliers that create the greatest possible deterrence (that is, the highest expected fine). Formally, the enforcement agency seeks to maximize the expected fine, $(1/n)\sum_{i=1}^{n} r_i$, subject to condition (1) and $r_m \in [0, 1]$ ($\forall\, m = 1, 2, \ldots, n$).

Proposition 3 characterizes the fine and leniency multipliers that create maximal deterrence:

Proposition 3. Take the enforcement effort e as fixed. Maximal deterrence is obtained with a maximal fine, $f = \overline{f}$, and the leniency multipliers $r_1 = p_0$ and $r_m = \min\{mp_{m-1} - (m-1)p_{m-2}, 1\}$ for $m = 2, \ldots, n$, where $r_1 < r_2 \leq \ldots \leq r_n$. The injurers commit the act and self-report at $t = 0$ if $b > \hat{b} = (1/n)\sum_{i=1}^{n} r_i = [(\overline{m}/n)p_{\overline{m}-1} + (n - \overline{m})/n]\overline{f}$, where $\overline{m} = \sup\{m \in \{1, \ldots, n\} \mid r_m < 1\}$, and do not commit the act otherwise.

Proposition 3 verifies the robustness of our previous findings to environments with two or more injurers (see proposition 1 and corollary 1). First, the greatest deterrence is achieved when the fine is set at the maximal level, $f = \overline{f}$. Second, the highest level of deterrence is achieved by implementing a weakly increasing ordered-leniency policy that gives injurers successive discounts for self-reporting on the basis of 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).[57] Third, all injurers self-report promptly in the CPNE. In other words, the optimal policy generates a race to the courthouse among the injurers. Importantly, as a result of prompt self-reporting, the harm inflicted on others conditional on acts being committed is minimized, $h = \underline{h}$.

Next we present a numerical example to illustrate our findings:

Example 2. Suppose that the maximal fine is $\overline{f} = 1$. Suppose also the group includes three members 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.[58] 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 $(r_1, r_2, r_3) = (.2, .6, .85)$.[59] In equilibrium, there is a race to the courthouse in which the three injurers self-report promptly. The expected fine is .55.[60] Suppose instead that the group includes four injurers and $(p_0, p_1, p_2, p_3) = (.2, .4, .55, .6625)$. The leniency multipliers that generate maximal deterrence are $(r_1, r_2, r_3, r_4) = (.2, .6, .85, 1)$.[61] In equilibrium, there is a race to the courthouse in which the four injurers self-report promptly. The expected fine is .6625.[62]

We derive important new insights regarding the effect of group size on the expected fine. 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, we find that when the group size grows, more coconspirators will report the act. As a result, the likelihood of detection and the expected fine will rise. More formally, suppose that $x_n < 1$; that is, leniency is also granted to the last position in the self-reporting queue. Since $\overline{m} = n$, the expected fine for a member of a group of size n is $\hat{b} = p_{n-1}\overline{f}$. As the size of the group increases, the probability p_{n-1} increases, and so the expected fine increases too. Suppose instead that $x_n > 1$. So $\overline{m} < n$. In this case, $\overline{m}$ does not change as the size of the group grows. Since $p_{\overline{m}-1} < 1$, the expected fine $\hat{b}$ increases as n increases. Corollary 2 summarizes this finding:

Corollary 2. The expected fine faced by an injurer, $\hat{b}$, strictly increases with the size of the group, n.

Intuitively, when optimal enforcement with ordered leniency is implemented, as the size of the group increases, the negative externalities in the self-reporting subgame rise, and the expected fine faced by an injurer increases. If the size of the

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

[58] The likelihood of detection of silent injurers is $p_1 = .4$. The expected fine is $(.2 + .4 + .4)/3 = \frac{1}{3}$.

[59] Note that $\overline{m} = 3 = n$.

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

[61] Note that $\overline{m} = 3 < n$.

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

group, n, is endogenously decided by the potential injurers, the choice of a large group would be strictly dominated by the choice of a small group. In other words, 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. In some real-world environments, the social harm might increase with the number of injurers involved in the activity.[63] Our previous result suggests that an additional social benefit of ordered leniency will be a further reduction in the harm inflicted on others.

Lemma 8, which follows from proposition 3, characterizes the expected fine function $\hat{b}(e, \pi)$. 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 $p_i(e, \pi) = e + (1 - e)[1 - (1 - \pi)^i]$ is the probability of detection if $i \in \{1, \ldots, n - 1\}$ injurers self-report.[64]

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 $m = 2, \ldots, n$. The expected fine is $\hat{b}(e, \pi) = [1 - (\overline{m}/n)(1 - e)(1 - \pi)^{\overline{m}-1}]\overline{f}$, where $\overline{m} = \sup\{m \in \{1, \ldots, n\} \mid m < 1/\pi\}$. The expected fine is continuous, piecewise differentiable, and satisfies $0 < \partial\hat{b}(e, \pi)/\partial e < \overline{f}$ and $\partial\hat{b}(e, \pi)/\partial\pi > 0$.

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

Finally, we analyze the optimal enforcement effort e. Proposition 4 establishes the necessary and sufficient conditions under which the enforcement agency can implement the first-best deterrence outcome at (almost) no cost with ordered leniency, describes the second-best enforcement policy when the first-best outcome cannot be achieved, and underscores that the socially optimal level of harm is always attained.

Proposition 4. An optimal enforcement policy with ordered leniency for self-reporting can implement the first-best deterrence outcome at (almost) no cost if and only if $\underline{h} \leq \hat{b}(0, \pi) = [1 - (\overline{m}/n)(1 - \pi)^{\overline{m}-1}]\overline{f}$, where $\overline{m} = \sup\{m \in \{1, \ldots, n\} \mid m < 1/\pi\}$. When $\underline{h} > \hat{b}(0, \pi)$, the second-best enforcement policy involves a maximal fine, higher enforcement costs, and underdeter-

[63] It is simple to show that our main qualitative results also hold in this environment.

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

rence relative to the first best. The socially optimal level of harm $\underline{h}$ is always implemented.

Proposition 4 generalizes proposition 2 to groups of injurers with more than two members.[65] The proof, which follows immediately from lemma 7, is omitted. Intuitively, when the externalities associated with the harmful activities are small enough, the optimal law enforcement policy with ordered leniency allows the enforcer to implement the first-best deterrence outcome at an arbitrarily low cost. The socially optimal level of harm is always attained with an ordered-leniency policy.

4. Extensions

This section investigates several relevant extensions. Our results demonstrate the tractability of our framework and the robustness of our previous findings.

4.1. Convicting the Innocent

Although our benchmark framework allows for type II errors in conviction (no conviction of guilty parties),[66] it abstracts from type I error in conviction (conviction of innocent parties). In real-world settings, however, we might also observe innocent parties erroneously convicted and punished. Our model can be easily modified to allow for type I errors in conviction. We will show that our previous qualitative results also hold in this environment.

Suppose that innocent potential injurers—those who decide not to participate in the harmful activity—are sometimes erroneously punished for acts that they did not commit. In particular, suppose that evidence that links innocent parties to crimes that they did not commit (type I errors in conviction) emerges with probability $q \in (0, 1)$. If this evidence emerges, the innocent parties face the same likelihood of detection and conviction as guilty parties: the probability of detection is p_0 if neither agent self-reports and p_1 if just one agent self-reports. Suppose further that the emergence of this evidence is observed by all the players, including the innocent parties. Subsequently, the innocent and guilty parties play the same self-reporting game previously described. The incentives for the parties regarding whether and when to report are similar to the ones described before. Hence, in equilibrium, all parties self-report promptly.

Consider now the potential injurer's decision to engage in a harmful activity. Take p_0 and p_1 as fixed and let $\hat{b}$ be the expected fine conditional on committing the act defined in lemma 3. The expected fine conditional on not committing the act is $q\hat{b}$.[67] The parties will commit the act if and only if their net benefit from do-

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

[66] Absent self-reporting, injurers who committed the act would remain undetected with probability $1 - p_0$.

[67] With probability q, the evidence emerges and the expected fine is $\hat{b}$, the same as for injurers

ing so exceeds their net benefit from not doing so, $b - \hat{b} > -q\hat{b}$, or $b > (1-q)\hat{b}$. Hence, the deterrence threshold is lower in this environment. In other words, type I errors in conviction increase the potential injurers' incentives to engage in harmful activities. Following the logic of our benchmark model, maximal deterrence and minimal harm are obtained with the ordered-leniency multipliers defined in proposition 1. The first-best deterrence outcome is implemented when $\underline{h} \leq (1-q)\hat{b}(0, \pi)$, and the socially optimal level of harm is always achieved.

4.2. Public Information about Positions in the Self-Reporting Queue

In our benchmark model, the injurers simultaneously decide whether and when to self-report. In equilibrium, there is a race to the courthouse in which both injurers self-report promptly, and an equally weighted coin flip determines who obtains the first and second positions in the self-reporting queue. When viewed from an ex post perspective, the injurer who gets the second position in the self-reporting queue is worse off when he self-reports. Since $r_2 > p_1$, the injurer would be better off remaining silent and paying $p_1\overline{f}$. The reason the second injurer is willing to self-report is that when she is making the decision about whether to self-report, she does not know whether she will obtain the first or the second position in the self-reporting queue. Hence, if the second injurer to report could see that the first position in the self-reporting queue is already occupied, she would choose to remain silent. Our model can be easily modified to relax the assumption of secrecy regarding the taken positions in the self-reporting queue by allowing for sequential moves at the self-reporting subgame. We will show that our previous qualitative results also hold in this environment.

Assume that the self-reporting subgame involves two stages. In the first stage, as before, the injurers decide whether and when to self-report simultaneously and, in case of a tie, a coin flip determines their positions in the self-reporting queue. In the second stage, after the injurers have been assigned and learn their positions, they choose whether to withdraw their leniency applications and return to obscurity.[68] The ability of the injurers to change their minds once their relative positions are revealed places an additional constraint on the design of ordered-leniency policies. In particular, to induce the injurers to self-report, the leniency multipliers must satisfy $r_1 \leq p_0$ and $r_2 \leq p_1$. If this were not true, then the injurer who is assigned to the second position in the queue would withdraw her application. In the two-injurer optimal mechanism, $r_1 = p_0$ and $r_2 = p_1$. More generally, with n injurers, the multipliers must satisfy $r_i = p_{i-1}$ for $i = 1, \ldots, n$.

who do commit the act; with probability $1 - q$, the evidence does not emerge, and the expected fine is 0.

[68] Assume that the enforcement agency implements ordered-leniency policies through an agency's computer system: the initial leniency applications are anonymously submitted to the enforcer's website, the computer assigns the positions and informs the applicants of their positions, and the applicants decide whether to withdraw their initial leniency applications. Only the applicants who decide not to withdraw their initial applications need to identify themselves to the enforcement agency.

Deterrence is weaker in this new environment. In our benchmark model, the injurers decide to commit the harmful act when $b \leq p_1 f$ when p_1 is large and $b \leq (1 + p_0)/2$ when p_1 is small (proposition 1). In this new setting, the injurers commit the harmful act if and only if $b > \hat{b} = (p_0 + p_1)/2$. So the expected fine is lower and deterrence is weaker. The enforcement agency has an obvious incentive to maintain secrecy about whether the positions in the self-reporting queue are taken. Hence, although our main qualitative results also hold in this new setting, our benchmark implementation of ordered-leniency policies is welfare superior.

4.3. Endogenous Group Size

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 or 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 or committing an (a possibly different) act alone. Suppose that 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 whether the act was pursued by an individual or by a group of individuals.[69] With no leniency for self-reporting, $r_1 = r_2 = 1$, the expected fine for an injurer is $p_0 f$ whether the injurer commits the act alone or as part of a group. Now suppose instead that the enforcement agency offers leniency for the first position in the self-reporting queue, $r_1 = p_0 - \varepsilon$ ($\varepsilon > 0$, an arbitrarily small number), but grants 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 group, however, at least one injurer self-reports (as suggested by lemma 2), and the expected fine is $[(1 + p_0 - \varepsilon)/2]f$, which is strictly higher than $p_0 f$. Intuitively, because of the negative externalities in the self-reporting subgame, an ordered-leniency policy will increase the expected sanction for acts committed by groups but will not affect the expected sanction for individually committed acts.

More generally, by corollary 1, as the size of the group increases, the negative externalities in the self-reporting subgame rise, and the expected fine faced by an injurer increases. Hence, if the size of the group is endogenously decided by the potential injurers, the choice of a large group will be strictly dominated by the choice a small group. Importantly, the choice of single-injurer activities might be the dominant strategy. In other words, ordered-leniency policies might induce injurers to substitute away from harmful group activities and toward harmful individual activities.

[69] In practice, the value of 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.

Formally, assume that $n = 2$. Assume also 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. Assume also that the harm associated with the act committed by an injurer alone is βh, where $\beta \in (0, 1]$. The injurers will choose to act individually if and only if $\alpha b - p_0 f \geq \max\{b - [(1 + p_0)/2]f, 0\}$. Intuitively, single-injurer activities will be chosen when they provide the injurer a higher net benefit than committing the act as part of a group or not committing the act at all.

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

4.4. Stochastic Detection Rate

Our benchmark framework assumes that the social planner perfectly controls the probabilities of detection, p_0 and p_1, via its enforcement effort e. In other words, probabilities of detection are deterministic. 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 equilibrium, injurers who decide to commit the act in stage 1 later decide to self-report in stage 2. Thus, the self-reporting of harmful acts is ubiquitous. Our framework can be extended to allow for probabilities of detection that depend on the enforcement effort in a stochastic way.

Consider first our benchmark environment. Suppose that the inculpatory evidence is strong enough to convict a silent injurer with almost certainty: $p_1 = 1 - \varepsilon$, where $\varepsilon > 0$ is an arbitrarily small number.[70] 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 is obtained with multipliers $(r_1, r_2) = (e, 1)$. Therefore, the act will be deterred if $b \leq \hat{b}(e) = [(1 + e)/2]\overline{f}$.

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

In stage 1, before learning the realization of the random variable p_0, the injurers

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

must decide whether to commit the act. They are deterred from committing the act when

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

Note that the deterrence threshold in this stochastic environment (right-hand side of the inequality) is smaller than $\hat{b}(e)$, the deterrence threshold with a certain detection rate. Thus, having an uncertain detection rate compromises deterrence in stage 1. Intuitively, when p_0 is stochastic with a median value of e rather than a deterministic value of e, the potential injurers 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 (minimal) harm is not too great. Hence, social welfare will be unambiguously lower in environments with stochastic detection rates.[71]

4.5. *Rewards for Self-Reporting*

Our benchmark model explores the optimal design of ordered-leniency policies in which injurers are offered reductions in fines for self-reporting. Formally, this corresponds to leniency multipliers $r_i \in [0, 1]$ for $i = 1, 2$. Our model can be easily modified to allow for rewards for self-reporting, that is, $r_i \leq 1$ for $i = 1, 2$.[72]

It is simple to show that our previous analysis also holds in this environment, except for case 1 of proposition 1. If we allow for rewards, then the family of solutions in case 1 of proposition 1 will (weakly) expand. In particular, the upper bound for Δ will (weakly) rise from $\min\{p_1, 1 - p_1\}$ to $1 - p_1$. Consider the upper bound of this expanded range, $\Delta = 1 - p_1$. For this value of Δ, the multipliers are $(r_1, r_2) = (2p_1 - 1, 1)$. When $p_1 < \frac{1}{2}$, the first injurer to report the act receives a reward $r_1 < 0$, and the second injurer to report receives no leniency at all. In other words, if we allow for rewards, then it is no longer necessary to grant leniency to the second injurer who self-reports. The average multiplier is still p_1, and both injurers commit the act and self-report if and only if $b > p_1 \bar{f}$. Hence, although allowing for rewards for self-reporting expands the set of optimal enforcement policies, the use of rewards for self-reporting does not improve social welfare.

Although environments involving erroneous conviction of innocent parties, public information about positions in the self-reporting queue, endogenous group size, stochastic detection rates, and rewards for self-reporting obviously raise some new and interesting issues, the main insights derived from our general

[71] 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 $l(p_0; e)$ and $g(b)$.

[72] We thank a referee for pointing out this relevant extension.

model and the implications for the design of optimal enforcement policies with ordered leniency remain relevant.

5. 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 race to the courthouse among the members of the group of injurers. Prompt self-reporting by all injurers occurs in equilibrium, and as a result, the harm inflicted on others is minimized. Our findings also suggest that the expected fine increases with the size of the group, and hence ordered-leniency policies discourage large illegal enterprises. Finally, our analysis shows, first, that the socially optimal level of deterrence can be obtained at an arbitrarily low cost with an enforcement policy with ordered leniency when the externalities associated with the harmful activities are not too great and, second, that the socially optimal level of harm is always attained. Thus, we provide a social welfare rationale for the use of ordered-leniency policies.

Our findings regarding enforcement policies with ordered leniency for groups of injurers complement the results in Kaplow and Shavell (1994) 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 for self-reporting is socially desirable because these policies reduce the number of harmful activities, increase the likelihood of detection of harmful activities, and reduce the social harm caused by those activities without increasing enforcement costs.

Several relevant extensions are investigated. First, we study a setting in which innocent parties can be erroneously convicted and show that, although our main findings regarding the design of an optimal ordered-leniency policy also hold in this environment, fewer harmful activities are deterred. Second, we investigate a framework in which the secrecy regarding positions already taken in the self-reporting queue is relaxed. Our findings suggest that our original implementation of ordered-leniency policies in which secrecy is preserved is welfare superior. Third, we explore an environment in which the potential injurers can decide whether to commit individual or group harmful acts, and we show that, under certain conditions, the social value of ordered-leniency policies might be reduced. Fourth, we consider an environment in which the detection rate depends on the enforcement effort in a stochastic way. In that setting, injurers who commit an act may refrain from self-reporting if the probabilities of detection are sufficiently low. As a result, deterrence might be compromised. Fifth, we study an environment that allows for rewards for self-reporting and show that, although the set of optimal enforcement policies is expanded, our main results hold in this

environment. Importantly, social welfare is not improved by the use of rewards for self-reporting. Our results demonstrate the tractability of our framework and the robustness of our previous findings.

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 for the design of enforcement mechanisms for environmental policies (see Malik 1993; Livernois and McKenna 1999). Our results suggest that the implementation of ordered-leniency policies by environmental agencies might induce early detection of environmentally harmful activities and, hence, might reduce the social harm associated with these activities. In addition, stronger deterrence of violations of environmental policies might be implemented with ordered leniency.[73]

Our work provides important lessons for the design of law enforcement policies involving corporate and individual criminal liability (see Arlen and Kraakman 1997; Kraakman 1986). 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, p. 166), 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." 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.[74] 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 reduction in sanctions granted to the first to report would not generally be full, and the sanctions faced by the second to report may not be maximal.

Our findings are also relevant to qui tam (whistle-blower) lawsuits brought under the US False Claims Act (FCA). The FCA allows regular citizens to bring lawsuits against federal contractors claiming fraud against the federal government (31 U.S.C. 3729–33). The qui tam provision of the act grants the whistle-blower a fraction of ultimate recovery, often on the order of 15–25 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" (31 U.S.C. 3730[b][5]).[75] Our framework can be easily modified to study the design of optimal qui tam policies.

[73] Our findings and insights might be also relevant for the design of law enforcement mechanisms associated with tax policies and the control of tax evasion. See Andreoni (1991) and Malik and Schwab (1991).

[74] 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.

[75] 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).

Finally, our paper 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.[76] Since then, the story has been told and retold countless times, and a Google Scholar search for the phrase "prisoners' dilemma" delivers tens of thousands of articles in academic fields as diverse as economics, biology, philosophy, sociology, political science, and of course law.[77] 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 wrongdoers are sufficiently distrustful of each other, the prosecutor could forgo the prisoners' dilemma and employ a coordination mechanism instead.

Appendix

Formal Proofs

Proof of Lemma 1

Denote the strategy of player j as $\sigma_j = (\rho_j, t_j)$, where $\rho_j \in \{\text{R, NR}\}$ is whether to report the act and $t_j \in [0, 1]$ is when to report the act. Suppose that $r_1 < r_2$. If $\sigma_{-j} = (\text{NR}, t_{-j})$, then player j is indifferent about his reporting time, $(\text{R}, 0) \sim (\text{R}, t_j)\ \forall\ t_j \in (0, 1]$. If $\sigma_{-j} = (\text{R}, t_{-j})$, then for player j we have $(\text{R}, 0) \sim (\text{R}, t_j)\ \forall\ t_j < t_{-j}$ and $(\text{R}, 0) \succ (\text{R}, t_j)\ \forall\ t_j \geq t_{-j}$. Therefore (R, 0) weakly dominates $(\text{R}, t_j)\ \forall\ t_j \in (0, 1]$ when $r_1 < r_2$. Suppose instead that $r_1 > r_2$. If $\sigma_{-j} = (\text{NR}, t_{-j})$, then player j is indifferent: $(\text{R}, 1) \sim (\text{R}, t_j)\ \forall\ t_j \in [0, 1)$. If $\sigma_{-j} = (\text{R}, t_{-j})$, then $(\text{R}, 1) \sim (\text{R}, t_j)\ \forall\ t_j > t_{-j}$ and $(\text{R}, 1) \succ (\text{R}, t_j)\ \forall\ t_j \leq t_{-j}$. Therefore, (R, 1) weakly dominates $(\text{R}, t_j)\ \forall\ t_j \in [0, 1)$ 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. Q.E.D.

Proof of Lemma 2

In case 1, $r_1 f \leq p_0 f$ and $[(r_1 + r_2)/2]f \leq p_1 f$. With the tie-breaking assumption, self-reporting is a dominant strategy, and (R, R) is the unique Nash equilibrium (NE). In case 4, $r_1 f > p_0 f$ and $[(r_1 + r_2)/2]f > p_1 f$, so not reporting is a dominant strategy, and (NR, NR) is the unique NE. In case 2, $r_1 f \leq p_0 f$ and $[(r_1 + r_2)/2]f > p_1 f$, 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). The NE (R, R) Pareto dominates (NR, NR) if $[(r_1 + r_2)/2]f \leq p_0 f$ or $(r_1 + r_2)/2 \leq p_0$. The NE (R, R) risk dominates (NR, NR) if the former is preferred by player j if player $-j$ is randomizing 50/50 be-

[76] "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" (Princeton University 1995).

[77] As of July 8, 2019.

tween R and NR, or $\frac{1}{2}(r_1 f) + \frac{1}{2}[(r_1 + r_2)/2]f \leq \frac{1}{2}(p_0 f) + \frac{1}{2}(p_1 f)$, or $(3r_1 + r_2)/4 \leq (p_1 + p_2)/2$. Q.E.D.

Proof of Lemma 3

Consider the four cases included in lemma 2. In case 1, (R, R) is the unique NE, and each injurer pays an expected fine of $[(r_1 + r_2)/2]f$. The injurers will commit the act if $b > [(r_1 + r_2)/2]f$. In case 2, there are two NE, (R, NR) and (NR, R), which cannot be ranked using conventional equilibrium refinements. In both equilibria, the expected fine is $[(r_1 + r_2)/2]f$, and the act is committed when $b > [(r_1 + r_2)/2]f$. In case 3, there are two NE, (R, R) and (NR, NR), which can be ranked using conventional equilibrium refinements. The act is committed if $b > p_0 f$ or $b > [(r_1 + r_2)/2]f$, 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 $b > p_0 f$. Q.E.D.

Proof of Proposition 1

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. We now characterize the values (r_1, r_2) that maximize the expected fine $[(r_1 + r_2)/2]f$ subject to the constraints that (i) $(r_1 + r_2)/2 \leq p_1$, (ii) $r_1 \in [0, p_0]$, and (iii) $r_2 \in [0, 1]$. Two subcases are considered:

Case 1.1. The first subcase refers to $p_1 \leq (1 + p_0)/2$. If $p_1 \leq (1 + p_0)/2$, then constraint i must hold with equality, $(r_1 + r_2)/2 = p_1$. Suppose not: suppose that $(r_1 + r_2)/2 < p_1$. This would imply that both $r_1 = p_0$ and $r_2 = 1$, for otherwise the expected fine $[(r_1 + r_2)/2]f$ could be increased. Then $(r_1 + r_2)/2 = (1 + p_0)/2 < p_1$, which is a contradiction. Therefore, $(r_1 + r_2)/2 = p_1$. 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, $\Delta \in [p_1 - p_0, \min\{p_1, 1 - p_1\}]$. The expression $p_1 \leq (1 + p_0)/2$ 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 subcase refers to $p_1 > (1 + p_0)/2$. If $p_1 > (1 + p_0)/2$, then constraint i does not bind at the optimum: $(r_1 + r_2)/2 < p_1$. Suppose not: suppose that $(r_1 + r_2)/2 = p_1$. 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 $p_1 > (1 + p_0)/2$ implies that $2p_1 > 1 + p_0$, which implies further that $p_1 - p_0 > \min\{p_1, 1 - p_1\}$. So no such value for Δ exists. Therefore, $(r_1 + r_2)/2 < p_1$. 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 $[(1 + p_0)/2]f < p_1 f$.

Case 2. There are multiple equilibria in this case. Neither the Pareto-dominance nor the risk-dominance refinements will eliminate either outcome:

both equilibria, (R, NR) and (NR, R), are Pareto or risk dominant. By assumption, equal probabilities are placed on both outcomes. Hence, the expected fine is $[(r_1 + p_1)/2]f$. 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 $[(p_0 + p_1)/2]f$. 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 $[(r_1 + r_2)/2] \leq p_0$. 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 $(r_1 + r_2)/2$ subject to the constraints that (i) $(3r_1 + r_2)/4 \leq (p_0 + p_1)/2$, (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 at which 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 $[r_1 + \min\{2(p_0 + p_1) - 3r_1, 1\}]/2$ Two subcases are considered:

Case 3.1. The first case refers to risk dominance and $p_1 \leq (1 + p_0)/2$. If $p_1 \leq (1 + p_0)/2$, 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 $p_1 > (1 + p_0)/2$. If $p_1 > (1 + p_0)/2$, then r_1 will be strictly greater than p_0 and the expected fine strictly higher than $p_1 \overline{f}$. To see why this is true, suppose that $r_1 = p_0 + \varepsilon$, where $\varepsilon > 0$. Since $p_1 > (1 + p_0)/2$ implies that $2p_1 > 1 + p_0$, 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 $[(r_1 + 1)/2]f$. Deterrence would be higher if r_1 were raised above p_0. The term r_1 will be raised to the point at which $2(p_0 + p_1) - 3r_1 = 1$, and so $r_1 = [2(p_0 + p_1) - 1]/3$ and $r_2 = 1$. The expected fine is $[(1 + p_0 + p_1)/3]f$. 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 $p_1 \leq (1 + p_0)/2$, the maximal expected fine corresponds to case 1 or case 3; when risk dominance is applied in case 3 and $p_1 > (1 + p_0)/2$, the maximal expected fine corresponds to case 3.

Second, since $r_1^j < r_2^j$ for $j = \mathrm{S}, \mathrm{M}$, 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 greatest deterrence is obtained with the maximal fine, $f = \overline{f}$. Q.E.D.

Proof of Lemma 4

Proposition 3 implies that if $p_1 \leq (1 + p_0)/2$, then $\hat{b}^S = \hat{b}^M = p_1\bar{f}$, and if $p_1 > (1 + p_0)/2$, then $\hat{b}^S = [(1 + p_0)/2]\bar{f}$, $\hat{b}^M = [(1 + p_0 + p_1)/3]\bar{f}$, and $\hat{b}^S < \hat{b}^M$. Substituting $p_0 = e$ and $p_1 = e + (1 - e)\pi$ gives cases 1 and 2 of lemma 4. Q.E.D.

Proof of Proposition 2

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 1 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 $\hat{b}^i(e, \pi)$, $i = S, M$, corresponds 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 $[\partial \hat{b}(e, \pi)/\partial e]g(\hat{b}^i(e, \pi))$ 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

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

We need to show that the second-best outcome involves $e^i > 0$. Suppose not: suppose that $e^i = 0$. In that case, $h > \hat{b}^i(0, \pi)$, since by assumption the first-best enforcement policy cannot be obtained, $\partial \hat{b}^i(e, \pi)/\partial e > 0$ 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 $[\partial \hat{b}^i(e, \pi)/\partial e]g(\hat{b}^i(e, \pi)) > 0$, the slope of the welfare function would be strictly negative. Social welfare would be higher if e were reduced. Q.E.D.

Proof of Proposition 3

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 $\boldsymbol{x}$ be the vector of multipliers for which condition (1) 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

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

The vector $\boldsymbol{x}$ has important properties: $x_1 = p_0 > 0$; $x_1 < x_2 < \dots x_n$, by our assumption that the sequence $\{ip_{i-1}\}_{i=1}^{n}$ is convex in i; and x_j $(j = 2, \dots, n)$ 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 proofs follow.

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

Proof. First, suppose that $\{r_i\}_{i=1}^{n}$ is constant in i, so $r_1 = \dots = r_n$. If $r_1 < p_0$, then there is a unique CPNE in which 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 in which no injurer self-reports the act, and the expected fine is p_0.

Next, suppose that $\{r_i\}_{i=1}^{n}$ 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 that there is a CPNE in which $m < n$ injurers self-report, and the remaining $n - m + 1$ injurers do not self-report. It must be true that $(1/m)\sum_{i=1}^{m} r_i \leq p_{m-1}$. 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 $p_m \leq [1/(m+1)]\sum_{i=1}^{m+1} r_i$, 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.

1. Suppose that $(1/n)\sum_{i=1}^{n} r_i > p_0$. There is a unique CPNE in which 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 $(1/m)\sum_{i=1}^{m} r_i > p_0$ 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.

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

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

Proof. The proof that condition (1) is sufficient is in the main text of the paper. We now prove that condition (1) is necessary.

Suppose that 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 $(1/n)\sum_{i=1}^{n} r_i \leq p_{n-1}$. 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.[78] The injurers in the 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, $[1/(m+1)]\sum_{i=1}^{m+1} r_i \leq p_m$ for all $m = 1, \ldots, n - 1$. There are two cases to consider.

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

First, suppose that $(1/n)\sum_{i=1}^{n} r_i > p_m$, 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: $[1/(m+1)]\sum_{i=1}^{m+1} r_i \leq p_m$. This is condition (1).

Second, suppose that $(1/n)\sum_{i=1}^{n} r_i \leq p_m$, 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 $[1/(m+1)]\sum_{i=1}^{m+1} r_i \leq p_m$. Again, this is condition (1). Q.E.D.

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

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

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

subject to

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

We start by demonstrating that if $\boldsymbol{r}$ is a solution to this program, then there is another (possibly different) solution $\boldsymbol{r}'$ with the property that $r'_m = 1$ if and only if $m > \bar{m}$ for some value $\bar{m}$. Suppose that the vector $\boldsymbol{r}$ 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 $\boldsymbol{r}'$ that is identical to $\boldsymbol{r}$ except that two values are swapped: $r'_{m-1} = r_m < 1$ and $r'_m = r_{m-1} = 1$. Notice that the expected fine associated with $\boldsymbol{r}'$ is the same as $\boldsymbol{r}$. We will now show that vector $\boldsymbol{r}'$ satisfies the system of equations. The only constraints we need to check are $m - 1$ and m. First, consider constraint $m - 1$: $r'_{m-1} \leq \min\{(m-1)p_{m-2} - \sum_{i=1}^{m-2} r_i, 1\}$. The right-hand side is the same with $\boldsymbol{r}'$ as with $\boldsymbol{r}$. Since $r'_{m-1} < r_{m-1}$, constraint $m - 1$ is satisfied by the new vector $\boldsymbol{r}'$ too. Next, consider constraint m: $r'_m \leq \min\{m p_{m-1} - \sum_{i=1}^{m-2} r_i - r'_{m-1}, 1\}$. The right-hand side is different with $\boldsymbol{r}'$ than with $\boldsymbol{r}$, 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 $\boldsymbol{r}$ in which $r_m \leq m p_{m-1} - \sum_{i=1}^{m-1} r_i$ if $m \leq \tilde{m}$ and $r_i = 1$ if $m > \tilde{m}$ for some value $\tilde{m}$. Importantly, constraint $\tilde{m}$ must bind, since otherwise $r_{\tilde{m}}$ could be raised without violating any constraint. So in the solution to the program, $r_{\tilde{m}} = \tilde{m} p_{\tilde{m}-1} - \sum_{i=1}^{\tilde{m}-1} r_i$ or, equivalently, $\sum_{i=1}^{\tilde{m}} r_i = \tilde{m} p_{\tilde{m}-1}$. (Constraints $i = 1, \ldots, \tilde{m} - 1$ need not bind,

and there is generally a continuum of solutions to the linear program, just as there is a continuum of solutions in proposition 1, case 1.) The solution to the program will therefore have

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

We now make use of the definitions of the vector $\boldsymbol{x}$ 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, $\bar{m}p_{\bar{m}-1} + (n - \bar{m})$. Q.E.D.

Claim 4. Suppose that $\{r_i\}_{i=1}^{n}$ 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 when all n injurers self-report.

Proof. Consider an ordered-leniency policy in which 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: $(1/n')\sum_{i=1}^{n'} r_i \leq p_{n'-1}$. 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 $(1/m)\sum_{i=1}^{m} r_i \leq p_{m-1}$ for all $m = 1, 2, \ldots, n'$. 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 $m = 2, \ldots, n'$, and the fines for the injurers who remain silent are $p_{n'}$.

With the ordered-leniency policy in which 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 $r_m \leq \min\{mp_{m-1} - (m - 1)p_{m-2}, 1\}$ for all $m = 2, \ldots, n$, the first n' injurers face weakly higher fines. Next, consider the $n - n'$ injurers who do not self-report. With the ordered-leniency policy in which 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 $i = n' + 1, n' + 2, \ldots, n$ are higher than $p_{n'}$ (the expected fine if exactly n' injurers self-report).

We conclude that any ordered-leniency policy in which $n' < n$ injurers self-report has a lower expected fine than an ordered-leniency policy in which all n injurers self-report. Q.E.D.

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, at $t = 0$. 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 3. Q.E.D.

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 $m \in \{2, \ldots, n - 1\}$. With the expressions included in proposition 3,

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

So we have $x_1 = e$ and

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

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

$$\bar{m} = \sup\left\{m \in \{1, 2, \ldots, n\} \,\middle|\, m < \frac{1}{\pi}\right\}.$$

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 gives

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

which has the same sign as $\pi - (1 - m\pi)\ln(1 - \pi)$. Since $\ln(1 - \pi) < 0$, we have that $\partial x_m/\partial m > 0$ when $m \leq \bar{m}$ and $\partial x_m/\partial m < 0$ 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 lemma 7 follows from substituting $p_{m-1} = e + (1-e)[1-(1-\pi)^{m-1}]$ into the expression in proposition 3. Taking the derivative with respect to e gives $\partial\hat{b}(e, \pi)/\partial e = (\bar{m}/n)(1-\pi)^{\bar{m}-1}\bar{f} \in (0, \bar{f})$. Q.E.D.

References

Andreoni, James. 1991. The Desirability of a Permanent Tax Amnesty. *Journal of Public Economics* 45:143–59.

Arlen, Jennifer. 2012. Corporate Criminal Liability: Theory and Evidence. Pp. 144–203 in *Research Handbook on Criminal Law,* edited by Alon Harel and Keith Hylton. Northampton, MA: Edward Elgar Publishing.

Arlen, Jennifer, and Reinier Kraakman. 1997. Controlling Corporate Misconduct: An

Analysis of Corporate Liability Regimes. *New York University Law Review* 72:687–779.
Aubert, Cécile, Patrick Rey, and William E. Kovacic. 2006. The Impact of Leniency and Whistle-Blowing Programs on Cartels. *International Journal of Industrial Organization* 24:1241–66.
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.
Buccirossi, Paolo, and Giancarlo Spagnolo. 2006. Leniency Policies and Illegal Transactions. *Journal of Public Economics* 90:1281–97.
Caplin, Andrew, and Andrew Schotter, eds. 2010. *The Foundations of Positive and Normative Economics: A Handbook.* Oxford: Oxford University Press.
Ceresney, Andrew. 2015. The SEC's Cooperation Program: Reflections on Five Years of Experience. Remarks presented at University of Texas School of Law's Government Enforcement Institute, Dallas, May 13. 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* 91:525–41.
Chen, Zhijun, and Patrick Rey. 2013. On the Design of Leniency Programs. *Journal of Law and Economics* 56:917–57.
Cooter, Robert D., and Nuno Garoupa. 2014. A Disruption Mechanism for Bribes. *Review of Law and Economics* 10:241–63.
Engstrom, David F. 2012. Harnessing the Private Attorney General: Evidence from Qui Tam Litigation. *Columbia Law Review* 112:1244–1325.
Federal Bureau of Investigation. 2012. *Financial Crimes Report 2010–2011.* Washington, DC: US Department of Justice, Federal Bureau of Investigation. https://www.fbi.gov/file-repository/stats-services-publications-financial-crimes-report-2010-2011-financial-crimes-report-2010-2011.pdf.
Feess, Eberhardt, and Markus Walzl. 2004. Self-Reporting in Optimal Law Enforcement When There Are Criminal Teams. *Economica* 71:333–48.
Garoupa, Nuno. 1997. The Theory of Optimal Law Enforcement. *Journal of Economic Surveys* 11:267–95.
———. 2000. The Economics of Organized Crime and Optimal Law Enforcement. *Economic Inquiry* 38:278–88.
Grossman, Gene M., and Michael L. Katz. 1983. Plea Bargaining and Social Welfare. *American Economic Review* 73:749–57.
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* 511–27.
Harsanyi, John C., and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.
Innes, Robert. 1999. Remediation and Self-Reporting in Optimal Law Enforcement. *Journal of Public Economics* 72:379-93.
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* 23:507–17.
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–

42.
Kraakman, Reinier H. 1986. Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy. *Journal of Law, Economics, and Organization* 2:53–104.
Landeo, Claudia M., and Kathryn E. Spier. 2009. Naked Exclusion: An Experimental Study of Contracts with Externalities. *American Economic Review* 99:1850–77.
———. 2012. Exclusive Dealing and Market Foreclosure: Further Experimental Results. *Journal of Institutional and Theoretical Economics* 168:150–70.
———. 2015. Incentive Contracts for Teams: Experimental Evidence. *Journal of Economic Behavior and Organization* 119:496–511.
———. 2018a. Optimal Law Enforcement with Ordered Leniency. Working Paper No. w25095. National Bureau of Economic Research, Cambridge, MA.
———. 2018b. Ordered Leniency: An Experimental Study of Law Enforcement with Self-Reporting. Working Paper No. w25094. National Bureau of Economic Research, Cambridge, MA.
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* 71:415–40.
Malik, Arun S. 1993. Self-Reporting and the Design of Policies for Regulating Stochastic Pollution. *Journal of Environmental Economics and Management* 24:241–57.
Malik, Arun S., and Robert M. Schwab. 1991. The Economics of Tax Amnesties. *Journal of Public Economics* 46:29–49.
Marvão, Catarina, and Giancarlo Spagnolo. 2018. Cartels and Leniency: Taking Stock of What We Learnt. Pp. 2:57–90 in *Handbook of Game Theory and Industrial Organization,* edited by Luis C. Corchón and Marco A. Marini. Northampton, MA: Edward Elgar Publishing.
Motta, Massimo, and Michele Polo. 2003. Leniency Programs and Cartel Prosecution. *International Journal of Industrial Organization* 21:347–79.
Piccolo, Salvatore, and Giovanni Immordino. 2017. Organized Crime, Insider Information and Optimal Leniency. *Economic Journal* 127:2504–24.
Polinsky, A. Mitchell, and Steven Shavell. 1984. The Optimal Use of Fines and Imprisonment. *Journal of Public Economics* 24:89–99.
Princeton University. 1995. Albert William Tucker. News release, January 26. https://www.princeton.edu/pr/news/95/q1/0126tucker.html.
Rasmusen, Eric B., J. Mark Ramseyer, and John S. Wiley, Jr. 1991. Naked Exclusion. *American Economic Review* 81:1137–45.
Reinganum, Jennifer F. 1988. Plea Bargaining and Prosecutorial Discretion. *American Economic Review* 78:713–28.
Segal, Ilya R., and Michael D. Whinston. 2000. Naked Exclusion: Comment. *American Economic Review* 90:296–309.
Siegel, Ron, and Bruno Strulovici. 2018. Judicial Mechanism Design. Unpublished manuscript. Pennsylvania State University, Department of Economics, University Park.
Silva, Francisco. 2019. If We Confess Our Sins. *International Economic Review* 60:1–24.
Spagnolo, Giancarlo. 2005. Divide et Impera: Optimal Leniency Programs. Unpublished manuscript. Stockholm School of Economics, Stockholm.
Spier, Kathryn E. 1994. A Note on Joint and Several Liability: Insolvency, Settlement, and Incentives. *Journal of Legal Studies* 23:559–68.