Ordered Leniency: An Experimental Study of Law Enforcement with Self-Reporting*

Claudia M. Landeo† Kathryn E. Spier‡

July 3, 2021

Abstract

This paper experimentally studies law enforcement policies with ordered leniency to detect and deter harmful short-term acts committed by groups of injurers. With ordered leniency, the degree of leniency granted to injurers who self-report depends on their positions in a self-reporting queue. In theory, with ordered leniency, injurers self-report faster, acts are detected sooner, and fewer acts are committed. Our findings are as follows. First, we show that ordered leniency raises the likelihood of self-reporting of the act. These results are aligned with the risk-dominance refinement when multiple equilibria arise. Second, ordered leniency raises the likelihood that injurers are detected and punished. Third, ordered leniency reduces the time to detect the act. This result might suggest that ordered leniency generates a race-to-the-courthouse effect where both injurers self-report promptly. Fourth, we show that the potential injurer's willingness to participate in the act is higher when the private economic benefit from the act is above the expected fine. Our findings provide support to the theory.

KEYWORDS: Law Enforcement; Ordered Leniency; Self-Reporting; Experiments; Non-Cooperative Games; Coordination Games; Prisoners' Dilemma; Risk Dominance; Pareto Dominance

JEL Categories: C72, C90, D86, K10, L23

---

*We acknowledge financial support from the National Science Foundation (NSF Grant SES-1155761). We thank Tom Brennan, Dan Coquillette, Nick Feltovich, Mike Gilbert, John Goldberg, Andrew Hayashi, Christine Jolls, Louis Kaplow, Mike Meurer, Max Nikitin, Jack Ochs, Abe Wickelgren, and Kathy Zeiler for insightful discussions and comments. We are grateful for suggestions from participants at the NBER Summer Institute in Law and Economics, the Harvard Law School Faculty Workshop, the University of Virginia and Boston University. We thank Susan Norton for administrative assistance, Tim Yuan for programming the software used in this study, and the Harvard Decision Science Laboratory for assistance in conducting the experimental sessions.

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

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

1 Introduction

Socially-harmful activities, including corporate crimes and securities fraud, are often committed by groups of wrongdoers rather than by individuals working in isolation. Law enforcement agencies have established leniency policies to encourage wrongdoers to self-report, with the goals of improving the early detection of these activities and strengthening deterrence.[1] A common feature of these policies is that wrongdoers who cooperate with the agency early in the process receive reduced sanctions, i.e., law enforcement agencies apply ordered-leniency mechanisms (Landeo and Spier, 2020). For instance, in 2009, the Securities and Exchange Commission (SEC) brought charges against Raj Rajaratnam and other co-conspirators for insider trading at several hedge funds including Galleon Management LP and New Castle Funds LLC.[2] Anil Kumar, an early cooperator, signed a cooperation agreement and was granted leniency. In contrast to other co-conspirators, he paid a reduced fine and received no prison time. At Kumar's sentencing, U.S. Attorney Prett Bharara stated: "Kumar's *immediate* cooperation warrants special mention and recognition" (Southern District of New York, Sentencing Memorandum, 2012, p. 12).[3]

Kaplow and Shavell (1994) provide seminal theoretical work on the control of harmful externalities with self-reporting.[4] In the context of single-injurer activities, they show that enforcement with self-reporting can induce individuals to report their harmful acts without compromising deterrence. Landeo and Spier (2020) extend this literature by studying the design of optimal enforcement policies with ordered leniency to detect and deter socially-harmful short-term activities committed by groups of injurers.[5] With ordered leniency, the degree of leniency granted to an injurer is determined by his or her position in a self-reporting queue. Granting a penalty reduction to the first injurer to report, and possibly (albeit lower) to the subsequent injurers, generates a so-called "race to the courthouse" where, in equilibrium, the injurers compete against each other to secure higher positions in the self-reporting queue.[6] By inducing prompt self-reporting, ordered leniency increases the likelihood of early detection of harmful acts without raising the enforcement costs, and hence allows the enforcement agency to mitigate the harm inflicted on society. As a result of the higher likelihood of detection, deterrence is strengthened and social welfare is fur-

[1] The Securities and Exchange Commission's Cooperation Program is an example of such a leniency policy.

[2] *SEC v. Rajaratnam,* 622 F.3d 159 (2nd Cir. 2010).

[3] Emphasis added; *United States v. Kumar*, Case 1:10-cr-00013-DC, Document 47-1 (S.D.N.Y. July 16, 2012). Additional legal cases involving ordered leniency are discussed in Landeo and Spier (2020, 2018).

[4] See Becker (1968) and Polinsky and Shavell (1984) for early work on law enforcement policies.

[5] Short-term illegal activities do not involve an ongoing relationship among group members. They are also referred as "occasional" activities (Buccirossi and Spagnolo, 2006). In game-theoretic terms, they correspond to one-shot strategic environments. Leniency programs have been also applied to long-term illegal activities. See Spagnolo and Marvão (2016) for a survey on this literature.

[6] In civil litigation cases, the expression "race to the courthouse" refers to the superior rights granted to the first action filed. In Landeo and Spier (2020), early reporting increases the likelihood of getting the first position in the self-reporting queue and, hence, of getting a larger reduction in the penalty.

ther improved. Our paper contributes to this literature by exploring these issues in a laboratory setting. No previous experimental test has been conducted to assess these theoretical predictions.

We apply experimental economics methods to study the effectiveness of ordered-leniency policies in detecting and deterring short-term group acts. The theoretical framework studied in this paper builds on Landeo and Spier (2020). The strategic environment involves two potential injurers. The potential injurers first decide whether to participate in a harmful group act. The act is committed only if both potential injurers decide to participate. 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 depends on the likelihood of detection if he remains silent, which itself depends on the self-reporting decision of the other injurer. The likelihood that an injurer will be detected and sanctioned is higher when the other injurer reports the act. In other words, there are negative externalities in the self-reporting stage.

Our experimental design includes three leniency treatments: Strong Leniency, Mild Leniency and No Leniency. In the first treatment, Strong Leniency, the first injurer to report receives a large reduction in the penalty. The strategic environment at the self-reporting stage resembles a prisoners' dilemma game with a unique (pure-strategy) Nash equilibrium in which both injurers self-report promptly. In the second treatment, Mild Leniency, the first injurer to report the act receives a small reduction in the penalty. The strategic environment at the self-reporting stage resembles a coordination game with two (pure-strategy) Nash equilibria: One in which both injurers self-report the act, the risk-dominant equilibrium (Harsanyi and Selten 1988), and the other in which neither injurer self-reports the act, the Pareto-dominant equilibrium. In the third treatment, No Leniency, penalty reductions for self-reporting are not granted. The strategic environment at the self-reporting stage involves a unique Nash equilibrium in which neither injurer self-reports the act. The participants, a pool of undergraduate and graduate students from Harvard University, Boston University, and Northeastern University, were paid according to their performance.

Experiments are particularly relevant to test different policy designs, and to observe the induced changes in the potential injurers' choices regarding participating in the act and cooperating with the enforcement agency by self-reporting the act. Furthermore, as Landeo and Spier (2020) demonstrate, the optimal ordered-leniency policy depends critically on the refinement criterion for equilibrium selection when multiple equilibria arise. 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). Hence, it is crucial to evaluate whether the choices of the injurers are aligned with either the risk-dominance or the Pareto-dominance refinements in environments with multiple equilibria.

Our main findings are as follows. First, Mild and Strong leniency raise the likelihood of self-reporting of the act. These results suggest that the choices at the self-reporting stage of the majority of injurers are aligned with the risk-dominance refinement when multiple equilibria arise.

Second, Mild and Strong Leniency raise the likelihood of detection and punishment of the injurer. Third, Mild and Strong Leniency reduce the time to detect the act. This finding might suggest that ordered-leniency policies generate a race-to-the-courthouse' effect where both injurers self-report promptly. Consequently, ordered leniency has the potential to mitigate social harm. Our results demonstrate the effectiveness of ordered-leniency policies as a detection mechanism.

Fourth, across leniency treatments, we show that the potential injurer's willingness to participate in the act is higher when the private economic benefit from the act is above the expected fine. Fifth, we find that the potential injurers are willing to participate in the act more frequently than predicted by the theory under Strong and Mild Leniency. In particular, the majority of potential injurers consider deterrence thresholds that are lower than those predicted by theory when making their participation decisions. These departures from the theory at the participation stage might originate in the potential injurers' beliefs about securing the first position in the self-reporting queue. As a result, the deterrence power of ordered leniency is weakened. Our findings underscore the importance of combining theoretical modeling with experimental and behavioral observation.

Our results challenge the common view among legal practitioners and policy makers that only the prisoners' dilemma game is relevant for the design of law enforcement policies and cooperation agreements with wrongdoers. We demonstrate that the enforcement agency might strengthen detection without increasing investigation costs by implementing either a coordination game or a prisoners' dilemma game through Mild Leniency or Strong Leniency, respectively. In particular, when Mild Leniency is implemented, although the injurers are jointly better off by refusing to cooperate with the enforcement authority, they are individually induced to cooperate due to strategic uncertainty. If the injurers are sufficiently distrustful of each other, then prompt self-reporting is also elicited under Mild Leniency. In consequence, the enforcement agency is able to detect the acts sooner and hence, mitigate social harm.

Our paper is motivated by public policies related to insider trading and securities fraud. We believe, however, that the theoretical environment studied in this paper and the insights derived from our experimental investigation might apply to the design of other public policies as well. The control of harmful externalities and the implementation of law-enforcement policies with self-reporting are relevant to environments such as plea bargaining with criminal defendants (Landes, 1971; Grossman and Katz, 1983; and, Kobayashi, 1992), corporate criminal liability and third-party enforcement (Kraakman, 1986; Arlen and Kraakman, 1997; and Arlen, 2012), federal government misbehavior and the *qui tam* whistleblower program (Engstrom, 2012), corporate misconduct and the whistleblower mechanisms under the Dodd-Frank Wall Street Reform and Consumer Protection Act (Greenberg, 2011), environmental policies and standards (Livernois and McKenna, 1999), and tax evasion policies (Andreoni, 1991; and Malik and Schwab, 1991), among others.[7]

[7] See Landeo and Spier (2020) for further discussion and theoretical analysis.

Related Literature.– Our paper is related to several strands of literature. The closest to our work is the theoretical literature on law enforcement and leniency for self-reporting. Kaplow and Shavell (1994) study a probabilistic enforcement model in which harmful acts are committed by individuals. They show that leniency for self-reporting directly reduces the enforcement costs without affecting deterrence. In their framework, injurers who self-report the act pay a sanction that is (slightly) lower than the expected sanction they would need to pay if they did not report the act. Given that the enforcement agency does not need to spend resources to identify the injurers who self-report, investigatory efforts are economized.[8] Landeo and Spier (2020) extend this literature by focusing on harmful acts committed by groups of injurers. They show that granting leniency to the first injurer to report, and possibly to the subsequent injurers, increases the likelihood of detection without raising investigatory costs, raises the expected sanctions, strengthens deterrence, and reduces social harm. In their model, the optimal enforcement policy with ordered leniency exploits the negative externalities between the injurers at the self-reporting stage. Our work complements Landeo and Spier (2020) by providing the first experimental evidence of the effects of ordered-leniency policies on detection and deterrence of short-term group acts.

Our work is also related to the experimental literature on competition policies and collusion. Motta and Polo (2003) and Spagnolo (2005) provide seminal theoretical work on leniency programs for cartels in repeated-game strategic environments.[9] They demonstrate that although leniency policies have the potential to strengthen deterrence, these policies might be also exploited by wrongdoers. In particular, Spagnolo (2005) shows that leniency policies might be used by co-conspirators as a punishment tool, and hence contribute to the stability of collusive agreements. Bigoni et al. (2012) study the effects of leniency programs in the lab. Their findings suggest that leniency policies strengthen deterrence (i.e., reduce cartel formation) but they also contribute to the stabilization of surviving cartels.[10] More recently, Feltovich and Hamaguchi's (2018) experimental findings suggest that leniency policies reduce both collusive behavior and average prices conditional on successful collusion.[11] None of these studies involve ordered leniency, i.e., a policy where the degree of leniency depends on the injurer's position in a self-reporting queue. Hence, the sanction reductions granted under these leniency policies do not depend on the probability of securing the first position in the self-reporting queue.

Our paper is also connected with the experimental economics literature on coordination games and equilibrium refinements. In seminal work, Cooper et al. (1990) suggest that risk-dominance is

[8]See also Fees and Walzl, 2004; Innes 1999; Livernois and McKenna 1999; Malik 1993; Andreoni 1991; Malik and Schwab 1991.

[9]See also Harrington (2008) and Chen and Rey (2013).

[10]See also Bigoni et al. (2015). See Apestegui et al. (2007) and Hinloopen and Soetevent (2008) for seminal experimental work on leniency and cartels.

[11]In theory, leniency policies reduce the stability of the cartel (direct effect). However, they also lower the cost of exiting the cartel and hence increase the incentives to form a cartel (indirect effect). Feltovich and Hamaguchi (2018) provide experimental evidence of the direct effect but not of the indirect effect.

generally the equilibrium selection criterion chosen by subjects in the lab when there are multiple equilibria.[12] Van Huyck et al. (1993) present experimental evidence of the effects of auctioning the right to participate in a median game on equilibrium selection. Their strategic environment involves a coordination game with endogenous participation of a subset of players. Their findings suggest that purchasing the right to participate communicates information about equilibrium selection, and hence allows the participants to coordinate on the Pareto-dominant equilibrium.[13] Landeo and Spier (2009, 2012) provide experimental evidence of the effects of exclusionary vertical contracts.[14] Their strategic environment involves a coordination game among buyers with payoffs endogenously determined by the incumbent monopolist. The market is foreclosed when the buyers fail to coordinate on their preferred equilibrium (the Pareto-dominant equilibrium). Their findings also suggest that the buyers' choices are aligned with the risk-dominance refinement. In environments involving ordered leniency, the expected fines depend on the probability of detection and on the probability of securing the first position in the self-reporting queue. Hence, equilibrium selection is a more complex task.

Our paper is also related to the empirical literature on crime and deterrence.[15] This literature underscores the importance of the potential injurers' beliefs about the consequences of participating in the illegal activity on the deterrence effect of enforcement policies. Two important findings from this previous work deserve special attention. First, these studies provide evidence of the difference between perceptual deterrence (subjective risk associated with the participation in the harmful act) and actual risk (Kleck et al., 2005; Kleck and Barnes, 2014; Lochner, 2007).[16] Second, previous experimental work suggests that there is an association between perceptual deterrence, self-serving bias,[17] and participation in harmful acts (Nagin and Pogarsky, 2003).[18] None of these

[12] Burton and Sefton (2004) provide powerful evidence of the role of riskiness in the choice of a strategy.

[13] Crawford and Broseta's (1998) theoretical model captures these findings. See Ochs (1995) for a survey of seminal work on coordination games. See Charness et al. (2007) for an experimental study of the prisoner's dilemma with endogenous transfers made in the first period.

[14] See the theoretical work of Rasmusen et al. (1991) and Segal and Whinston (2000). See Smith (2011) for additional experimental work on exclusive contracts and Landeo and Spier (2015) for experimental work on the design of incentive contracts for teams in the presence of externalities among team members.

[15] See Chalfin and McCrary (2017), Apel (2013) and Pogarsky (2009) for comprehensive surveys.

[16] Using the National Longitudinal Survey of Youth (NLSY), Lochner (2007) finds that there is no correlation between perceived and actual risk. Interestingly, his results regarding a sample of active offenders (who self-report criminal involvement) suggest that these individuals, on average, have more accurate perceptions about the risk of arrest than non-criminal individuals.

[17] The self-serving bias, a class of cognitive bias, is attributed to motivated reasoning. In the context of pretrial bargaining experiments, Babcock et al. (1995) show that even when the parties are exposed to the exact same information, they arrive at expectations of an adjudicated settlement that are biased in a self-serving manner.

[18] Nagin and Pogarsky's (2003) findings provide robust evidence of the association between self-serving bias and participation in harmful activities, and self-serving bias and perceptual deterrence. Their results suggest that individuals with high self-serving bias are significantly more likely to participate in the act (cheat in the lab), and individuals with high self-serving bias view themselves less likely to be detected.

studies consider leniency for self-reporting. Hence, the sanctions imposed under these environments do not depend on the injurers' self-reporting decisions and the probability of securing the first position in the self-reporting queue.

The rest of the paper is organized as follows. Section 2 presents the theoretical framework, the model parameterization, and the equilibrium analysis. Section 3 describes the experimental design. Section 4 discusses the qualitative hypotheses. Section 5 examines the results from the experimental sessions and provides evidence of the detection power and the race-to-the-courthouse effect of ordered-leniency policies. Section 6 concludes the paper and discusses avenues for future research.

2 Theoretical Framework

The strategic environment studied in this paper builds on Landeo and Spier's (2020) theoretical work on ordered leniency. They present a general model of law enforcement involving an enforcement agency (the designer of the ordered-leniency policy) and groups of injurers with multiple members. The potential injurers maximize their net benefits, or the private economic benefit of committing the act minus the expected value of the fine or sanction. The enforcement agency maximizes social welfare, which includes the injurers' private benefits, the harms inflicted on others (externalities associated with the harmful activities), and the costs of enforcement. Landeo and Spier (2020) show that the socially optimal ordered-leniency policy grants a cascade of reduced sanctions corresponding to the positions in the self-reporting queue.[19] Ordered leniency policies create a "race to the courthouse" where injurers self-report promptly, increasing the rate and speed of detection. We focus on groups of potential injurers with two members.[20]

Consider the following timing. First, the enforcement agency publicly commits to an enforcement policy with ordered leniency. The policy has three components. (1) A fine or monetary sanction $f > 0$ (measured per injurer). (2) Leniency multipliers $r_1, r_2 \in [0, 1]$ that correspond to the first and second positions in the self-reporting queue, respectively. The discount for position i in the self-reporting queue is $1 - r_i, i = 1, 2$. (3) The enforcement effort (investigation cost) $e \geq 0$ determines the probability harmful acts are detected.

Second, after observing the enforcement policy, the potential injurers play a two-stage game. In Stage 1, the Participation Stage, the potential injurers observe the private benefit of committing the act $b > 0$ (measured per injurer), and decide simultaneously and independently whether to

[19] Leniency for the first injurer to report may be less than full, and the second and subsequent injurer to self-report may receive lenient treatment as well (albeit to a lesser degree).

[20] Although Landeo and Spier's (2020) results are general, the key insights are captured in a strategic environment with just two group members. This simplifying assumption avoids unnecessary complexity in the experimental design and reduces subjects' computational costs.

participate.[21] If both injurers decide to participate, then the act is committed and Stage 2 begins; otherwise, the game ends. In Stage 2, the Self-Reporting Stage, the injurers decide simultaneously and independently whether and when to report the harmful act to the enforcement agency. The injurers select their reporting times, $t_1, t_2 \in [0, \bar{t}]$ where $t_i = 0$ corresponds to immediate self-reporting. The social harm associated with the act $h > 0$ (measured per injurer) is smaller when injurers self-report promptly and acts are detected sooner.[22] Prompt self-reporting, and hence faster detection, allows the enforcer to mitigate the harm h inflicted on others.

Third, the injurers are detected by the enforcement agency and sanctioned. If neither injurer self-reports, the injurers are detected with probability p_0 so each injurer pays expected fine $p_0 f$. If exactly one injurer reports the act, the injurer who self-reports pays fine $r_1 f$ and the silent co-conspirator is detected with probability $p_1 > p_0$ and pays expected fine $p_1 f > p_0 f$. In other words, there are negative externalities between the injurers.[23] If both injurers report the act, the injurers pay fines $r_1 f$ and $r_2 f$ depending on the order their reports are received. In the event of a tie, a coin flip (equally weighted) determines their positions in the self-reporting queue and the expected fine is $\frac{1}{2}(r_1 + r_2)f$.

Finally, the act is detected with certainty at time $\min\{t_1, t_2\}$ when both injurers decide to self-report and with certainty at time t_i when only injurer i self-reports.[24] If neither injurer self-reports, the act is detected with probability p_0 at time $t = \bar{t}$.

2.1 Model Parameterization

To minimize subjects' computational costs, and given that the purpose of this study is to assess the effectiveness of ordered-leniency mechanisms in detecting and deterring short-term group acts, we take the enforcement agency's decisions as exogenous and focus on the potential injurers' participation and self-reporting decisions. We proceed as follows. First, we assign particular numerical values to the model parameters.[25] Second, following Landeo and Spier (2020), we identify the optimal ordered-leniency policies for the parameter values. Third, we use these leniency policies to construct the experimental treatments.

The following parameter values are considered: Economic benefits from the act $b \in [200, 1600]$; a sanction or fine $f = 900$; probabilities of detection of an injurer $p_0 = 0.4$ and $p_1 = 0.9$, when no injurer self-reports and when the other injurer self-reports, respectively; time to self-report the

[21] Landeo and Spier's (2020) framework allows for benefit asymmetries and Coasian bargaining in Stage 1.

[22] In Landeo and Spier (2020), $\bar{t}$ is normalized to 1 and the harm h is a function of the time of detection, which is $\min\{t_1, t_2\}$ when the injurers both self-report at times t_1 and t_2.

[23] The information provided by the first injurer to report helps the enforcement agency to detect and punish the silent conspirator. In Landeo and Spier (2020), p_0 and p_1 depend on the enforcement agency's effort e and the strength of inculpatory evidence.

[24] When one injurer self reports, the silent accomplice is detected and punished with probability p_1.

[25] The strategic environment corresponds to Landeo and Spier (2020) Proposition 1, Case 2.

act $t \in [0, 90]$ (measured in seconds); and, time to detect the act (conditional on detection) when neither injurer decides to self-report $t = 90$.[26]

With these parameter values, the optimal ordered-leniency policy gives no leniency to the second injurer to report the act, $r_2 = 1$. The optimal degree of leniency granted to the first injurer who self-reports, r_1, depends on the criterion for equilibrium selection when multiple equilibria arise.[27] As shown in Landeo and Spier (2020), when $r_1 \in (0.400, 0.800)$ the Self-Reporting Stage resembles a coordination game with two pure-strategy Nash equilibria: (1) the Pareto-dominant equilibrium where neither injurer self-reports and (2) the risk-dominant equilibrium (Harsanyi and Selten 1988) where both injurers self report. If the injurers coordinate on Pareto-dominant equilibria, the enforcement agency must grant a large fine reduction for self-reporting (Strong Leniency), transforming the coordination game into a prisoners' dilemma where both injurers self-report. If the injurers coordinate on risk-dominant equilibria, then a small fine reduction will suffice to induce self-reporting (Mild Leniency).

Three strategic environments are studied in this experiment: Strong Leniency ($r_1 = 0.333$ and $r_2 = 1$), Mild Leniency ($r_1 = 0.466$ and $r_2 = 1$), and No Leniency ($r_1 = r_2 = 1$).[28] The Strong and Mild Leniency treatments correspond to the optimal ordered-leniency policies under the Pareto- and risk-dominance refinements.[29] Each leniency environment involves two Stages, the Participation Stage (Stage 1) and the Self-Reporting Stage (Stage 2).

2.2 Equilibrium Analysis

We characterize the equilibria for each strategic environment: Strong Leniency, Mild Leniency, and No Leniency. The equilibrium concept is subgame-perfect Nash equilibrium. We apply backward induction and start by analyzing the equilibrium behavior of the injurers in Stage 2. We then study the potential injurers' equilibrium decisions regarding participating in the act in Stage 1. We focus on pure-strategy equilibria that survive the elimination of weakly dominated strategies. Without loss of generality, we assume that when the potential injurer is indifferent between participating and not participating in the act, they decide not to participate in the act. Formal proofs are presented in Appendix 1.[30]

[26] The parameter values satisfy the assumptions of Landeo and Spier's (2020) model, and the predictions derived from these assumptions hold.

[27] See Landeo and Spier's (2020) Proposition 1, Case 2.

[28] The large fine reduction under Strong Leniency, $1 - r_1 = 0.667$, creates a prisoners' dilemma in Stage 2. The smaller fine reduction under Mild Leniency, $1 - r_1 = 0.533$, creates a coordination game in Stage 2.

[29] See Landeo and Spier's (2020) Proposition 1, Case 2. With Pareto-dominance, the optimal multipliers are $(r_1^S, r_2^S) = (p_0, 1) = (0.400, 1)$; with risk dominance they are $(r_1^M, r_2^M) = \left(\frac{1}{3}[2(p_1 + p_0) - 1], 1\right) = (0.533, 1)$. To break indifference in the self-reporting stage when Strong Leniency is applied, while keeping the difference between r_1^S and r_2^S constant, we adopt leniency multipliers $r_1 - \varepsilon$ where $\varepsilon = 0.067$.

[30] For general versions of lemmas and propositions, see Landeo and Spier (2020, 2018).

Self-Reporting Stage.– In Stage 2, each injurer decides simultaneously and independently whether to self report and the time to submit the report, $t \in [0, 90]$. We first analyze the length of time taken by the injurers to report the harmful act.

Lemma 1. *When Strong or Mild Leniency for self-reporting is granted, an injurer who reports the act will do so promptly, $t = 0$. When No Leniency for self-reporting is granted, an injurer who reports the act may do so at any time, $t \in [0, 90]$.*

According to Lemma 1, with Strong Leniency and Mild Leniency, any injurer who reports the act will so immediately, at time $t = 0$.[31] Since the first injurer to report the act receives leniency, $r_1 < 1$, but the second to report does not, $r_2 = 1$, late self-reporting at time $t > 0$ is a weakly-dominated strategy.[32] Lemma 1 implies that if both injurers self-report they are equally likely to get the first position in the self-reporting queue.[33]

Second, we study the injurers' decision whether to report the act. The strategic-form representation of the self-reporting subgame is shown in Table 1. When Strong Leniency is implemented, the injurers confront a prisoners' dilemma game. If neither injurer self-reports, the expected payoffs are $b - p_0 f = b - 360$; if one injurer self-reports the injurer who reports gets $b - r_1 f = b - 300$ and the second injurer gets $b - p_1 f = b - 810$; if both injurers self-report they are equally likely to be first in the self-reporting queue and so each gets expected payoff $b - 600$.[34] With Strong Leniency, no reporting is a strictly dominated strategy. Hence, the unique Nash equilibrium is the less-preferred outcome where both injurers self-report the act, i.e., the (R, R) outcome.

As shown in Table 1, When Mild Leniency is implemented, the injurers confront a coordination game. If neither injurer self-reports, the expected payoffs are $b - p_0 f = b - 360$; if one injurer self-reports the injurer who reports gets $b - r_1 f = b - 420$ and the second injurer gets $b - p_1 f = b - 810$; if both injurers self-report they are equally likely to be first in the self-reporting queue and so each gets expected payoff $b - 660$.[35] The Pareto-dominant Nash equilibrium is (NR, NR) and the risk-dominant Nash equilibrium is (R, R). Then, self-reporting by both injurers occurs only when both injurers fail to coordinate on their preferred outcome. When leniency for self-reporting is

[31] As shown in Landeo and Spier (2020), this follows from the elimination of weakly dominated strategies. 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.

[32] With No Leniency, there is no advantage from being the first to report since the fine remains at 900.

[33] By assumption, an equally weighted coin flip determines who obtains the first position.

[34] With Strong Leniency, $(r_1, r_2) = (0.333, 1)$. If neither self reports, the expected fine is $p_0 f = (0.4)900 = 360$. If one injurer self reports the (expected) fines are $r_1 f = (0.333)900 = 300$ and $p_1 f = (0.9)900 = 810$. If both self-report the (expected) fines are $\frac{1}{2}(r_1 + r_2)f = \frac{1}{2}(0.333 + 1)900 = 600$.

[35] With Mild Leniency, $(r_1, r_2) = (0.466, 1)$. If neither self reports, the expected fine is $p_0 f = (0.4)900 = 360$. If one injurer self reports the (expected) fines are $r_1 f = (0.467)900 = 420$ and $p_1 f = (0.9)900 = 810$. If both self-report the (expected) fines are $\frac{1}{2}(r_1 + r_2)f = \frac{1}{2}(0.466 + 1)900 = 660$.

Table 1 – Strategic-Form Representation of the Self-Reporting Subgame
(Expected Payoffs under Parameter Values)

Strong Leniency (S)

	No Report (NR)	Report (R)
No Report (NR)	$b-360$, $b-360$	$b-810$, $b-300$
Report (R)	$b-300$, $b-810$	$b-600$, $b-600$

Mild Leniency (M)

	No Report (NR)	Report (R)
No Report (NR)	$b-360$, $b-360$	$b-810$, $b-420$
Report (R)	$b-420$, $b-810$	$b-660$, $b-660$

No Leniency (N)

	No Report (NR)	Report (R)
No Report (NR)	$b-360$, $b-360$	$b-810$, $b-900$
Report (R)	$b-900$, $b-810$	$b-900$, $b-900$

Notes: For Mild Leniency, the expected payoffs under the risk-dominance refinement are presented.

not granted (No Leniency), self-reporting is a strictly dominated strategy, and the unique Nash equilibrium is (NR, NR). Hence, neither injurer self-reports the act.

Proposition 1 characterizes the equilibria of the self-reporting subgame.

Proposition 1. *The Nash equilibria for the self-reporting subgame for the three leniency environments are as follows.*

1. *Suppose Strong Leniency for self-reporting is granted. Both injurers self-report (R, R). Each injurer is detected with 100% likelihood at* $t = 0$*. The expected fine is 600.*
2. *Suppose Mild Leniency for self-reporting is granted.*
   (a) *Pareto-dominance refinement: Neither injurer self-reports (NR, NR). Each injurer is detected with 40% likelihood at* $t = 90$*. The expected fine is 360.*
   (b) *Risk-dominance refinement: Both injurers self report (R,R). Each injurer is detected with 100% likelihood at* $t = 0$*. The expected fine is 660.*
3. *Suppose No Leniency for self-reporting is granted. Neither injurer self-reports, (NR, NR). Each injurer is detected with 40% likelihood at* $t = 90$*. The expected fine is 360.*

Proposition 1 underscores two potential social benefits from ordered-leniency polices. First, ordered-leniency policies facilitate earlier detection of socially harmful acts. In the absence of ordered leniency, neither injurer self-reports and the harmful act is detected with 40% likelihood at the latest possible time, $t = 90$. Since the harmful act is detected late (or not at all), the level of social harm is high. Conversely, when Strong Leniency or Mild Leniency is granted and the risk-dominance refinement is applied, both injurers self-report immediately. As a result of this race to the courthouse, the harmful act is detected at the earliest possible time, $t = 0$. Importantly, early detection of the harmful act allows public and private actors to take steps to mitigate the social harms.

Corollary 1. *A harmful act is detected at $t = 0$ with 100% likelihood under Strong Leniency or under Mild Leniency with the risk-dominance refinement. A harmful act is detected at $t = 90$ with 40% likelihood under No Leniency or Mild Leniency with the Pareto-dominance refinement.*

Second, ordered-leniency policies increase the expected fines paid by the injurers. With No Leniency, neither injurer has an incentive to report the act and so each injurer pays an expected fine $p_0 f = (0.40)(900) = 360$. When Strong or Mild Leniency is granted and the risk-dominance refinement is applied, both injurers self-report and the probability that an injurer is detected increases (from 40% to 100%). Hence, although ordered leniency gives reduced fines to injurers who self-report and secure high positions on the self-reporting queue, the injurers will on average pay higher fines. This should, in theory, deter potential injurers from participating in socially harmful activities.

Participation Stage.– In Stage 1, the potential injurers decide simultaneously and independently whether to participate in the act. If both potential injurers decide to participate in the act, the act is committed and the injurers play the self-reporting subgame shown in Table 1. The potential injurers will choose to participate in the harmful act if (and only if) the private benefit of committing the act, b, is larger than the equilibrium expected fine characterized in Proposition 1. Thus, the expected fines in Proposition 1 represent the "deterrence thresholds" and are denoted by $\hat{b}$.[36] The deterrence thresholds for each leniency environment are as follows: $\hat{b} = 600$ for Strong Leniency; $\hat{b} = 660$ for Mild Leniency when the risk-dominance refinement is applied; and $\hat{b} = 360$ for No Leniency and Mild Leniency when the Pareto-dominance refinement is applied.

Proposition 2 characterizes the equilibria of the participation stage.

Proposition 2. *The Nash equilibria of the participation stage for the three leniency environments are as follows.*

[36] When the individual benefit of committing the act, b, is greater than the deterrence threshold, $\hat{b}$, then participating in the activity is a weakly-dominant strategy. When b is smaller than the deterrence threshold, $\hat{b}$, then participating in the activity is a weakly-dominated strategy. Without loss of generality, we assume that when the potential injurer is indifferent between participating and not participating in the act, he decides not to participate in the act.

1. *Suppose Strong Leniency for self-reporting is granted. If $b > 600$ then both potential injurers participate; if $b \leq 600$ then neither potential injurer participates.*

2. *Suppose Mild Leniency for self-reporting is granted.*

   (a) *Pareto-dominance refinement: If $b > 360$ then both potential injurers participate; if $b \leq 360$ then neither potential injurer participates.*

   (b) *Risk-dominance refinement: If $b > 660$ both potential injurers participate; if $b \leq 660$ then neither potential injurer participates.*

3. *Suppose No Leniency for self-reporting is granted. If $b > 360$ then both potential injurers participate; if $b \leq 360$ then neither potential injurer participates.*

Finally, note that the harmful group act is committed if and only if *both* potential injurers are willing to participate in the act. Hence, in theory, across leniency environments, a harmful act will always be deterred when the private benefit from committing the act is smaller than the deterrence threshold (the expected fine). Corollary 2 outlines the equilibrium act deterrence rates.[37]

Corollary 2. *A harmful act is always deterred when $b \leq 600$ under Strong Leniency, when $b \leq 660$ under Mild Leniency with the risk-dominance refinement, and when $b \leq 360$ under No Leniency or Mild Leniency with the Pareto-dominance refinement.*

Table 2 summarizes the theoretical point predictions for individual injurers for the three leniency environments, including the rates and times of self-reporting and detection, the expected fines, and the injurer participation rates.

3 Experimental Design

Our experiment is designed to assess the validity of the qualitative predictions of the theory. First, our study analyzes the effects of ordered-leniency mechanisms on the self-reporting and the detection of injurers. Second, our study evaluates whether ordered leniency reduces the time to detect the act, i.e., whether ordered leniency creates a race-to-the-courthouse effect. Third, it analyzes whether the potential injurer's willingness to participate in the act is higher when the

[37] In theory, the decision to participate in the act is the same for both potential injurers: Both potential injurers refuse to participate in the act when the benefit associated with the act is below the deterrence threshold (expected fine). Therefore, the potential injurer's refusal to participate rate (individual-level analysis) and the act deterrence rate (group-level analysis) are the same. In the lab, the decisions of the two potential group members might differ. Hence, the individual-level and group-level rates might be different.

Table 2 – Theoretical Point Predictions - Individual Injurers

Leniency Environment	Report		Injurer Detection		Expected Fine[e]	Participation Rate	
	Rate[a]	Time[b]	Rate[c]	Time[d]	($\hat{b}$)	$b \leq \hat{b}$	$b > \hat{b}$
Strong Leniency (S)	1	0	1	0	600	0	1
Mild Leniency (M)							
• Risk Dominance	1	0	1	0	660	0	1
• Pareto Dominance	0	–	.40	90	360	0	1
No Leniency (N)	0	–	.40	90	360	0	1

Notes: For Mild Leniency, theoretical predictions under Pareto-dominance and risk-dominance refinements are presented.

[a]Report rate conditional on committing the act.

[b]Report time (in seconds) conditional on committing the act and reporting the act.

[c]Detection rate conditional on committing the act.

[d]Detection time (in seconds) conditional on committing the act and detecting the act.

[e]Expected fine conditional on committing the act.

private economic benefit from the act is above the deterrence threshold (expected fine). Fourth, our study evaluates the effects of ordered-leniency policies on the potential injurers' willingness to participate in the act.

We specify the experimental setting in a way that satisfies the assumptions of the theory. We use abstract (neutral) and meaningful (but not emotion-evocative) labels. In other words, we use a combination of first- and second-level contexts (Alekseev et al., 2017).[38] Specifically, neutral labels are used to denote the subjects' roles: Players B1 and B2 (which refer to injurers 1 and 2, respectively).[39] The economic decision involving economic benefits (realized in Stage 1) and economic losses (realized in Stage 2) is described as "The Act." The economic losses are described as "Paying a Fine," and the likelihood of detection is described as the "Chance to Pay a Fine." The players' choices in Stage 1 are described as "To Agree to Jointly Commit the Act" and "Not to Agree to Jointly Commit the Act," and the players' choices in Stage 2 are described as "To

[38]As defined by Alekseev et al. (2017), first-level contexts use abstract or neutral labels. Second-level contexts "[use] terms that can be related to a real-life situation ... These terms have a meaning, but do not evoke strong emotions or connotations" (p. 56). Alekseev and colleagues study the role of evocative terms in experimental economics studies, and conclude that "the use of evocative terms is more likely to be irrelevant in simple games, whose structure is suggestive of the situation being modeled [such as p]articular forms of bribery games and tax-compliance games" (p. 57). Previous studies regarding the effects of context in experiments on corruption suggest that neutral contexts (instead of framed contexts) do not have any effects (Abbink and Hennig-Schmidt, 2006) or have just a small effect (Barr et al., 2004).

[39]Labels such as "Wrongdoer 1" and "Wrongdoer 2" will not contribute to the subjects' understanding of the experimental environment, and might generate noise in the subjects' responses due to the degree of identification with the role described by the label.

Report the Act" and "Not to Report the Act."[40] Our simple contextual environment ensures the subjects' understanding of the environment as well as control and replicability.[41]

The experimental design consists of three leniency environments: Strong Leniency (S), where the fine reduction for self-reporting is large; Mild Leniency (M), where the fine reduction for self-reporting is small; and, No Leniency (N), where a fine reduction for self-reporting is not granted. A concern with our study, a concern that is common to all experimental research, is its external validity. Although our experiment cannot predict the effects of ordered-leniency policies in richer environments, the experiment provides evidence regarding whether the three leniency mechanisms in an environment such as the one we have structured here will have the predicted effects.

The Games.– Procedural regularity is accomplished by developing a software program that allows the subjects to play the game by using networked personal computers. The software, constructed using the Java programming language, consists of 3 versions of the game reflecting the three experimental conditions: Strong Leniency, Mild Leniency, and No Leniency.[42]

The experiment involves a two-player, two-stage game of complete information. The benchmark game corresponds to the Strong Leniency condition (S). Subjects play the role of Player B1 or Player B2. The roles of Players B1 and B2 are similar. Each match involves two stages. In Stage 1, each player independently decides whether to participate or not participate in the act. The players have 90 seconds to submit their decisions in Stage 1. After the decisions are made, both players are informed about the other player's decision. If both players agree to participate in the act, then the act is committed and Stage 2 starts. Otherwise, the game ends. In Stage 2, each player independently decides whether to report or not report the act. The players have 90 seconds to submit their decisions in Stage 2.[43] When both players decide to report at the same time, the computer randomly assigns the first position in the self-reporting queue to each player with equal probability. After the decisions are made, both players are informed about the decision of the other player and the payoffs for both players, and the game ends.

The subjects play a similar game across experimental conditions. The only difference refers to the fine reduction granted to the first player to self-report the act. Specifically, in the Strong Leniency condition (S), the first player to self-report receives a large fine reduction (67%); in the Mild Leniency condition (M), the first player to self-report receives a small fine reduction (53%);

[40]See Appendix 2 for a sample of the instructions (Mild Leniency condition).

[41]Note that a minimum context is required to replicate the theoretical environment in the lab. If the findings in this simple environment do not conform to the theory, there is little hope that this theory can explain subjects' behavior in more complex settings (see Douglas Davis and Charles Holt 1993). Hence, experiments provide useful feedback to theorists.

[42]The use of a software especially designed for this study allows us to have flexibility in the design of randomization processes and the design of user-friendly screens. A complete set of software screens is available from the authors upon request.

[43]This information will be used to assess whether ordered-leniency policies reduce the time to detect the act.

Table 3 – Participants per Experimental Condition

b-Value Segments	Strong Leniency (S)	Mild Leniency (M)	No Leniency (L)
$b \in [200, 360]$	8	8	8
$b \in (360, 600]$	22	22	22
$b \in (600, 660]$	22	22	22
$b \in (660, 1600]$	44	44	44
Total Number of Individuals	96	96	96

and, in the No Leniency condition (N), the first player to self-report does not receive any fine reduction.

The Experimental Sessions.– We ran twelve 80-minute sessions of 24 subjects each (four sessions per condition; 96 subjects per condition; 288 subjects in total) at Harvard University. Table 3 summarizes the information regarding participants per experimental condition and the observations per b-value segment.[44]

Each session was conducted by two research assistants at the Harvard Decision Science Laboratory. Subjects were recruited using the lab's Sona computer program and the lab's subject pool. Subjects were allowed to participate in one experimental session only, and received information only about the game version that they were assigned to play. The participant pool included undergraduate and graduate students from Harvard University, Boston University and Northeastern University, from a wide variety of fields of study. A laboratory currency called the "token" (29 tokens = 1 U.S. dollar) was used in our experiment. To avoid negative payoffs, each subject received an initial endowment equal to 700 tokens.[45] The show-up fee was equal to $10. The average game earnings were equal to $32. Hence, the average total payment was equal to $42 (average game earnings plus show-up fee) for an 80-minute session.

At the beginning of each session, written instructions were provided to the subjects (see Appendix 2). The instructions about the game and the software were verbally presented by the research assistant to create common knowledge. Subjects were informed: (1) about the game structure, possible choices, and payoffs; (2) about the random process of allocating roles; (3) about the randomness and anonymity of the process of forming pairs;[46] (4) about the token/dollar

[44]The theoretical deterrence thresholds guide the design of the distribution of b-values. The adopted distribution of b-values allows us to collect a sufficiently high number of observations to perform statistical analyses of deterrence and detection across conditions. To ensure comparability across conditions, we randomly predetermine the b-values used in the actual match and practice matches, and apply these values across conditions.

[45]Note that the minimum possible b-value was equal to 200 tokens and the maximum possible fine was equal to 900 tokens. Then, with an initial endowment of 700, the minimum possible match payoff was equal to 700 + 200 - 900 = 0 tokens.

[46]Subjects were informed: (1) that they would not play with the same partner in any practice match; and, (2) that they would not play with any of their previous partners in the actual match.

equivalence, and that they would receive the dollar equivalent of the tokens they held at the end of the session. The subjects were then asked to complete a set of exercises to ensure their ability to read the information tables. The answers to the exercises were read aloud by the research assistant. Questions about the written instructions and questions about the exercises were answered by the research assistant privately and before the beginning of the practice matches. The rest of the session was entirely played using computer terminals and the software designed for this experiment.

Each experimental session included five practice matches and one actual match. The practice matches allowed the subjects to experiment with the different options and hence, learn about the consequences of their choices. Only the actual match was considered in the subject's payment. After the actual match, subjects were required to fill out a short questionnaire with general demographic questions. At the end of each experimental session, subjects privately received their monetary payoffs in cash.

4 Qualitative Hypotheses

The qualitative hypotheses derived from the theory are presented below. The hypotheses related to Mild Leniency are constructed under the risk-dominance refinement. A discussion of the relevant comparisons under the risk-dominance and Pareto-dominance refinements is also presented.

Hypothesis 1. *Strong and Mild Leniency increase the report rate, with respect to No Leniency. Strong and Mild Leniency exhibit the same report rate.*

When Strong or Mild Leniency for self-reporting is granted and the risk-dominance refinement is applied, both injurers self-report the act. In contrast, when leniency for self-reporting is not granted or when Mild Leniency for self-reporting is granted and the Pareto-dominance refinement is applied, neither injurer self-reports the act. Therefore, both ordered-leniency policies raise the report rate when the risk-dominance refinement is applied.

Hypothesis 2. *Strong and Mild Leniency increase the injurer's detection rate, with respect to No Leniency. Strong and Mild Leniency exhibit the same detection rate.*

When Strong Leniency or Mild Leniency for self-reporting is granted and the risk-dominance refinement is applied, an injurer is always detected. In contrast, when leniency for self-reporting is not granted or Mild Leniency for self-reporting is granted and the Pareto-dominance refinement is applied, an injurer is detected with 40% probability. Hence, both ordered-leniency policies raise the detection rate when the risk-dominance refinement is applied.

Hypothesis 3. *Within each leniency environment, the potential injurer's willingness to participate in the act rate is higher when the benefit from the act is greater than the deterrence threshold.*

Across leniency treatments, the potential injurer decides to participate in the act when the benefit is greater than the deterrence threshold (expected fine). Hence, the participation rate is higher when the benefit is greater that the deterrence threshold (100% versus zero).

Hypothesis 4. *Mild Leniency decreases the potential injurer's willingness to participate in the act rate, with respect to Strong Leniency and No Leniency. Strong Leniency decreases the potential injurer's willingness to participate in the act rate, with respect to No Leniency.*

The deterrence threshold for Mild Leniency when the risk-dominance refinement is applied is higher than the deterrence thresholds for Strong Leniency and No Leniency (660 v. 600, 660 v. 360, respectively); and, the deterrence threshold for Strong Leniency is higher than the deterrence threshold for No Leniency (600 v. 360). In contrast, the deterrence thresholds for No Leniency and Mild Leniency when the Pareto-dominance refinement is applied are equal (360). Hence, both ordered-leniency policies reduce the participation rates when the risk-dominance refinement is applied.[47]

Hypothesis 5. *Strong and Mild Leniency decrease the time to detect the act, with respect to No Leniency.*

Strong and Mild Leniency when the risk-dominance refinement is applied generate a "race-to-the-courthouse" where both injurers compete against each other to secure the first position in the self-reporting queue and, hence, both self-report the act immediately. Consequently, the act is detected at $t = 0$ with certainty. In contrast, when leniency for self-reporting is not granted or Mild Leniency is granted and the Pareto-dominance refinement is applied, neither injurer self-reports the act. As a result, the act is detected at $t = 90$ with 40% probability. Hence, both ordered-leniency policies have the potential to mitigate social harm through early detection.

5 Results

The main experimental findings are presented in a series of results.

5.1 Data Summary

Table 4 provides descriptive statistics (individual-level data) for the report rate, report time, detection rate, fine, and willingness to participate in the act rate (for b-values below, above the theoretical deterrence threshold $\hat{b}$, and overall b-values). The report rate is defined as the

[47] Under Strong and Mild Leniency, the expected fine depends on the probability of detection and the players' beliefs about securing the first position in the self-reporting queue. Judgement errors in the players' beliefs might arise in these complex environments (Larrick et al., 2007; Kunda, 1990, 1987).

Table 4 – Descriptive Statistics - Individual Injurers

Cond.	Report Rate[a]	Report Time[b] R	Report Time[b] NR	Injurer Detection Rate[c]	Fine[d]	Willingness-to-Participate Rate $b \leq \hat{b}$	$b > \hat{b}$	All b	Th.[e]
Strong Leniency	.90	2.14	15.25	.99	588.46	.63	.97	.86	.69
[96]		(1.30)	(19.62)		(307.47)				
Mild Leniency	.76	2.71	12.29	.96	638.25	.79	1.00	.89	.46
[96]		(2.70)	(17.28)		(268.71)				
No Leniency	.06	16.75	16.21	.53	475.71	.50	.87	.84	.92
[96]		(22.23)	(17.67)		(452.51)				

Notes: Total number of individuals are in brackets, and standard deviation are in parentheses.
[a]Report rate conditional on committing the act.
[b]Mean time to submit the report decision (R or NR) in Stage 2 conditional on committing the act (in seconds).
[c]Injurer detection rate conditional on committing the act.
[d]Mean fine conditional on committing the act.
[e]Theoretical willingness-to-participate rate across b-values; theoretical willingness-to-participate rate for Mild Leniency under the risk-dominant refinement.

percentage of individuals who decided to report the act, conditional on committing the act. The report time is defined as the average individual time to submit the report decision (R or NR) in Stage 2 (in seconds), conditional on committing the act.[48] The detection rate is defined as the percentage of individuals who were detected, conditional on committing the act. The fine corresponds to the average individual fine, conditional on committing the act. The willingness-to-participate rate when $b \leq \hat{b}$ (or $b > \hat{b}$ or All b) is defined as the percentage of individuals with $b \leq \hat{b}$ (or $b > \hat{b}$ or $b \in [200, 1600]$) who decided to participate in the act. The theoretical willingness-to-participate rate corresponds to the rate predicted by the theory across b-values.[49]

Table 5 provides descriptive statistics (group-level data) for the act detection time, act detection rate, and act deterrence rate across b-values. The act detection time is defined as the average act detection time, conditional on committing the act.[50] The act detection rate is defined as the percentage of acts (groups of injurers) that were detected, conditional on committing the act. The act deterrence rate is defined as the percentage of acts (groups of potential injurers) that

[48]To control for any possible differences in connectivity time between the lab server and the terminals, we decided to place all participants that submitted their decision in Stage 2 within the same second in the same "time interval." For instance, if two participants submitted their decisions between the first and second seconds, i.e., $t \in [1, 2)$ seconds, then the recorded "time interval" will be the same for these two participants. We use the midpoint of each time interval in the statistical analysis of results.

[49]Given the b-values per segment and the total number of individuals per condition, the theoretical willingness-to-participate rates are equal to 69% (66/96), 46% (44/96) and 92% (88/96), for Strong Leniency, Mild Leniency and No-Leniency, respectively.

[50]The act detection time is equal to the minimum time to submit the decision in Stage 2 among group members (in seconds) when one or both members decided to self-report the act, and equal to 90 seconds when both members decided not to report the act.

Table 5 – Descriptive Statistics - Act (Group of Injurers)

Cond.	Act Detection		Act Deterrence	
	Time[a]	Rate[b]	Rate	Th. Rate[c]
Strong Leniency	1.78	1.00	.19	.31
[48]	(.86)			
Mild Leniency	4.87	.98	.17	.54
[48]	(14.35)			
No Leniency	74.58	.54	.27	.08
[48]	(32.00)			

Notes: Total number of groups are in brackets, and standard deviation are in parentheses.
[a]Act detection time conditional on committing the act and detection.
[b]Act detection rate conditional on committing the act.
[c]Theoretical act deterrence rate across b-values; act deterrence for Mild Leniency under the risk-dominant refinement.

are not committed, across b-values. The theoretical act deterrence rate corresponds to the group deterrence rate predicted by the theory across b-values.[51]

Overall, our results are aligned with the theoretical predictions under the risk-dominance refinement. The data indicate that the implementation of ordered-leniency policies increased self-reporting (90 versus 6%, Strong Leniency and No Leniency, respectively; and, 76 versus 6%, Mild Leniency and No Leniency, respectively). The average times to submit the report decision were equal to 2.14, 2.71, and 16.75 seconds, for Strong Leniency, Mild Leniency, and No Leniency, respectively. These findings suggest that the ordered-leniency policies induced the injurers to self-report promptly to increase their chances to get the first position in the self-reporting queue. The data also indicate that the implementation of ordered-leniency policies increased the rate of detection of injurers (99 versus 53%, for Strong Leniency and No Leniency, respectively; and, 96 versus 53%, for Mild Leniency and No Leniency, respectively). The high detection rates under Strong and Mild Leniency are explained by the relatively high self-reporting rates under ordered leniency. Importantly, the data suggest that ordered leniency policies dramatically reduced the time to detect the act (1.78 versus 74.58, for Strong Leniency and No Leniency, respectively; and, 4.87 versus 74.58, for Mild Leniency and No Leniency, respectively). These results provide evidence of a race-to-the-courthouse effect under ordered-leniency policies. By inducing early detection of the act, ordered-leniency policies have the potential to mitigate the harm inflicted on society.

As predicted by the theory, the data indicate that the potential injurers' willingness to participate in the act when $b > \hat{b}$ was higher than the rate when $b \leq \hat{b}$, across leniency conditions

[51]Given the b-values per segment and the total number of groups per condition, the theoretical act deterrence rates are equal to 31% (15/48), 54% (26/48) and 8% (4/48), for Strong Leniency, Mild Leniency and No-Leniency, respectively. In theory, the individual reluctance to participate rate and the group deterrence rate are the same.

Table 6 – Effects of Treatments on the Probability of Self-Reporting
(Tests of Differences Across Conditions)

	N versus S (Mg. Effect)		N versus M (Mg. Effect)		M versus S (Mg. Effect)
Strong Leniency	.84***	Mild Leniency	.71***	Strong Leniency	.13*
	(.06)		(.05)		(.08)
Observations	148	Observations	150	Observations	158

Notes: The columns report the change in the probability of self-reporting (conditional on committing the act) due to S (with respect to N), M (with respect to N), and S (with respect to M); probit analysis using sessions as clusters; marginal effects are reported; robust standard errors are in parentheses; observations correspond to number of individuals who committed the act.

***Significant at the 1 percent level.

*Significant at the 10 percent level.

(63 versus 97%, 79 versus 100%, and 50 versus 87%, for Strong Leniency, Mild Leniency and No Leniency, respectively). In other words, the benefits from the act induced the potential injurers to participate in the act. The data also suggest that the potential injurers' willingness-to-participate rates when $b \leq \hat{b}$ for Strong Leniency and Mild Leniency were higher than the predicted by the theory when the risk-dominance refinement is applied (63 and 79% instead of zero, for Strong Leniency and Mild Lenience, respectively). These regular departures from the theory at the participation stage might suggest the presence of systematic underestimation of the deterrence threshold (expected fine).

5.2 Data Analysis

Our regression analysis involves standard errors that are robust to general forms of heteroskedasticity and hence, account for the possible dependence of observations within session.

Self-Reporting.– Table 6 reports the results of the analysis of the effects of each leniency treatment on self-reporting. Each probit model includes a treatment dummy variable as its regressor. The treatment dummy variable is constructed as follows. For example, for the case of the probit model that assess the effect of Strong Leniency (with respect to No Leniency), the dummy variable will take a value equal to one if the observation pertains to the Strong Leniency condition, and a value equal to zero if the observation pertains to the No Leniency condition.[52] Marginal effects are reported here.[53]

The effect of Strong Leniency on the self-reporting rate is reported in the second column. Strong Leniency significantly increases the likelihood of self-reporting. In fact, as a result of implementing Strong Leniency, higher self-reporting rates are observed: 90 versus 6% for the

[52]The data for the Strong Leniency and No Leniency conditions are pooled to estimate this probit model.

[53]Given that probit magnitudes are difficult to interpret, we report the marginal effects.

Table 7 – Effects of Treatments on the Probability of Injurer's Detection
(Tests of Differences Across Conditions)

	N versus S (Mg. Effect)		N versus M (Mg. Effect)		M versus S (Mg. Effect)
Strong Leniency	.46***	Mild Leniency	.43***	Strong Leniency	.03**
	(.14)		(.13)		(.02)
Observations	148	Observations	150	Observations	158

Notes: The columns report the change in the probability of injurer's detection (conditional on committing the act) due to S (with respect to N), M (with respect to N), and S (with respect to M); probit analysis using sessions as clusters; marginal effects are reported; robust standard errors are in parentheses; observations correspond to number of individuals who committed the act.
***Significant at the 1 percent level.
**Significant at the 5 percent level.

Strong Leniency and No Leniency conditions, respectively. Hence, there is a clear support to Hypothesis 1.

Result 1. *Strong Leniency significantly increases the report rate, with respect to No Leniency.*

The results about the effect of Mild Leniency on the probability of self-reporting is reported in the fourth column. Mild Leniency significantly increases the likelihood of self-reporting. This result supports Hypothesis 1 under the risk-dominance refinement. The comparisons are 76 versus 6% for Mild Leniency and No Leniency conditions, respectively. Importantly, our results suggest that the enforcement agency can elicit injurers' cooperation without relying on large discounts for self-reporting (Strong Leniency), i.e., by implementing a coordination-game environment through Mild Leniency.

Result 2. *Mild Leniency significantly increases the report rate, with respect to No Leniency.*

The sixth column reports the results of Strong Leniency on the probability of self-reporting, with respect to Mild Leniency. Strong Leniency marginally increases the likelihood of self-reporting ($p = .078$): 90 versus 76%, for Strong and Mild Leniency, respectively. This result can be explained as follows. Under the Mild Leniency condition, although the majority of subjects chose the strategy associated with the risk-dominant equilibrium, (R, R), almost a quarter of subjects chose the strategy associated with the Pareto-dominance equilibrium, (NR, NR).

Result 3. *Strong Leniency marginally increases the report rate, with respect to Mild Leniency.*

Injurer's Detection.– Table 7 reports the results of the analysis of the effect of the leniency treatments on the probability of detection of an injurer, i.e., probit estimations. The second column indicates that Strong Leniency significantly increases the likelihood of injurer's detection (99 versus 53%, for Strong and No Leniency, respectively). The results of the effect of Mild Leniency on the likelihood of detection are presented in the fourth column. Mild Leniency significantly increases the detection rate (96 versus 53%, for Mild and No Leniency, respectively).[54] The sixth column reports the results of Strong Leniency on the probability of detection, with respect to Mild Leniency. Strong Leniency significantly increases the detection rate (99 versus 96%).[55] Our findings provide strong support to Hypothesis 2 under the risk-dominance refinement, and suggest that ordered leniency is a very effective detection mechanism.

Result 4. *Strong Leniency significantly increases the injurer's detection rate, with respect to No Leniency.*

Result 5. *Mild Leniency significantly increases the injurer's detection rate, with respect to No Leniency.*

Result 6. *Strong Leniency significantly increases the injurer's detection rate, with respect to Mild Leniency.*

Potential Injurer's Willingness to Participate in the Act.– We will now evaluate whether the injurer's willingness to participate in the act is higher when the private economic benefit from the act is above the deterrence threshold ($b > \hat{b}^i$, $i = S, M, N$). We define "above-threshold benefit" as a benefit above the deterrence threshold ($b > \hat{b}^i$, $i = S, M, N$) and construct a dummy variable that takes a value equal 1 if the observation involves a benefit above the deterrence threshold ($b > \hat{b}^i$, $i = S, M, N$) and zero otherwise.[56] Table 8 reports our findings regarding the effect of above-threshold benefit on the likelihood of potential injurer's willingness to participate in the act (within-condition analysis), i.e., probit estimations.

Our results are aligned with the theory. The effect of above-threshold benefit on the probability of potential injurer's willingness to participate in the act is positive and significant, across conditions: 63 versus 97%, 79 versus 100%, and 50 versus 87%, for Strong, Mild and No Leniency, respectively.[57] In other words, as predicted by the theory, the potential injurer's willingness to participate in the act rate is higher when the private economic benefit from the act is above the deterrence threshold (expected fine). Our findings provide clear support for Hypothesis 3.[58]

[54]Similarly, Mild Leniency significantly increases the probability of detection of the act (group-level analysis), with respect to No Leniency ($p < .000$). Given that the act detection rate is equal to 100% under Strong Leniency, probit estimations of the effects of Strong Leniency (with respect to No Leniency and Mild Leniency) cannot be implemented.

[55]p-value = .024.

[56]Under Mild Leniency, when $b > \hat{b}^M = 660$, the injurer's willingness-to-participate rate is equal to 100%. As a result, a probit analysis cannot be conducted. We use instead the deterrence rate for Strong Leniency, $\hat{b}^S = 600$. The injurer's willingness-to-participate rates are 70 and 97%, for $b \leq 600$ and $b > 600$, respectively.

[57]$p - value = .010$ for the probit model corresponding to No Leniency (N).

[58]Probit analysis with "benefit" b as a covariate provides similar results. The effect of the economic private

Table 8 – Effect of Above-Threshold Benefit on the Probability of Potential Injurer's Willingness to Participate (Within-Condition Analysis)

	S (Mg. Effect)		M (Mg. Effect)		N (Mg. Effect)
Above-Threshold b	.34***	Above-Threshold b	.27***	Above-Threshold b	.38**
	(.04)		(.07)		(.18)
Observations	96	Observations	96	Observations	96

Notes: The columns report the change in the probability of potential injurer's willingness to participate in the act due to above-threshold benefit (probit analysis using sessions as clusters); marginal effects are reported; robust standard errors are in parentheses; observations correspond to total number of individuals.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

Result 7. *Within each leniency condition, the potential injurer's willingness-to-participate rate is significantly higher when the benefit associated with the act is above the deterrence threshold.*

Table 9 reports the results of the analysis of the effects of the leniency treatments on the likelihood of potential injurer's willingness to participate in the act, i.e., probit estimations. Our findings indicate that the majority of potential injurers consider deterrence thresholds lower than the theoretical expected fines when making their participation decisions: when $b \leq \hat{b}$, the willingness-to-participate rates are equal to 63 and 79% (instead of zero), for Strong Leniency and Mild Leniency, respectively.[59] As a result, departures from the theory are observed in the lab: Strong Leniency does not significantly affect the willingness-to-participate rate, with respect to No Leniency and Mild Leniency; and, Mild Leniency significantly increases the willingness-to-participate rate, with respect to No Leniency (89 versus 84%, for Mild Leniency and No Leniency, respectively).[60]

Result 8. *Mild Leniency significantly increases the potential injurer's willingness-to-participate*

benefit from the act on the probability of potential injurer's willingness to participate in the act is positive and significant, across conditions.

[59]Given that 78% of injurers with $b \leq 660$ self-reported under Mild Leniency, the willingness to participate findings cannot be explained by the choice of the deterrence threshold when the Pareto-dominance refinement is applied.

[60]Kunda (1990, 1987) argues that environments characterized by strong ambiguity are more prone to elicit judgment errors. In our experiment, subjects might perceive a strong degree of ambiguity under the Strong and Mild Leniency conditions because the expected payoffs also depend also on the subjects' beliefs about their chances to be the first to report. Moreover, given that the Mild Leniency environment also involves multiple equilibria, it might be perceived as the one with the strongest degree of ambiguity. Consequently, we might expect more pervasive judgment errors and more prominent departures from the theoretical point predictions on deterrence in the Mild Leniency condition. Failure to apply backward induction due to limited computational abilities might explain some of the deviations from the theoretical predictions in the participation stage (Camerer and Johnson, 2004; Johnson et al., 2002).

Table 9 – Effects of Treatments on the Probability of Potential Injurer's Willingness to Participate (Tests of Differences Across Conditions)

	N versus S (Mg. Effect)		N versus M (Mg. Effect)		M versus S (Mg. Effect)
Strong Leniency	.02	Mild Leniency	.04**	Strong Leniency	−.02
	(.03)		(.02)		(.02)
Observations	192	Observations	192	Observations	192

Notes: The columns report the change in the probability of potential injurer's willingness to participate due to S (with respect to N), M (with respect to N), and S (with respect to M); probit analysis using sessions as clusters; marginal effects are reported; robust standard errors are in parentheses; observations correspond to total number of individuals.
**Significant at the 5 percent level.

rate, with respect to No Leniency.

The underestimation of the deterrence threshold (expected fine) might be explained by the potential injurers' beliefs about their chances to secure the first position in the self-reporting queue. A group member who believes that he has the ability to report the act faster than other members will place better-than-even odds on securing the first position. Potential injurers who believe that their chance of securing the first position is greater than 50% will consider a lower deterrence threshold than predicted. At the extreme, a potential injurer who is 100% confident of reporting the act first would consider fines equal to 300 and 420 instead of the theoretical expected fines of 600 and 660 when making the participation decision in the Strong and Mild Leniency environments, respectively.[61]

Time to Detect the Act.– Thus far, we have assessed the effects of ordered leniency on self-reporting, injurer's detection, and potential injurer's willingness to participate in the act using an individual-level analysis. We will now turn to a group-level analysis of the effects of ordered leniency on the time to detect the committed act. In theory, ordered leniency generates a "race to the courthouse" where prompt self-reporting occurs. As a result, an early detection of the act occurs.

Table 10 reports the results of the analysis of the effects of of each leniency treatment on the time to detect the act, i.e., linear regression estimations. The second column indicates that Strong Leniency significantly decreases the time to detect the act. In fact, the average times to detect the act under Strong Leniency is less than 2 seconds and the average time to detect the act under

[61]Some subjects might be more able than others and correctly believe that their odds of reporting first are better than fifty-fifty. Some other subjects might exhibit overconfidence in their abilities to report fast (Larrick et al., 2007). Benoît and Dubra (2011) and Benoît et al. (2015) argue that only beliefs that cannot be rationalized (i.e., beliefs that are not formed in an unbiased Bayesian manner) correspond to "true" overconfidence (i.e., judgment errors). An assessment of overconfidence is beyond the scope of this paper.

Table 10 – Effects of Treatments on the Time to Detect the Act
(Tests of Differences Across Conditions)

	N versus S (Mg. Effect)		N versus M (Mg. Effect)		M versus S (Mg. Effect)
Strong Leniency	-72.80^{***}	Mild Leniency	-69.71^{***}	Strong Leniency	-3.09
	(8.03)		(7.95)		(2.82)
Observations	58	Observations	58	Observations	78

Notes: The columns report the change in the mean time to detect the act due to S (with respect to N), M (with respect to N), and S (with respect to M); linear regression analysis using sessions as clusters; coefficients are reported; robust standard errors are in parentheses; observations correspond to total number of groups.
***Significant at the 1 percent level.

No Leniency is more than 74 seconds. These results, which are aligned with Hypothesis 5, might suggest that Strong Leniency generates a race-to-the-courthouse effect where self-reporting occurs promptly.

Result 9. *Strong Leniency significantly reduces the time to detect the act, with respect to No Leniency.*

The fourth column indicates that Mild Leniency significantly decreases the time to detect the act. In fact, as a result of the implementation of Mild Leniency, a lower time to detect the act is observed: The average time to detect the act under Mild Leniency is less than 5 seconds and the average time to detect the act under No Leniency is more than 74 seconds. These findings, which are aligned with Hypothesis 5, might suggest that Mild Leniency generates a race-to-the-courthouse effect where both injurers self-report promptly. These results also provide additional evidence of the subjects' choice of the risk-dominant strategy.

Result 10. *Mild Leniency significantly reduces the time to detect the act, with respect to No Leniency.*

The sixth column indicates that the times to detect the act under Strong and Mild Leniency are not significantly different. This finding is aligned with the theory under the risk-dominance refinement.

Overall, our results demonstrate the effectiveness of ordered-leniency policies as detection mechanisms and provide strong support to the theory. Importantly, our findings indicate that the decisions of the majority of injurers in case of multiplicity of equilibria are aligned with the risk-dominant refinement. As a result, the implementation of either Mild Leniency or Strong Leniency policies leads to a higher likelihood of self-reporting and detection and punishment of injurers. By inducing a faster detection of the act, ordered-leniency policies have the potential to mitigate the harm inflicted on society.

6 Summary and Conclusions

This paper experimentally studies the effects of ordered-leniency policies on detecting and deterring short-term acts committed by groups of injurers. The strategic environment studied in this paper builds on Landeo and Spier's (2020) theoretical work on law enforcement with ordered leniency for self-reporting. In theory, ordered-leniency policies, where the degree of leniency depends on the injurer's position in a self-reporting queue, induce prompt self-reporting. As a result, early detection of harmful acts is enhanced and social harm is mitigated. Given the higher detection rates under ordered leniency, the expected sanctions also rise, and hence, deterrence and social welfare are further enhanced. Importantly, the design of optimal law enforcement policies with ordered leniency depends crucially on the refinement criterion for equilibrium selection when multiple equilibria arise. When the risk-dominance refinement is applied, the enforcement authority can induce all injurers to prompt self-report by offering only small penalty discounts (Mild Leniency). When the Pareto-dominance refinement is applied instead, the enforcement agency must offer bigger penalty discounts to induce self-reporting (Strong Leniency).

Our experimental analysis suggests that ordered-leniency policies are effective detection mechanisms. First, our findings indicate that the decisions of the majority of injurers in case of multiplicity of equilibria at the self-reporting stage are aligned with the risk-dominance refinement. As a result, the implementation of either Strong or Mild Leniency policies leads to a higher likelihood of self-reporting. Second, as predicted by the theory, we show that ordered-leniency policies enhance the ability of the enforcement agency to detect and punish the injurers. Third, our experiment suggests that ordered-leniency mechanisms elicit early self-reporting and prompt detection of the act. These findings might suggest that ordered leniency generates a race-to-the-courthouse effect. Consequently, ordered-leniency policies have the potential to mitigate the harm inflicted on society.

Fourth, as predicted by the theory, our results indicate that the potential injurer's willingness to participate is significantly higher when the benefit associated with the act is above the deterrence threshold (expected fine), across leniency conditions. Fifth, our findings suggest that the potential injurers are willing to participate in the act more frequently than predicted by the theory under Strong and Mild Leniency. These departures from the theory at the participation stage might originate in the potential injurers' beliefs about securing the first position in the self-reporting queue. As a result, acts are committed more frequently than predicted.

Future experimental work might assess the performance of ordered-leniency policies in environments with asymmetric private benefits from committing the act. In theory, the results derived from the benchmark model are robust to asymmetric benefits as long as the potential injurers can write side contracts with each other and negotiate transfer payments. In experimental settings, however, including a transfer negotiation stage might also affect the potential injurers' willingness to participate and their self-reporting decisions. It might be also relevant to experimentally study environments with more than two potential injurers, and to assess how the group size affects self-reporting, detection and deterrence. In theory, the findings regarding willingness to participate and immediate self-reporting by all injurers are robust to group size. In the lab, however, larger groups may find even harder to coordinate their actions on their preferred outcome i.e., the outcome where neither injurer self-reports. As a result, the detection power of Mild Leniency policies might be stronger than the observed here.[62]

Future theoretical work might assess the social welfare effects of ordered-leniency policies in environments that allow for cognitive biases and repeat offenders. Previous empirical studies on crime and deterrence suggest the presence of judgment errors in the decisions about participating in harmful activities and provide evidence of cognitive biases (Kleck et al., 2005; Kleck and Barnes, 2014; Lochner, 2007; Nagin and Pogarsky, 2003). The findings from this literature also indicate that crime detection experiences attenuate cognitive biases (Pogarsky et al., 2005, 2004). Therefore, ordered-leniency policies might have an additional welfare-enhancing effect: By increasing the likelihood of detection, ordered-leniency policies might reduce the injurers' cognitive biases, i.e., ordered leniency might act as a debiasing mechanism. As a result, higher deterrence of unlawful activities by repeat offenders might be observed. These, and other extensions, may be fruitful topics for future research.

[62]See Landeo and Spier (2020) for theoretical work on these environments and other relevant settings.

References

Abbink, Klaus. 2006. "Laboratory Experiments on Corruption." In *International* , Chapter 14. In Susan Rose-Ackerman, ed., *International Handbook on the Economics of Corruption*. Massachusetts: Edward Elgar Publishing.

Abbink, Klaus and Heike Hennig-Schmidt. 2006. "Neutral versus Loaded Instructions in a Bribery Experiment." *Experimental Economics*, 9: 103–121.

Alekseev, Aleksandr, Gary Charness, and Urin Gneezy. 2017. "Experimental Methods: When and Why Contextual Instructions Are Important." *Journal of Economic Behavior and Organization*, 134: 48–59.

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

Apel, Robert. 2013. "Sanctions, Perceptions, and Crime: Implications for Criminal Deterrence." *Journal of Quantitative Criminology*, 29: 67–101.

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

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

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. "Biased Judgments of Fairness in Bargaining." *American Economic Review*, 11: 109–26.

Barr, Abigail, Magnus Lindelöw, and Pieter Serneels. 2004. "To Serve the Community or Oneself: The Public Servant's Dilemma." Policy Research Working Paper No.3187. World Bank, Washington, D.C.

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

Benoît, Jean-Pierre and Juan Dubra. 2011. "Apparent Overconfidence." *Econometrica*, 79: 1591–1625.

Benoît, Jean-Pierre, Don A. Moore, and Juan Dubra. 2015. "Does The Better-Than-Average Effect Show that People Are Overconfident?: Two Experiments." *Journal of the European Economic Association*, 13: 293–329.

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

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

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

Burton, Anthony and Martin Sefton. 2004. "Risk, Pre-Play Communication and Equilibrium." *Games and Economic Behavior*, 46: 23–40.

Camerer, Colin and Eric J. Johnson. 2004. "Thinking about Attention in Games: Backward and Forward Induction." In Isabelle Brocas and Juan D. Carrillo, eds., *The Psychology of Economic Decisions. Volume Two: Reasons and Choices.* New York: Oxford University Press Inc.

Chalfin, Aaron and Justin McCrary. 2017. "Criminal Deterrence: A Review of the Literature." *Journal of Economic Literature*, 55: 5–48.

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

Cooper, Russell W., Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. 1992. "Communication in Coordination Games." *Quarterly Journal of Economics*, 107: 739–771.

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

Charness, Gary, Guillaume R. Frechette, and Cheng-Zhong Qin. 2007. "Endogenous Transfers in the Prisoner's Dilemma Game: An Experimental Test of Cooperation and Coordination." *Games and Economic Behavior*, 60: 287–306.

Crawford, Vincent P., and Bruno Broseta. 1998. "What Price Coordination? The Efficiency-Enhancing Effect of Auctioning the Right to Play." *American Economic Review*, 88: 198–225.

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

Friedman, Daniel and Shyam Sunder. 1994. *Experimental Methods – A Primer for Economists.* New York: Cambridge University Press.

Greenberg, Michael D. 2011. *For Whom the Whistle Blows. Advancing Corporate Compliance and Integrity Efforts in the Era of Dodd-Frank.* California: The RAND Corporation.

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

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*, 61: 511?27.

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

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

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

Johnson, Eric J., Colin Camerer, Sankar Sen, and Talya Rymon. 2002. "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining." *Journal of Economic Theory*, 104: 16–47.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93: 1449–1475.

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

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

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

Kleck, Gary, and J. C. Barnes. 2014. “Do More Police Lead to More Crime Deterrence?” *Crime and Delinquency*, 60: 716–38.

Kleck, Gary, Brion Sever, Spencer Li, and Marc Gertz. 2005. “The Missing Link in General Deterrence Research.” *Criminology*, 43: 623–60.

Kunda, Ziva. 1987. “Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories.” *Journal of Personality and Social Psychology*, 53: 636–47.

Kunda, Ziva. 1990. “The Case of Motivated Reasoning.” *Psychological Bulletin*, 108: 480–98.

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

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

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

Landeo, Claudia M. and Kathryn E. Spier. 2018. “Optimal Law Enforcement with Ordered Leniency.” Working Paper No. w25095. National Bureau of Economic Research, Cambridge, MA.

Landeo, Claudia M. and Kathryn E. Spier. 2020. “Optimal Law Enforcement with Ordered Leniency.” *Journal of Law and Economics*, 63: 71–111.

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

Larrick, Richard, Katherine Burson, and Jack Soli. 2007. “Social Comparison and Confidence: When Thinking You’re Better than Average Predicts Overconfidence (and when it does not).” *Organizational Behavior and Human Decision Processes*, 102, 76–94.

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

Lochner, Lance. 2007. “Individual Perceptions of the Criminal Justice System.” *American Economic Review*, 97: 444–6 0.

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

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

Martin, Stephen, Hans-Theo Normann and Christopher M Snyder. 2001. “Vertical Foreclosure in Experimental Markets” *RAND Journal of Economics*, 32: 466–496.

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

Nagin, Daniel S. and Greg Pogarsky. 2003. “An Experimental Investigation of Deterrence: Cheating, Self-Serving Bias, and Impulsivity.” *Criminology*, 41: 167–194.

Ochs, Jack. 1995. "Coordination Problems." In J.H. Kagel and A.E. Roth, eds., *Handbook of Experimental Economics*. New Jersey: Princeton University Press Inc.

Pogarsky, Greg. 2009. "Deterrence and Decision Making: Research Questions and Theoretical Refinement." In Krohn, Marvin D., Alan J. Lizzote and Gina Penly Hall, eds., *Handbook on Crime and Deviance*. New York: Springer.

Pogarsky, Greg, Alex R. Piquero, and Ray Paternoster. 2004. "Modeling Change in Perceptions about Sanction Threats: The Neglected Linkage in Deterrence Theory." *Journal of Quantitative Criminology*, 20: 343–69.

Pogarsky, Greg, KiDeuk Kim, and Ray Paternoster. 2005. "Perceptual Change in the National Youth Survey: Lessons for Deterrence Theory and Offender Decision-Making." *Justice Quarterly*, 22: 1–29.

Smith, Angela S. 2011. "An Experimental Study of Exclusive Contracts." *International Journal of Industrial Organization*, 29: 4–13.

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

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

Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science,* 185: 1124–1131.

Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil. 1993. "Asset Markets as an Equilibrium Selection Mechanism: Coordination Failure, Game Form Auctions, and Tacit Communication." *Games and Economic Behavior*, 5: 485–504.

Ordered Leniency: An Experimental Study of Law Enforcement with Self-Reporting
Appendix 1

Claudia M. Landeo* Kathryn E. Spier†

May 6, 2021

This Appendix presents formal proofs of Lemma 1, Proposition 1, and Proposition 2.

Proof of Lemma 1.[1] Denote the strategy of player j as $\sigma_j = (\rho_j, t_j)$ where $\rho_j \in \{R, NR\}$ is whether to report the act and $t_j \in [0, 90]$ is when to report the act. With Strong and Mild Leniency, $r_1 < r_2 = 1$. If $\sigma_{-j} = (NR, t_{-j})$, then player j is indifferent about their reporting time, $(R, 0) \sim (R, t_j)\ \forall t_j \in (0, 90]$. If $\sigma_{-j} = (R, t_{-j})$, then for player j we have $(R, 0) \sim (R, t_j)\ \forall t_j < t_{-j}$ and $(R, 0) \succ (R, t_j)\ \forall t_j \geq t_{-j}$. Therefore $(R, 0)$ weakly dominates $(R, t_j)\ \forall t_j \in (0, 90]$ for Strong and Mild Leniency. With No Leniency, $r_1 = r_2 = 1$ then there is no advantage to being first or second and so the players are indifferent as to the reporting time. ■

Proof of Proposition 1.[2] With Strong Leniency, self-reporting is a dominant strategy and (R, R) is the unique Nash equilibrium (NE).[3] Since injurers self-report at $t = 0$ (Lemma 1), a coin flip determines which injurer is in the first position so the expected fine is $\frac{r_1+r_2}{2} f = 600$. With Mild Leniency, neither player has a dominant strategy and there are two pure-strategy NEs, (NR, NR) and (R,R).[4] With the Pareto-dominance refinement, the injurers coordinate on (NR, NR) because the expected fine is smaller, $360 < 660$. Since the injurers do not self-report, they are detected at $t = 90$ with probability $p_0 = .40$ and the expected fine is 360. With the risk-dominance refinement, the injurers coordinate on (R, R) because R is preferred by player j if player $-j$ is randomizing 50/50 between R and NR or $\frac{1}{2}(420) + \frac{1}{2}(660) < \frac{1}{2}(360) + \frac{1}{2}(810)$, or $540 < 585$. With No Leniency, NR is a dominant strategy and therefore (NR, NR) is the unique NE.[5] Since the injurers do not

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

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

[1]This proof corresponds to Landeo and Spier's (2020) proof of Lemma 1.

[2]This proof corresponds to Landeo and Spier's (2020) proof of Lemma 2, Cases 1, 3, and 4 for Strong Leniency, Mild Leniency, and No Leniency, respectively.

[3]This follows from Table 1 where $b - 300 > b - 360$ and $b - 600 > b - 810$.

[4]This follows from Table 1 where $b - 420 < b - 360$ and $b - 660 > b - 810$.

[5]This follows from Table 1 where $b - 900 < b - 360$ and $b - 900 < b - 810$.

self-report, they are detected at $t = 90$ with probability $p_0 = .40$ and the expected fine is 360. ■

Proof of Proposition 2.[6] Consider Strong Leniency. If both players choose to participate, the act is committed and the expected fine is 600 (see Proposition 1 Case 1). If $b > 600$ then player j strictly prefers to participate if player $-j$ participates, and is indifferent if player j does not participate. Therefore when $b > 600$ participating weakly dominates not participating. If $b < 600$ then player j strictly prefers to not participate if player $-j$ participates, and is indifferent if player j does not participate. Therefore when $b < 600$ not participating weakly dominates participating. When $b = 600$ the player's payoff is zero, and by assumption they do not participate in the act. Therefore $\hat{b} = 600$ is the deterrence threshold. The proofs for Mild Leniency under the Pareto-dominance refinement ($\hat{b} = 360$), Mild Leniency under the risk-dominance refinement ($\hat{b} = 660$), and No Leniency ($\hat{b} = 360$) are similar and not presented. ■

[6] This proof corresponds to Landeo and Spier's (2020) proof of Lemma 3, Cases 1, 3, and 4 for Strong Leniency, Mild Leniency, and No Leniency, respectively.

Appendix 2: Sample Instructions (Mild-Leniency Condition)

PLEASE GIVE THIS MATERIAL TO THE EXPERIMENTER AT THE END OF THE EXPERIMENT

INSTRUCTIONS

This is an experiment in the economics of decision-making. The National Science Foundation has provided the funds for this research.

In this experiment you will be asked to play an economic decision-making computer game. The experiment currency is the "token." The instructions are simple. If you follow them closely and make appropriate decisions, you may make a large amount of money. At the end of the session you will be paid your game earnings in CASH. If you have any questions at any time, please raise your hand and the experimenter will go to your desk.

PROBABILITY OR CHANCE

The concept of probability or chance will be used in this experiment. **PROBABILITY OR CHANCE (EXPRESSED IN PERCENTAGES)** indicates the likelihood of occurrence of uncertain events. The concept of probability or chance can be illustrated as follows. Suppose that an urn contains 100 balls: 20 out of 100 balls are white, and 80 out of 100 balls are black. Suppose that you randomly extract one ball from the urn. The chance that the ball will be white is equal to 20% because 20 out of 100 balls in the urn are white. Similarly, the chance that the ball will be black is equal to 80%, because 80 out of 100 balls in the urn are black.

SESSION AND PLAYERS

The session is made up of 6 matches. The first 5 matches are practice matches. After the last practice match, **ONE ACTUAL MATCH** will be played.

1) At the beginning of the session, every participant will be randomly assigned a role. The equally likely roles are: **Player B1** and **Player B2**.

   The **ROLES WILL REMAIN THE SAME until the end of the session**.

2) Before the beginning of **EACH PRACTICE MATCH**, the computer will randomly form pairs of **TWO PEOPLE**: **Player B1** and **Player B2**.

   YOU WILL NOT KNOW THE IDENTITY OF YOUR PARTNER.
   YOU WILL PLAY WITH A DIFFERENT PARTNER IN EVERY PRACTICE MATCH.

3) Before the beginning of the **ACTUAL MATCH**, the computer will randomly form pairs of **TWO PEOPLE**: **Player B1** and **Player B2**.

   YOU WILL NOT KNOW THE IDENTITY OF YOUR PARTNER.
   YOU WILL NOT PLAY WITH ANY OF YOUR PREVIOUS PARTNERS.

MATCH STAGES

STAGE 1: DECISION WHETHER TO JOINTLY COMMIT THE ACT

1) Each player has an **initial endowment** equal to **700 tokens**.

2) **THE DECISION TO JOINTLY COMMIT THE ACT** refers to an economic decision involving potential economic benefits and potential economic losses.

   - **ECONOMIC BENEFITS** might occur in **STAGE 1**.
   - **ECONOMIC LOSSES** might occur in **STAGE 2**.

3) **THE COMPUTER** randomly determines the **NUMBER OF TOKENS X** that each player will get **IF BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT.** Both players will receive the same number of tokens X.

   - The number of tokens X can be equal to **200, …, 1598, 1599, 1600 tokens**.
   - The number of tokens X will be revealed to both players.

4) **Player B1** and **Player B2** decide whether to **JOINTLY COMMIT THE ACT**

- Each player will have **90 SECONDS** **TO DECIDE WHETHER TO AGREE TO JOINTLY COMMIT THE ACT OR NOT AGREE TO JOINTLY COMMIT THE ACT AND PRESS THE NEXT BUTTON.**
    - If a player **FAILS TO MAKE A CHOICE AND TO PRESS THE NEXT BUTTON WITHIN THE 90-SECOND PERIOD,** it will be implied that he/she decided **NOT TO AGREE TO JOINTLY COMMIT** the act.

5) The possible outcomes are as follows.

- **BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT:** Each player gets **X TOKENS** (in addition to the initial endowment of 700 tokens) and **STAGE 2 BEGINS**.

- **ONLY ONE PLAYER AGREES TO JOINTLY COMMIT THE ACT:** Each player gets **ZERO TOKENS** and the **MATCH ENDS**. The match payoff for each player will be 700 tokens (initial endowment).

- **NEITHER PLAYER AGREES TO JOINTLY COMMIT THE ACT:** Each player gets **ZERO TOKENS** and the **MATCH ENDS**. The match payoff for each player will be 700 tokens (initial endowment).

STAGE 2: DECISION WHETHER TO REPORT THE ACT

1) If **both payers agreed to jointly commit the act**, then Stage 2 begins.

2) **A FINE EQUAL TO 900 TOKENS MIGHT BE DEDUCTED FROM A PLAYER'S TOKEN BALANCE AS A CONSEQUENCE OF JOINTLY COMMITTING THE ACT.**

    - A player's decision to report the act **MIGHT DECREASE THE FINE HE/SHE WILL PAY** from 900 tokens to 420 tokens.

    - A player's decision to report the act **MIGHT INCREASE HIS/HER PARTNER'S CHANCE TO PAY A FINE** from 40% to 90% or from 40% to 100%.

    THE SPECIFIC FINE AND THE CHANCE OF PAYING THAT FINE DEPEND ON THE DECISIONS OF BOTH PLAYERS ABOUT REPORTING THE ACT.

3) Each player will have **90 SECONDS TO DECIDE WHETHER TO REPORT OR NOT TO REPORT THE ACT AND PRESS THE NEXT BUTTON.**

    - If a player **FAILS TO MAKE A CHOICE AND TO PRESS THE NEXT BUTTON WITHIN THE 90-SECOND PERIOD,** it will be implied that he/she decided **NOT REPORT** the act.

4) The possible outcomes and match payoffs are presented below.

POSSIBLE OUTCOME 1: BOTH PLAYERS DECIDE NOT TO REPORT THE ACT

- **<u>NEITHER PLAYER GETS A FINE REDUCTION</u>**
- **<u>EACH PLAYER'S CHANCE OF PAYING A FINE EQUAL TO 900 TOKENS IS 40%</u>**: Each player pays a fine equal to **900 tokens** with **40% chance** and **does not pay any fine** with **60% chance**.

Hence, the match payoffs are as follows.

With a **40% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 900 tokens Player B2's match payoff = 700 tokens + X tokens – 900 tokens

With a **60% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 0 tokens Player B2's match payoff = 700 tokens + X tokens – 0 tokens

POSSIBLE OUTCOME 2: ONLY PLAYER B1 DECIDES TO REPORT THE ACT

- **ONLY PLAYER B1 GETS A FINE REDUCTION**: Instead of paying a fine equal to 900 tokens, Player B1 **always** pays only **420 tokens**.
- **PLAYER B2'S CHANCE OF PAYING A FINE EQUAL TO 900 TOKENS IS 90%**: Player B2 pays a fine equal to **900 tokens** with **90% chance** and **does not pay any fine** with **10% chance**.

Hence, the match payoffs are as follows.

With a **90% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 420 tokens
Player B2's match payoff = 700 tokens + X tokens – 900 tokens

With a **10% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 420 tokens
Player B2's match payoff = 700 tokens + X tokens – 0 tokens

POSSIBLE OUTCOME 3: ONLY PLAYER B2 DECIDES TO REPORT THE ACT

- **ONLY PLAYER B2 WILL GET A FINE REDUCTION**: Instead of paying a fine equal to 900 tokens, Player B2 **always** pays only **420 tokens**.
- **PLAYER B1'S CHANCE OF PAYING A FINE EQUAL TO 900 TOKENS IS 90%**: Player B1 pays a fine equal to **900 tokens** with **90% chance** and **does not pay any fine** with **10% chance**.

Hence, the match payoffs are as follows.

With a **90% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 900 tokens Player B2's match payoff = 700 tokens + X tokens – 420 tokens

With a **10% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 0 tokens Player B2's match payoff = 700 tokens + X tokens – 420 tokens

POSSIBLE OUTCOME 4: BOTH PLAYERS DECIDE TO REPORT THE ACT

- **IF PLAYER B1 REPORTS FIRST**
  - **<u>ONLY PLAYER B1 GETS A FINE REDUCTION</u>**: Instead of paying a fine equal to 900 tokens, Player B1 **always** pays only **420 tokens**.
  - **<u>PLAYER B2'S CHANCE OF PAYING A FINE EQUAL TO 900 TOKENS IS 100%</u>**: Player B2 **always** pays a fine equal to **900 tokens.**

Hence, the match payoffs are as follows.

With a **100% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 420 tokens Player B2's match payoff = 700 tokens + X tokens – 900 tokens

- **IF PLAYER B2 REPORTS FIRST**
  - **<u>ONLY PLAYER B2 GETS A FINE REDUCTION</u>**: Instead of paying a fine equal to 900 tokens, Player B2 **always** pays only **420 tokens**.
  - **<u>PLAYER B1'S CHANCE OF PAYING A FINE EQUAL TO 900 TOKENS IS 100%</u>**: Player B1 **always** pays a fine equal to **900 tokens**.

Hence, the match payoffs are as follows.

With a **100% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 900 tokens Player B2's match payoff = 700 tokens + X tokens – 420 tokens

- **IF BOTH PLAYERS REPORT AT THE SAME TIME**
  - **<u>EACH PLAYER GETS A FINE REDUCTION WITH 50% CHANCE</u>**: Instead of paying a fine equal to 900 tokens, each player pays only **420 tokens** with **50% chance**.
  - **<u>EACH PLAYER'S CHANCE OF PAYING A FINE EQUAL TO 900 TOKENS IS 50%:</u>** Each player pays a fine equal to **900 tokens** with **50% chance**.

Hence, the match payoffs are as follows.

With a **50% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 420 tokens
Player B2's match payoff = 700 tokens + X tokens – 900 tokens

With a **50% CHANCE,** the match payoffs will be:

Player B1's match payoff = 700 tokens + X tokens – 900 tokens
Player B2's match payoff = 700 tokens + X tokens – 420 tokens

EXERCISES

Suppose that the number of tokens that each player gets **IF BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT** is equal **to X tokens.**

Nine exercises, based on the possible outcomes, are presented below. Please fill the blanks.

Exercise 1

Suppose that **BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT**. Then, each player gets ____ tokens. Suppose also that **Player B1** decides **NOT TO REPORT** the act and **Player B2** decides **NOT TO REPORT** the act.

The **MATCH PAYOFFS (IN TOKENS)** are as follows.

Player B1:

Chance	Payoff	Chance	Payoff

Player B2:

Chance	Payoff	Chance	Payoff

Exercise 2

Suppose that **BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT**. Then, each player gets ____ tokens. Suppose also that **Player B1** decides **TO REPORT** the act and **Player B2** decides **NOT TO REPORT** the act.

The **<u>MATCH PAYOFFS</u> (IN TOKENS)** are as follows.

Player B1:

Chance	**Payoff**	**Chance**	**Payoff**

Player B2:

Chance	**Payoff**	**Chance**	**Payoff**

Exercise 3

Suppose that **BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT**. Then, each player gets ____ tokens. Suppose also that **Player B1** decides **NOT TO REPORT** the act and **Player B2** decides **TO REPORT** the act.

The **<u>MATCH PAYOFFS</u> (IN TOKENS)** are as follows.

Player B1:

Chance	**Payoff**	**Chance**	**Payoff**

Player B2:

Chance	**Payoff**	**Chance**	**Payoff**

Exercise 4

Suppose that **BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT**. Then, each player gets ____ tokens. Suppose also that **Player B1** decides **TO REPORT** the act and **Player B2** decides **TO REPORT** the act, and that **Player B1 IS THE FIRST TO REPORT**.

The **MATCH PAYOFFS (IN TOKENS)** are as follows.

Player B1:

Chance	Payoff	Chance	Payoff

Player B2:

Chance	Payoff	Chance	Payoff

Exercise 5

Suppose that **BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT**. Then, each player gets ____ tokens. Suppose also that **Player B1** decides **TO REPORT** the act and **Player B2** decides **TO REPORT** the act, and that **Player B2 IS THE FIRST TO REPORT**.

The **MATCH PAYOFFS (IN TOKENS)** are as follows.

Player B1:

Chance	Payoff	Chance	Payoff

Player B2:

Chance	Payoff	Chance	Payoff

Exercise 6

Suppose that **BOTH PLAYERS AGREE TO JOINTLY COMMIT THE ACT**. Then, each player gets ____ tokens. Suppose also that **Player B1** decides **TO REPORT** the act and **Player B2** decides **TO REPORT** the act, and that **BOTH PLAYERS REPORT AT THE SAME TIME**.

The **MATCH PAYOFFS (IN TOKENS)** are as follows.

Player B1:

Chance	**Payoff**	**Chance**	**Payoff**

Player B2:

Chance	**Payoff**	**Chance**	**Payoff**

Exercise 7

Suppose that **Player B1 AGREES TO JOINTLY COMMIT THE ACT** and **Player B2 DOES NOT AGREE TO JOINTLY COMMIT THE ACT**.

The **MATCH PAYOFFS (IN TOKENS)** are as follows.

Player B1:

Chance	**Payoff**	**Chance**	**Payoff**

Player B2:

Chance	**Payoff**	**Chance**	**Payoff**

Exercise 8

Suppose that **Player B1 DOES NOT AGREE TO JOINTLY COMMIT THE ACT** and **Player B2 AGREES TO JOINTLY COMMIT THE ACT**.

The **<u>MATCH PAYOFFS</u> (IN TOKENS)** are as follows.

Player B1:

Chance	**Payoff**	**Chance**	**Payoff**

Player B2:

Chance	**Payoff**	**Chance**	**Payoff**

Exercise 9

Suppose that **NEITHER PLAYER AGREES TO JOINTLY COMMIT THE ACT**.

The **<u>MATCH PAYOFFS</u> (IN TOKENS)** are as follows.

Player B1:

Chance	**Payoff**	**Chance**	**Payoff**

Player B2:

Chance	**Payoff**	**Chance**	**Payoff**

SESSION PAYOFF

The game earnings in tokens will be equal to the **PAYOFF FOR THE ACTUAL MATCH**. The game earnings in dollars will be equal to: (game earnings in tokens)/29 (29 tokens = 1 dollar). The session payoff will be equal to the game earnings in dollars plus the $10 participation fee.

GAME SOFTWARE

The game will be played using a computer terminal. You will need to enter your decisions by using the mouse. In some instances, you will need to wait until the other players make their decisions before moving to the next screen. Please **BE PATIENT**. There will be a box, displayed in the upper right-hand side of your screen, which indicates the "Match Number," "Your Role," and "Your Balance."

Please press the **NEXT >>** button to move to the next screen. **DO NOT TRY TO GO BACK TO THE PREVIOUS SCREEN AND DO NOT CLOSE THE BROWSER**: The software will stop working.

Next, the **5 PRACTICE MATCHES** will begin. After that, the **ACTUAL MATCH** will be played. **YOU CAN CONSULT THESE INSTRUCTIONS AT ANY TIME DURING THE EXPERIMENT**.

THANKS FOR YOUR PARTICIPATION IN THIS STUDY!!

PLEASE GIVE THIS MATERIAL TO THE EXPERIMENTER AT THE END OF THE EXPERIMENT