Crime Entanglement, Deterrence, and Witness Credibility*

Harry Pei Bruno Strulovici

July 14, 2019

Abstract: When a defendant is accused of multiple crimes, one may consider punishing him if the probability that he has committed *at least one* of these crimes is high. We show that entangling criminal accusations in this way can severely harm deterrence and reduce witnesses' credibility. When conviction entails a large punishment, an individual's decisions to commit distinct crimes are substitutes and witnesses' reports are complements. This tension induces negatively correlated private information and a coordination motive among witnesses, which leads to uninformative reports, ineffective deterrence, and frequent crimes in equilibrium. We discuss various remedies to restore credibility and reduce crime.

Keywords: soft evidence, deterrence, strategic restraint, coordination, information linkage.
JEL Codes: D82, D83, K42.

1 Introduction

Abuses of power, extortions, assaults, and other crimes are often hard to prove with incontrovertible evidence. To address this difficulty and to mitigate the risk that some accusations are driven by spite, grudges, or ulterior motives, the number of accusations leveled against an individual is sometimes used to assess this individual's likelihood of guilt. The presumption is that the accumulation of accusations against an individual makes it more likely that he is guilty of at least some of the crimes.

This approach supposes that convictions and punishments can be decided based on an individual's overall probability of guilt rather than on the probabilities with which he has committed each specific crime. This feature, which we call *crime entanglement*, has some intuitive appeal that we illustrate with the following example. Consider two defendants, each of whom is suspected of having abused some ex ante identical victims with the following probabilities:

*Preliminary versions of this project have been circulated under various titles. We thank S. Nageeb Ali, Bocar Ba, Sandeep Baliga, Arjada Bardhi, Mehmet Ekmekci, Alex Frankel, Chishio Furukawa, George Georgiadis, Bob Gibbons, Andrei Gomberg, Yingni Guo, Anton Kolotinin, Frances Xu Lee, Annie Liang, Matt Notowidigdo, Wojciech Olszewski, Alessandro Pavan, Joyce Sadka, Larry Samuelson, Ron Siegel, Vasiliki Skreta, Juuso Toikka, Rakesh Vohra, Alex White, Alex Wolitzky, Boli Xu, and our seminar audiences for helpful comments. Strulovici acknowledges financial support from the National Science Foundation (NSF Grant No.1151410). Pei: Department of Economics, Northwestern University. harrydp@northwestern.edu. Strulovici: Department of Economics, Northwestern University. b-strulovici@northwestern.edu

	Prob of abusing victim 1	Prob of abusing victim 2	Prob of abusing at least one victim
Defendant 1	49 %	49 %	98 %

	Prob of abusing victim 3	Prob of abusing victim 4	Prob of abusing at least one victim
Defendant 2	60 %	1 %	61 %

The first defendant is almost surely guilty and seems more deserving of punishment than the second defendant. From this perspective, crime entanglement appears to be desirable and may affect how individuals accused of multiple crimes are treated in business decisions (such as the termination of employment or advertising contracts) and in the court of public opinion.

However, crime entanglement is not a feature of criminal justice systems. These systems treat each crime as a separate count and guilt is decided on each count. In the example, the second defendant, who is more likely of being guilty of a specific crime (abusing victim 3) is more likely to be convicted than the first defendant, whose probability of abusing any specific victim is lower. This feature of criminal justice systems may seem counterintuitive and perhaps suboptimal from a deterrence and welfare perspective, especially considering the fact crimes are defined as acts *against society*, whose harm is independent of the victim's identity.

This paper studies the desirability of crime entanglement when the incentives to commit and report crimes are endogenous. We analyze a game between a judge, a potential criminal (*principal*) and multiple potential victims or witnesses (*agents*) whose reports are influenced by three considerations: (1) a preference for punishing criminal behavior, (2) a risk of retaliation if accusations fail to get the principal convicted, and (3) some (possibly small) idiosyncratic private benefits or costs of getting the principal convicted, which are independent of whether crimes have taken place. The crimes studied here include abuses of power, extortions, workplace bullying, and various forms of harassment.[1]

Our focus is on a potential criminal's incentive to commit crimes and the informativeness of witnesses' reports—accusations, or the *absence thereof*[2]—when conviction is based on the probability

[1]In our model, (1) some abuses go unreported and (2) some charges of abuse are not deemed credible enough to lead to a conviction. Both features are consistent with the empirical evidence on abuses and reports of abuse. For instance, a 2016 survey conducted by the USMSPB concluded that 21% of women and 8.7% of men experienced at least one of 12 categorized behaviors of sexual harassment, of which only a small fraction was followed by charges. According to data released by USMSPB (2018), of the harassment charges filed in 2017, only 16% led to "*merit resolutions*," i.e., to outcomes favorable to the charging parties. A similar pattern was found in a study of harassment in the U.S. military by the RAND corporation (2018) and studies of police brutality or inaction by Ba (2018) and Ba and Rivera (2019) using data from the city of Chicago.

[2]Empirically, the main problem concerning assaults is not false accusations but the underreporting of crime. For example, 58% of violent crimes and 77% of sexual assaults were unreported to the police according to the US Bureau

that a defendant committed *at least one* crime. We show, perhaps paradoxically, that when conviction entails a large punishment for the principal, crime entanglement destroys the informativeness of reports, and leads to a high equilibrium probability of crime.[3] By contrast, we show that treating each crime separately induces highly informative reports and a low probability of crime. Applied to abuses of power, this implies that treating each alleged victim's report as a stand-alone case, investigated independently of other accusations, leads to better outcomes than does using the sum of accusations against a defendant to determine his punishment. Moreover, we show that criminal accusations can be disentangled without increasing the maximal punishment needed to deter crime.

In our baseline model, the principal has several opportunities to commit a crime, each of which is associated with a distinct agent who privately observes whether the corresponding crime takes place. In some applications, the principal is an employer who has multiple opportunities of violating the law and agents are employees who may witness violations and become whistleblowers. In other applications, such as abuses of power, extortions and harassments, agents are the direct victims of the principal's acts. Crime entanglement means that the principal is convicted if the probability that he has committed some crime, given all agents' reports, exceeds some exogenous conviction threshold.

We show in Theorem 1 that when the punishment faced by a defendant is large, a rational principal strategically commits few crimes to reduce the expected number of accusations filed against him and reduce the odds of being punished. This strategic restraint undermines the informativeness of agents' reports and, paradoxically, leads to a high probability of crime.

To see this, let us first consider the case of two agents. We show that when punishment in case of conviction is large, the principal is convicted in equilibrium only if two reports are made against him.[4] When the conviction probabilities have this property, the principal's decisions to abuse each agent are *strategic substitutes* because he is unpunished if one accusation is filed against him, but is punished when he faces two accusations. In equilibrium, the principal either commits no crime or only one crime. The agents' decisions to report crimes are *strategic complements* because one

of Justice Statistics (2016), whereas roughly 2% of criminal complaints and 8% of forcible rape complaints were "unfounded" (FBI CIUS report, 1996, Section II, p. 24). This evidence is consistent with our model, which predicts that victims underreport crime in equilibrium.

[3]Our results are derived from an equilibrium analysis and are thus ill-suited to describe behavior during unexpected legal and social changes, in which people entertain wrong beliefs about the consequences of their actions. See Section 6.

[4]Intuitively, suppose that one report were enough to convict the principal. Then, two reports would get him convicted for sure. As the punishment in case of conviction increases, the principal will have a strict incentive not to commit any crime, which is incompatible with any punishment occurring in equilibrium given that the judge is Bayesian. More precisely, we show that crime must occur with strictly positive probability in every equilibrium that survives a presumption of innocence refinement (Lemma 3.1).

report is insufficient to have the principal convicted and the reporting agent incurs a retaliation cost whenever the principal is acquitted. Given the principal's equilibrium strategy, an agent who has been abused knows that the other agent is unlikely to accuse the principal and, hence, that his report is unlikely to result in a conviction. This weakens the abused agent's incentive to accuse the principal. A similar argument applies in the reverse direction: an agent who has not been abused assigns a higher probability to the other agent being abused, which encourages him to accuse the principal.

Conceptually, the equilibrium behavior of the principal induces *negative correlation* in agents' private observations of crime, and the judge's conviction rule induces *complementarity* of agents' reports. Although an abused agent is strictly more likely to accuse the principal than an agent who has not been abused, the negative correlation and report complementarity *reduce the difference* between the reporting probabilities of abused and unabused agents, undermining the informativeness of agents' reports. This lack of credibility has a further perverse effect: it increases the equilibrium probability of crime. We show in Theorem 2 that when the punishment in case of conviction becomes arbitrarily large relative to the benefit from committing crime, the agents' reports become *arbitrarily uninformative* and the probability that the principal commits crime converges to the exogenous conviction threshold. This finding stands in sharp contrast with the single-agent benchmark, in which the potential victim's report becomes arbitrarily informative and the probability of crime vanishes to zero as the punishment to the convicted criminal becomes arbitrarily large (Proposition 1).

This adverse logic of crime entanglement has another implication when there are more agents: as the punishment to a convicted principal becomes large enough, the informativeness of individual reports becomes so weak that conviction occurs only when *all* agents accuse the principal (Proposition 6). To see this, suppose that one report of abuse were enough to convict the principal in equilibrium. This would expose the principal to a high probability of wrongful conviction, reflecting the high probability that at least one agent falsely accuses him, and violate the conviction threshold that must be met in equilibrium. Thus, more accusations must be filed in order to convict the principal. However, this requirement introduces a coordination motive among agents which, combined with the principal's strategic restraint, reduces the informativeness of individual reports. As a result, even more reports are needed, other things being equal, to reach the conviction threshold. In turn, requiring more reports for conviction renders the coordination problem even more severe across agents and reduces the informativeness of their individual reports still further, and so on. Pushes to its extreme, crime entanglement implies that the principal is convicted only when all agents report against him.

Following the same reasoning, we show in Proposition 7 that as the number of agents increases, reports' informativeness decreases even when all the reports are aggregated. As a result, the equilibrium probability of crime increases. Intuitively, a larger number of potential victims magnifies the effects of negative correlation and coordination motives. Furthermore, the probability that any given agent accuses the principal *increases* with the number of agents. This distinguishes our results from theories of public good provision, in which contributions become scarcer as the number of agents increases.

In order to describe the forces driving our results in their simplest form, we omit in our baseline model any heterogeneity (or uncertainty) in the principal's benefits from committing crimes. In practice, one rationale for aggregating reports about a given defendant is to ascertain the defendant's propensity to commit crime. To account for this heterogeneity, we generalize our analysis by allowing the principal to be either a *virtuous* type, with zero benefit from committing crime; or an *opportunistic* one with the same payoff function as in our baseline model. The principal's type is privately observed.

Focusing the heterogeneous-principal analysis on the case of two agents, we show that when the punishment to a convicted principal is sufficiently large, two reports are required to convict the principal. When the prior probability that the principal is virtuous exceeds a cutoff,[5] the opportunistic type commits at least one crime and commits crime against both agents with positive probability. However, the agents' private observations of crime remain negatively correlated: a victim assigns a lower probability to the other agent being abused, compared to the belief of a non-victim. Therefore, the strategic substitutability in the incentives to commit crimes, the endogenous negative correlation across the agents' private observations, and the agents' coordination motives in filing reports still arise in this setting. Also robust is our result that convicting the principal based on the overall probability of guilt lowers the informativeness of reports and leads to a high probability of crime.

We explore several remedies to restore report the informativeness and reduce crime. First, we consider an alternative conviction rule: after receiving agents' reports, the judge computes the probability that the principal has abused *each agent*, and convicts the principal if the *maximum of these probabilities* across agents exceeds an exogenous threshold. Proposition 3 shows that in equilibrium, the probability of conviction is linear in the number of accusations, agents' private signals are uncorrelated, agents do not benefit from coordinating their reports, and the probability of crime vanishes as the punishment

[5]This cutoff probability equals one minus the conviction threshold. When the prior probability of the virtuous principal falls below this cutoff, the equilibrium probability of crime and the informativeness of the agents' reports coincides with the baseline model.

in case of conviction becomes large.

Disentangling crimes against different individuals is consistent with existing procedures in the criminal justice system, in which crimes against different victims (or different crimes) are examined separately. Our analysis provides a rationale for using the *probability of each individual crime* when making conviction decisions, rather than the *overall probability of guilt*. Applied to the context of workplace bullying and sexual assault, it suggests that the accusation of every potential victim should be treated independently of the probabilities with which the accused individual abuses other potential victims. This implication echoes some critiques of the procedures that link accusations across potential victims.[6] However, this new conviction rule is unappealing from an ex post perspective. As illustrated by our earlier numerical example, the rule can result in convicting defendants whose likelihood of guilt is significantly lower than some defendants who are acquitted. As a result, its implementation requires commitment power from the individuals that make conviction decisions, such as judges and firm executives, and raises serious ethical concerns.

When crime entanglement is unavoidable, a second way of reducing crime is, paradoxically, to avoid extreme punishments.[7] For intermediate punishment levels, we show that the equilibrium conviction probabilities is concave in the number of accusations and a single report suffices to convict the principal with positive probability (Proposition 4). The principal's decisions to abuse different agents are now *strategic complements*, which induces positive correlation in agents' private signals and leads to *more* informative reports and a lower probability of crime. This remedy comes at the cost of increasing the number of victims conditional on abuse taking place. Our analysis thus unveils a tradeoff between reducing the *probability of crime* and reducing the *severity of crime*.

Finally, we consider the use of monetary transfers. We show that by compensating an agent who stands alone in accusing the principal, one can offset the coordination motives among agents.[8]

The insights from our baseline model can be extended along several directions. The principal could face a larger punishment if the probability that he has committed multiple crimes is high

[6] Keith Hiatt, the director of the technology program at the Human Rights Center at UC Berkeley School of Law notes, concerning the multiple-accusations approach taken by the online platform Callisto that "it may also codify an entrenched attitude that women need to have corroborating evidence to be believed." *New York Times, "The War on Campus Sexual Assault Goes Digital," Nov. 13, 2015.*

[7] This remedy may be best suited for minor crimes and stands in contrast to earlier works, such as Becker (1968), who finds that maximal punishment is optimal regardless of the crime.

[8] As the punishment to a convicted principal becomes arbitrarily large, the agents' reports become arbitrarily informative under such a transfer scheme (Proposition 5). However, this transfer scheme is not budget-balanced: it is costly to the social planner and creates incentives for the principal and the agents to collude, and we show that budget-balanced transfers fail to fully suppress the harmful channel uncovered in the analysis of our benchmark model.

enough, or he could have decreasing marginal benefits from committing multiple crimes. These features strengthen the principal's incentive for strategic restraint and, hence, strengthens the negative correlation in agents' private signals. In other applications, the retaliation cost suffered by each reporting agent may decrease in the number of reporting agents. This feature strengthens agents' coordination motives, which exacerbates the effects described in the baseline model.

Our results also extend when false accusations can be exposed ex post with positive probability. Punishing false accusers is formally equivalent to increasing agents' propensity to punish criminal behavior, and our results require no assumption on the latter's magnitude. The results also hold when agents directly care about crimes committed against (or observed by) other agents. The endogenous negative correlation emphasized earlier has a more pronounced effect because each agent's payoff now directly depends on other agents' private signals. When agents' signals are negatively correlated, each agent's reporting decision attaches more weight to his belief about other agents' signals, and therefore, becomes less responsive to his own private signal. Alternatively, agents may have a preference for reporting the truth, independent of the principal's punishment. We provide an upper bound on the reports' informativeness and a lower bound on the probability of crime when the direct benefit from truth-telling is small relative to the loss from retaliation.

Finally, the principal may hold private information concerning the number of potential victims. In this case, we show that the principal's incentive to commit crime is stronger when this number is smaller. Intuitively, the expected number of reports is smaller when there are fewer potential victims, other things being equal. The negative correlation emphasized throughout this paper arises in this setting as well, because an agent who has been abused infers that the set of potential victims is small and expects few reports to be filed by other agents.

Related Literature: This paper contributes to literatures on information aggregation, games of coordination, law and economics, and the elicitation of correlated private information.

First, we identify a novel channel to explain the failure of information aggregation, which is different from the channel behind the social learning models of Scharfstein and Stein (1990), Banerjee (1992), Ottaviani and Sørensen (2000), and Smith and Sørensen (2000). In those models, agents fail to act on their private information because they can observe informative actions taken by their predecessors. By contrast, in our model, agents cannot observe one another's actions, and the failure of information aggregation is driven by the negative correlation between the agents' private signals and their incentives to coordinate their reports.[9]

In models of collective decision making, information aggregation may fail due to individual biases (Morgan and Stocken 2008) or voters' use of pivotal reasoning (Austen-Smith and Banks 1996, Bhattacharya 2013). When voters' *payoffs* from a reform are negatively correlated, Schmitz and Tröger (2012) show that majority rule is dominated by other voting rules, and Ali, Mihm, and Siga (2018) show that supermajority rules lead to inefficient collective decisions.[10] In our paper, the negative correlation concerns agents' private information, rather than their payoffs, and moreover, both the *voting rule* and the *correlation structure* are endogenous.[11]

Second, our paper is related to the literature on strategic information transmission with multiple senders, which includes Battaglini (2002, 2017), Ambrus and Takahashi (2008), and Ekmekci and Lauermann (2019). In contrast to these works, senders in our model communicate information about the principal's action, and the senders' private signals are endogenously correlated. We examine how this endogenous correlation structure, combined with the senders' endogenous coordination motive, affects communication informativeness, and propose ways to improve informativeness by mitigating senders' endogenous negative correlation and coordination motive.

Third, our paper contributes to law and economics by (1) studying conviction rules that entangle multiple crimes, (2) endogenizing the informativeness and credibility of witness testimonies, and (3) analyzing the interplay between an individual's incentive to commit crimes and witnesses' incentives to tell the truth.[12] Our results justify a key feature of criminal justice systems, which is to treat distinct accusations separately in conviction decisions. Our finding that a lower punishment can reduce the probability of crime stands in contrast to Becker's (1968) well-known observation that maximal punishments save on law-enforcement costs.[13] Our results highlight the challenges of designing

[9] Strulovici (2018) studies a sequential learning model in which an agent is less likely to have an informative signal, other things being equal, if another agent has found such a signal. This *information attrition* may be viewed as a form of negative correlation across agents' signals, which also has adverse effects on learning.

[10] Intuitively, if a voter is pivotal, then a sizable share of other voters is in favor of the reform. This, together with the assumed negative correlation of payoffs across voters, implies that the pivotal voter is unlikely to benefit from the reform.

[11] This contrasts to previous models of strategic voting, both with and without information acquisition (e.g., Feddersen and Pesendorfer 1998, Persico 2004), in which these elements are typically exogenous.

[12] In settings that contrast with ours, Silva (2018) and Baliga, Bueno de Mesquita and Wolitzky (2019) consider models with multiple potential suspects and at most one crime. Silva (2018) constructs a scheme that elicits truthful confessions among suspects. In Baliga, et al. (2019), one of the potential assailants has an opportunity to commit crime. In both papers, the negative correlation in suspects' types is exogenous.

[13] Stigler (1970) observes that several punishment levels should optimally be used when criminals can choose between different levels of crime. The rationale is to provide marginal incentives not to commit the worst crimes. In this scenario, applying the maximal punishment to the worst crimes remains optimal from the perspective of deterring crimes. Siegel and Strulovici (2018), find that extreme punishments are optimal among judicial mechanisms in which the defendant has

punishment when crimes are nonverifiable and unveil a trade-off between reducing the probability of crime and reducing the severity of crime.

Lee and Suen (2019) study the timing of reports by victims and false accusers in a model in which a criminal commits a crime against each of the two agents with some exogenous probability, and provide an explanation for the well-documented fact that victims sometimes delay their accusations. Their analysis and ours consider complementary aspects of the victims' reporting incentives. In particular, the principal's strategic restraint that emerges endogenously in our model and the negative correlation it induces on the agents' private information are distinctive features of our analysis.

The rest of the paper is organized as follows. We introduce our baseline model in section 2. We study the effects of crime entanglement on deterrence and reporting credibility in section 3. Section 4 proposes solutions to restore informativeness and reduce crime, which includes treating crimes against different agents separately. Section 5 studies extensions of the baseline model. Section 6 concludes and puts our results in a broader perspective.

2 Model

We consider a three-stage game between a principal (e.g., the manager of a firm, a police officer), n agents (e.g., the principal's subordinates, citizens) and an evaluator (e.g., a judge, a board of trustees, the supervisor of the police officer). In stage 1, the principal chooses an n-dimensional vector $\boldsymbol{\theta} \equiv (\theta_1, ..., \theta_n) \in \{0,1\}^n$, where $\theta_i = 1$ (0) means that the principal commits a crime (does not commit any crime) against agent i (e.g., extortion, bullying).

In stage 2, each agent $i \in \{1, 2, ..., n\}$ privately observes θ_i and the realization of a private shock $\omega_i \in \mathbb{R}$, which represent agent i's private benefit in case the principal is acquitted. The agents then simultaneously decide whether to file a report against the principal. We let $a_i \in \{0,1\}$ denote agent i's decision: $a_i = 1$ if he accuses the principal and $a_i = 0$ if he does not. We assume that $\omega_1, \omega_2, \ldots, \omega_n$ are independently and normally distributed with mean μ and variance σ^2, and denote their cdf and pdf by $\Phi(\cdot)$ and $\phi(\cdot)$, respectively.[14] For technical purposes, we assume that each agent is strategic with

a binary type.

[14]The assumption that the support of ω_i's distribution is unbounded from below guarantees that, as long as conviction decisions are responsive to agents' reports, a victim reports with positive probability regardless of the odds of retaliation. In Online Appendix I, we show that the results extend to a large class of non-Gaussian distributions whose support is unbounded from below. When the support of ω is bounded from below and the punishment to the convicted principal is large, we list several undesirable and unrealistic features arising in equilibria of the game (Online Appendix I).

probability $\delta \in (0,1)$ and is mechanical otherwise.[15] A strategic agent chooses whether to report or not in order to maximize his payoff. A mechanical agent reports with some exogenous probability $\alpha \in (0,1)$. Whether an agent is strategic or mechanical is independent of $\{\omega_1, ..., \omega_n\}$ and of whether other agents are strategic or mechanical. Our primary focus is on cases in which δ is close to 1.

In stage 3, the evaluator observes the vector of reports $\boldsymbol{a} \equiv (a_1, ..., a_n) \in \{0,1\}^n$, and decides whether to *convict* or *acquit* the principal. Formally, his decision is denoted by $s \in \{0,1\}$, with $s = 1$ if he convicts the principal and $s = 0$ if he acquits the principal.

Payoffs: Let $\overline{\theta} \equiv \max_{i \in \{1,2,...,n\}} \theta_i$. By definition, the principal is innocent when $\overline{\theta} = 0$, and is guilty (of at least some crime) when $\overline{\theta} = 1$. The evaluator has a quadratic payoff function:

$$-\Big(s + (\pi^* - \frac{1}{2}) - \overline{\theta}\Big)^2, \tag{2.1}$$

where $\pi^* \in (0,1)$ is an exogenous parameter, called the *conviction threshold*. Therefore, the evaluator's optimal decision is:

$$s \begin{cases} = 1 & \text{if} \quad \Pr\left(\overline{\theta} = 1 \middle| \boldsymbol{a}\right) > \pi^* \\ \in \{0,1\} & \text{if} \quad \Pr\left(\overline{\theta} = 1 \middle| \boldsymbol{a}\right) = \pi^* \\ = 0 & \text{if} \quad \Pr\left(\overline{\theta} = 1 \middle| \boldsymbol{a}\right) < \pi^*, \end{cases} \tag{2.2}$$

that is, he prefers to convict the principal when his belief attaches probability at least π^* to the principal being guilty and vice versa. This payoff structure captures crime entanglement: the evaluator cares about the principal's overall probability of guilt, which is assessed based on all agents' reports. The principal's payoff is:

$$\sum_{i=1}^{n} \theta_i - sL, \tag{2.3}$$

with $L > 0$. According to (2.3), the principal's marginal benefit from committing each crime is normalized to 1, and his loss from conviction is L. Strategic agent i's payoff is:

$$u_i(\omega_i, \theta_i, a_i) \equiv \begin{cases} 0 & \text{if} \quad s = 1 \\ \omega_i - b\Big((1-\gamma)\theta_i + \gamma\overline{\theta}\Big) - ca_i & \text{if} \quad s = 0, \end{cases}$$

[15]We show in section 3 that having a positive fraction of non-strategic agents, no matter how small, (1) rules out bad trivial equilibria in which the principal abuses all agents for sure and is convicted no matter what (Lemma 3.1), and (2) guarantees the existence of equilibrium that survives our refinement. Our insights are robust under alternative specifications of the mechanical types' strategies (section 5 and Online Appendix H.2).

in which $b > 0$ measures his preference for punishing criminals, and $c > 0$ is his loss from the principal's retaliation, which is incurred when his accusation fails to convict the principal. Our formulation allows agents to have *social preferences* in the sense that they directly care about crimes against (or witnessed by) other agents, where the parameter $\gamma \in [0,1]$ measures the intensity of their social preferences. For illustration purposes, we focus on settings in which $\gamma = 0$ in the baseline model, namely,

$$u_i(\omega_i, \theta_i, a_i) = \begin{cases} 0 & \text{if} \quad s = 1 \\ \omega_i - b\theta_i - ca_i & \text{if} \quad s = 0. \end{cases} \tag{2.4}$$

In section 5.3, we provide intuition for why the effects identified in our results are stronger when γ is strictly positive. The proof of which can be found in Online Appendix H.1.

Interpretation of Payoff Functions: In our model, the principal faces a trade-off between the benefit from committing more crime and the loss from being convicted. Our main interest concerns settings in which L is large enough. This is the case, for example, when conviction entails the loss of one's job and/or opportunities for promotion (e.g., when the principal is a manager or a police officer), significant harm to the principal's public image (e.g., if the principal is a public figure), and a loss of power (e.g., when the principal is a politician or bureaucrat).

An agent's relative payoff when the principal is acquitted rather than convicted depends on three terms. First, it is affected by an idiosyncratic taste towards the principal, modeled by the i.i.d. preference shocks $\{\omega_i\}_{i=1}^n$. Second, each agent is more inclined to have the principal convicted when the principal is guilty, which is modeled by the strictly positive b.[16] Since (2.4) is agent i's utility at the reporting stage *after* the abuse has taken place, b is not the direct utility loss from the abuse, which has been sunk. Rather, it captures an agent's preference for justice, and his disutility from continuing to interact with a principal who has abused him before.

Third, filing a report incurs a cost c if the principal is acquitted. More generally, each agent's reporting cost is *strictly higher* when the principal is acquitted (i.e., agents may incur a smaller reporting cost when the principal is convicted). When the principal exerts some formal authority over the agents, the cost c may be interpreted as a retaliation cost (as in Chassang and Padró i Miquel 2018) if the accusations fail to remove the principal from power.[17]

[16] An alternative way to model this is to assume that the distribution of ω_i shifts to the left when θ_i increases.

[17] In section 5, we show that our results continue to hold in the following scenarios: (1) agents receive utility gains from telling the truth; (2) the principal faces a larger punishment than L when he is believed to have committed multiple

Depending on the application, the evaluator could be a judge when abuses are of a criminal nature and the punishment is enforced by the criminal justice system. The evaluator could also be the owner of a firm, of which the principal is a manager whose misbehavior can hurt the firm's public image. When the principal's behavior is legal but immoral, the evaluator stand for other members of the principal's community, in which case the punishment takes the form of ostracism or the loss of status within the community.

Since agents can accuse the principal no matter whether they have been abused or not, the evaluator must take into account this credibility problem when deciding whether to convict or acquit the principal. The conviction threshold π^* measures the evaluator's attitude towards the trade-off between convicting an innocent person and acquitting a guilty one. We take π^* as an exogenous parameter to reflect the evaluator's or the society's preference over this tradeoff, instead of as a design instrument.

3 Results

We focus on Bayesian Nash equilibria that satisfy two refinements: *presumption of innocence* and *symmetry*, defined in section 3.1. We first derive and compare the equilibrium outcomes with a single agent and with two agents, and highlight the significant differences between the two cases. We generalize our analysis to three or more agents in section 5.1 and derive comparative statics with respect to the number of agents.

3.1 Equilibrium Refinement & Preliminary Analysis

An equilibrium consists of $\big\{(\sigma_i)_{i=1}^n, \boldsymbol{\pi}, q\big\}$, in which $\sigma_i : \mathbb{R} \times \{0,1\} \to \Delta\{0,1\}$ is (strategic) agent i's strategy, $\boldsymbol{\pi} \in \Delta\Big(\{0,1\}^n\Big)$ is the principal's strategy, and $q : \{0,1\}^n \to [0,1]$ is the evaluator's strategy, that maps the agents' reporting profile to the probability of conviction. We introduce our presumption of innocence refinement:

Axiom 1 (Presumption of Innocence). $q(0,0,...,0) = 0$.

Axiom 1 requires that the principal is not convicted when nobody accuses him. It rules out equilibria in which the principal abuses all agents and is convicted no matter what. Those equilibria

crimes; (3) the principal has decreasing marginal returns from committing crimes; and (4) the principal's utility from abusing agents is privately known and can be arbitrarily small or even negative.

are not only unrealistic, but are also against the idea that a defendant should be convicted based on the accusations filed against him rather than on the sole basis of a prior belief. Lemma 3.1 shows that under presumption of innocence, the ex ante probability of crime is interior and the conviction probabilities are responsive to every agent's report.

Lemma 3.1. *In every Bayesian Nash equilibrium that satisfies presumption of innocence,*

1. $\overline{\theta} = 1$ *occurs with probability strictly between* 0 *and* 1.[18]

2. *For every* $i \in \{1, ..., n\}$, *there exists* $a_{-i} \in \{0,1\}^{n-1}$ *such that* $q(1, a_{-i}) \neq q(0, a_{-i})$.

The proof of this lemma and the next are in Appendix A.1 and A.2. An equilibrium is *symmetric* if all agents adopt the same strategy (i.e., $\sigma_1 = \sigma_2 = ... = \sigma_n$), and the principal's strategy $\boldsymbol{\pi}$ treats the agents symmetrically.[19]

Lemma 3.2. *In every symmetric Bayesian Nash equilibrium that satisfies presumption of innocence, for every* $\boldsymbol{a}, \boldsymbol{a'} \in \{0,1\}^n$ *with* $\boldsymbol{a} \succ \boldsymbol{a'}$, *we have* $q(\boldsymbol{a}) \geq q(\boldsymbol{a'})$.

According to Lemma 3.2, each agent's report (weakly) increases the conviction probability when the equilibrium is symmetric and satisfies presumption of innocence. A further implication of this lemma is: every agent's equilibrium strategy is characterized by two cutoffs: ω^* and ω^{**}, such that (strategic) agent i reports when $\omega_i \leq \omega^*$ and $\theta_i = 1$, or when $\omega_i \leq \omega^{**}$ and $\theta_i = 0$.

3.2 Single-Agent Benchmark

If there is only one agent, then according to Axiom 1, the principal is acquitted unless the agent files an accusation. Let q_s denote the probability that the principal is convicted if the agent reports. If the agent is abused ($\theta = 1$), then he prefers to report when:

$$\omega - b \leq (1 - q_s)(\omega - b - c), \quad \text{or equivalently,} \quad \omega \leq \omega_s^* \equiv b - c\frac{1 - q_s}{q_s}. \tag{3.1}$$

[18] This is related to albeit different from a well-known result in the literature on inspection games (Dresher 1962), in which crime occurs with positive probability due to the inspector's cost of inspection. In our model, it is because the principal is convicted only when the posterior probability of crime exceeds an interior cutoff.

[19] Online Appendix A shows that symmetry emerges endogenously in all sequential equilibria under two arguably weak refinements. A *monotonicity* refinement, which requires that the principal be convicted with a weakly higher probability when a larger subset of agents report against him. A *properness* refinement in the spirit of Myerson (1978), which requires that at every off-path history, agents believe that the principal is arbitrarily less likely to make more costly mistakes.

Similarly, if the agent is not abused ($\theta = 0$), then he prefers to report when:

$$\omega \leq (1-q_s)(\omega - c), \quad \text{or equivalently,} \quad \omega \leq \omega_s^{**} \equiv -c\frac{1-q_s}{q_s}. \tag{3.2}$$

Subtracting (3.1) from (3.2), we obtain $\omega_s^* - \omega_s^{**} = b$. The *informativeness* of the agent's report is measured by the likelihood ratio that the agent reports as a function of whether he has observed a crime or not:

$$\mathcal{I}_s \equiv \frac{\Pr(\text{agent reports} \mid \theta = 1)}{\Pr(\text{agent reports} \mid \theta = 0)} = \frac{\delta\Phi(\omega_s^*) + (1-\delta)\alpha}{\delta\Phi(\omega_s^{**}) + (1-\delta)\alpha}. \tag{3.3}$$

Thus, his report is uninformative when this ratio is 1, and is perfectly informative when this ratio is infinity. As the probability with which the agent is strategic goes to 1, this informativeness ratio converges to $\Phi(\omega_s^*)/\Phi(\omega_s^{**})$. Let π_s denote the probability that the principal commits crime in equilibrium. We characterize these equilibria and study their properties when the punishment to the convicted becomes arbitrarily large and the fraction of mechanical types becomes arbitrarily small.

Proposition 1. *There exists $\overline{L} > 0$ such that for every $L > \overline{L}$, there exists a unique equilibrium that satisfies Axiom 1. This unique equilibrium is characterized by $(\omega_s^*, \omega_s^{**}, q_s, \pi_s)$, which satisfies (3.1), (3.2),*

$$\delta q_s\Big(\Phi(\omega_s^*) - \Phi(\omega_s^{**})\Big) = 1/L \quad \textit{and} \quad \mathcal{I}_s\frac{\pi_s}{1-\pi_s} = \frac{\pi^*}{1-\pi^*}. \tag{3.4}$$

Moreover,

$$\lim_{L\to\infty}\lim_{\delta\to 1}\mathcal{I}_s = \infty \textit{ and } \lim_{L\to\infty}\lim_{\delta\to 1}\pi_s = 0. \tag{3.5}$$

The proof is in Appendix A.6. The highlight of this proposition is (3.5), which says that as the punishment to the convicted principal becomes arbitrarily large, the agent's report becomes arbitrarily informative and the equilibrium probability of crime vanishes to 0. In another word, large punishments help to deter crimes.

3.3 Two-Agent Environment

We now study the properties of *all* symmetric Bayesian Nash equilibria that satisfy presumption of innocence (or *equilibrium* for short) when there are two agents and compare it to the single-agent benchmark. We start by establishing the existence of such equilibria:

Proposition 2. *If L is large enough, then there exists an equilibrium.*

The proof is based on Brouwer's fixed point theorem (Online Appendix B). It generalizes to any arbitrary number of agents as well as to alternative specifications of the mechanical types' strategies. In what follows, we focus on values of L for which an equilibrium exists. Theorem 1 characterizes the common properties of *all* equilibria when L is large enough and compares them to the unique equilibrium in the single-agent benchmark.

Theorem 1. *There exists $\overline{L} \in \mathbb{R}_+$ such that when $L > \overline{L}$, every equilibrium is characterized by a tuple $(\omega_m^*, \omega_m^{**}, q_m, \pi_m) \in \mathbb{R}_- \times \mathbb{R}_- \times [0,1] \times [0,1]$ such that*

1. *For every $i \in \{1, 2\}$, agent i reports when $\{\omega_i \leq \omega_m^*$ and $\theta_i = 1\}$ or $\{\omega_i \leq \omega_m^{**}$ and $\theta_i = 0\}$.*
2. *The principal chooses $(\theta_1, \theta_2) = (0, 0)$ with probability $1-\pi_m$, $(\theta_1, \theta_2) = (1, 0)$ with probability $\pi_m/2$, and $(\theta_1, \theta_2) = (0, 1)$ with probability $\pi_m/2$.*
3. *The conviction probabilities satisfy $q(0, 0) = q(0, 1) = q(1, 0) = 0$ and $q(1, 1) = q_m$.*
4. *Compared to the unique equilibrium in the single-agent benchmark, we have: $\omega_m^* > \omega_s^*$, $\omega_m^{**} > \omega_s^{**}$, $q_m > q_s$, $\pi_m > \pi_s$ and $\mathcal{I}_s > \mathcal{I}_m$ in which*

$$\mathcal{I}_m \equiv \frac{\Pr(\textit{both agents report} \mid \overline{\theta} = 1)}{\Pr(\textit{both agents report} \mid \overline{\theta} = 0)} = \frac{\Pr(\textit{agent } i \textit{ reports} \mid \theta_i = 1)}{\Pr(\textit{agent } i \textit{ reports} \mid \theta_i = 0)}, \quad \textit{for } i \in \{1, 2\}.$$

The proof consists of three parts, which are in Appendix A.3 (comparison to the single-agent benchmark), Appendix A.4 (principal's incentives), and Online Appendix C.2 (equilibrium conviction probabilities). According to Theorem 1, all equilibria share the following properties when L is large: First, the principal abuses each agent with positive probability but never abuses both agents at the same time. Second, the principal is convicted with strictly positive probability only when two reports are filed against him. Third, compared to the single-agent benchmark, reports are less informative both at the individual and at the aggregate level, the equilibrium probability of crime is strictly higher, and more interestingly, strategic agents are *more* likely to file reports compared to the single-agent benchmark. The last feature distinguishes our result from the theories of public good provision (Chamberlin 1974), in which inefficiency is driven by the scarcity of contributions. We will generalize these comparative statics to any finite number of agents in Proposition 7.

Our second theorem examines the informativeness of the agents' reports and the equilibrium probability of crime as the punishment to benefit ratio, L, becomes arbitrarily large.

Theorem 2. *For every $\epsilon > 0$, there exists $\overline{L}_\epsilon \geq 0$ such that when $L > \overline{L}_\epsilon$, every equilibrium satisfies: $\omega_m^* < -1/\epsilon$, $\omega_m^{**} < -1/\epsilon$, $\mathcal{I}_m < 1 + \epsilon$, and $\pi_m \geq \pi^* - \epsilon$.*

The proof is in Appendix A.5. Theorem 2 suggests that regardless of the relative magnitudes of b and c, as the punishment to the convicted principal becomes arbitrarily large relative to his benefit of committing crime, agents' reports become *arbitrarily uninformative* about whether a crime has occurred or not, and the probability of crime converges to the conviction threshold π^*.[20] This stands in sharp contrast to the single-agent benchmark, in which report becomes arbitrarily informative and the probability of crime vanishes to zero.[21]

Given the presence of mechanical type agents whose reports transmit no information, one may suspect that $\mathcal{I}_m \to 1$ is driven by the scarcity of reports filed by the strategic types. Statement 4 of Theorem 1 refutes this conjecture. In particular, the probability with which a strategic type files an accusation is *strictly higher* in the two-agent case compared to the single-agent benchmark. Since the agent's report is arbitrarily informative as $\delta \to 1$ and $L \to \infty$ in the single-agent benchmark, their reports becoming arbitrarily uninformative under the same limit in the two-agent case cannot be driven by the scarcity of reports filed by strategic types.

Intuition: We argue that Theorems 1 and 2 are driven by the tension between the following two forces, both of which arise endogenously when L is large enough. First, the negative correlation in agents' private signals, θ_1 and θ_2; and second, the agents' coordination motive in reporting.

We begin by explaining why, when L is large enough, the principal is convicted only when he has been accused by both agents. Suppose by way of contradiction that a single accusation suffices to generate a conviction with positive probability. Then, the principal must be surely convicted if two accusations are leveled against him, because two accusations make him strictly more likely of having committed crime, compared to a single one. We show in Proposition C.2 (Online Appendix C.2) that the above observation implies a uniform lower bound on the marginal increase in conviction probabilities when the principal commits an extra crime. When L is large enough, this uniform lower bound implies that the principal has a strict incentive not to commit any crime, contradicting the

[20] Although a lower threshold π^* reduces the probability of crime, it increases the fraction of wrongful convictions, which must be equal to $1 - \pi^*$ in equilibrium. This illustrates a classic tradeoff between improving deterrence and reducing wrongful convictions.

[21] For this comparison, we are referring to the double limit in which $\lim_{L\to\infty} \lim_{\delta\to 1}$. Economically, this captures situations in which mechanical type agents are arbitrarily rare and the majority of agents are strategic.

conclusion of Lemma 3.1 that the equilibrium probability of crime is interior. Therefore, $q(0,0) = q(1,0) = q(0,1) = 0$.

Since the principal's marginal benefit from committing a crime is constant, whether his choices of θ_1 and θ_2 are strategic complements or strategic substitutes depends on the sign of

$$q(1,1) + q(0,0) - q(1,0) - q(0,1), \tag{3.6}$$

which determines whether the conviction probability is convex or concave in the number of reports. We show in Lemma A.1 of Appendix A.4 that

1. If (3.6) is strictly positive, then the choices of θ_1 and θ_2 are strategic substitutes.

2. If (3.6) is strictly negative, then the choices of θ_1 and θ_2 are strategic complements.

The previous step implies that (3.6) is strictly positive under a large L. This, together with Lemma 3.1, implies that the principal must be indifferent between abusing no agent and abusing only one agent, but he has a strict incentive not to abuse both agents. This leads to an *endogenous negative correlation* between θ_1 and θ_2.

From the perspective of agent 1, the equilibrium conviction probabilities imply that he has more incentive to report when he believes that agent 2 is more likely to report, since it increases his chances of avoiding the retaliation cost c. Based on the principal's equilibrium strategy, if $\theta_1 = 1$, then agent 1 believes that agent 2 is not abused and is less likely to report; if $\theta_1 = 0$, then agent 1 believes that agent 2 is likely to be abused and is more likely to report. In summary, agent 1's incentives to coordinate with agent 2 discourages him from reporting when he has been abused, and encourages him to report when he has not been abused. In summary, the agents have negatively correlated private information but their decisions to accuse the principal are strategic complements. These forces undermine the credibility of their reports and increase the probability of crime.

As $L \to \infty$, two competing effects arise. First, as in the single-agent benchmark, the probability of false accusations decreases, which improves informativeness. Second, the distance between the reporting cutoffs of abused and intact agents decreases, which is caused by the negative correlation and the agents' coordination motives. This undermines informativeness. Theorem 2 shows that the second effect dominates the first one, irrespective of the magnitude of b and c.

To better understand the interplay between these two forces, we sketch a proof of Theorem 2 by deriving formulas for the agents' reporting cutoffs, the informativeness ratio, and the conviction probabilities. Throughout these derivations, we take statements 1, 2, and 3 of Theorem 1 as given. First, conditional on $\theta_i = 0$, the probability with which the principal is innocent is:

$$\beta \equiv \frac{1-\pi_m}{1-\pi_m/2}. \tag{3.7}$$

For agent i, if $\theta_i = 1$, then he prefers to report when

$$\omega \le \omega_m^* \equiv b - c\frac{1-q_m Q_1}{q_m Q_1} = b + c - \frac{c}{q_m Q_1}, \tag{3.8}$$

where Q_1 is the probability that the other agent accuses the principal *conditional on* $\theta_i = 1$. Similarly, if $\theta_i = 0$, then agent i prefers to report when

$$\omega \le \omega_m^{**} \equiv -c\frac{1-q_m Q_0}{q_m Q_0} = c - \frac{c}{q_m Q_0}, \tag{3.9}$$

where Q_0 is the probability that the other agent accuses the principal *conditional on* $\theta_i = 0$. The expressions for Q_1 and Q_0 are given by

$$Q_1 = \delta\Phi(\omega_m^{**}) + (1-\delta)\alpha \tag{3.10}$$

and

$$Q_0 = \delta\Big(\beta\Phi(\omega_m^{**}) + (1-\beta)\Phi(\omega_m^*)\Big) + (1-\delta)\alpha. \tag{3.11}$$

The comparisons between Q_1 and Q_0 and between ω_m^* and ω_m^{**} unveil the key difference between the two-agent scenario and the single-agent benchmark. Instead of having a constant distance b, the distance between the two reporting cutoffs is:

$$\omega_m^* - \omega_m^{**} = b - \frac{c}{q_m} \cdot \frac{-1 + Q_0/Q_1}{Q_0}. \tag{3.12}$$

This leads to the following lemma:

Lemma 3.3. $\omega_m^* - \omega_m^{**} \in (0, b)$.

Proof of Lemma 3.3: From (3.10) and (3.11), $\omega_m^* - \omega_m^{**} > 0$ is equivalent to $Q_0 > Q_1$. To see this, suppose by way of contradiction that $Q_0 \leq Q_1$. Equation (3.12) then implies that $\omega_m^* \geq \omega_m^{**} + b > \omega_m^{**}$. The comparison between (3.8) and (3.9) then yields $Q_0 > Q_1$, the desired contradiction. Since $Q_0 > Q_1$, the term $\frac{-1+Q_0/Q_1}{Q_0}$ is strictly positive. Therefore, $\omega_m^* - \omega_m^{**} < b$. □

Next, we explore how the decrease in $\omega_m^* - \omega_m^{**}$ (relative to the single-agent benchmark) affects the informativeness of reports and the equilibrium probability of crime. First, we provide a formula for the informativeness ratio $\mathcal{I}_m$, which has been defined in the fourth statement of Theorem 1:

$$\mathcal{I}_m \equiv \frac{\big(\delta\Phi(\omega_m^*) + (1-\delta)\alpha\big)\big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\big)}{\big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\big)^2} = \frac{\delta\Phi(\omega_m^*) + (1-\delta)\alpha}{\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha}. \tag{3.13}$$

Since $q_m \in (0,1)$, the evaluator believes that $\theta_1\theta_2 = 0$ with probability π^* after observing two reports. Therefore, the equilibrium probability of crime π_m satisfies:

$$\frac{\pi_m}{1-\pi_m} = \frac{l^*}{\mathcal{I}_m}, \quad \text{in which} \quad l^* \equiv \frac{\pi^*}{1-\pi^*}. \tag{3.14}$$

Plugging (3.14) into (3.7), we have the following expressions for β and $1-\beta$:

$$\beta = \frac{2\mathcal{I}_m}{l^* + 2\mathcal{I}_m} \text{ and } 1-\beta = \frac{l^*}{l^* + 2\mathcal{I}_m}. \tag{3.15}$$

Plugging (3.15) into (3.10) and (3.11), we obtain the following expression for the ratio between Q_0 and Q_1:

$$\frac{Q_0}{Q_1} = \beta + (1-\beta)\mathcal{I}_m = \frac{(l^*+2)\mathcal{I}_m}{l^* + 2\mathcal{I}_m}. \tag{3.16}$$

Plugging (3.8) and (3.9) into (3.16), we obtain

$$\frac{|\omega_m^* - c - b|}{|\omega_m^{**} - c|} = \frac{c/q_m Q_1}{c/q_m Q_0} = \frac{Q_0}{Q_1} = \frac{(l^*+2)\mathcal{I}_m}{l^* + 2\mathcal{I}_m}. \tag{3.17}$$

This leads to the following lemma:

Lemma 3.4. *If $\omega_m^* \to -\infty$, then $\mathcal{I}_m \to 1$ and $\pi_m \to \pi^*$.*

Proof of Lemma 3.4: Since $\omega_m^* - \omega_m^{**} \in (0, b)$, the difference between $|\omega_m^* - c - b|$ and $|\omega_m^{**} - c|$ is at most b. The LHS of (3.17) converges to 1 as $\omega_m^* \to -\infty$. Since the RHS of (3.17) is strictly increasing in $\mathcal{I}_m$, we know that the limiting value of $\mathcal{I}_m$ is 1. According to (3.14), the limiting value of π_m is π^*. □

To complete this heuristic proof of Theorem 2, we now argue that ω_m^* and ω_m^{**} converge to $-\infty$ as $L \to \infty$. In equilibrium, the principal is indifferent between committing one crime and committing no crime:

$$\frac{1}{\delta L} = q_m\Big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\Big)\Big(\Phi(\omega_m^*) - \Phi(\omega_m^{**})\Big). \tag{3.18}$$

If ω_m^{**} converges to some finite number ω^{**}, then either $q_m \to 0$ or $|\omega_m^* - \omega_m^{**}| \to 0$. In the first case where $q_m \to 0$, (3.8) and (3.9) suggest that both ω_m^* and ω_m^{**} converge to $-\infty$. This contradicts the hypothesis that ω_m^{**} converges to some finite number. In the second case where $|\omega_m^* - \omega_m^{**}| \to 0$ and ω_m^{**} converges to some finite number, then $\mathcal{I}_m \to 1$ and $Q_0/Q_1 \to 1$. Expression (3.12) then implies that the limiting value of $|\omega_m^* - \omega_m^{**}|$ is b. This contradicts that $|\omega_m^* - \omega_m^{**}|$ converging to 0.

4 Restoring the Credibility of Reports

We propose several solutions to restore the credibility of crime reports and reduce the probability of crime, focusing for simplicity on the case of two agents. In section 4.1, we study an alternative conviction rule that disentangles crimes by treating the probability of committing each crime separately. Our result provides support to a key feature of criminal justice systems, which is to treat distinct accusations separately in conviction decisions. In section 4.2, we show that under the conviction rule in the baseline model, the probability of crime is lower under some intermediate levels of punishment. In section 4.3, we construct a transfer scheme that restores reporting credibility and reduces the probability of crime.

4.1 Alternative Conviction Rule: Disentangling Crimes

We consider an alternative conviction rule that treats crimes against different individuals separately. Conviction is decided by first evaluating the probability with which the principal has abused each agent, and then comparing the maximum of these probabilities to some conviction threshold. Formally, let $\pi_i \equiv \mathbb{E}[\theta_i|\boldsymbol{a}]$ be the probability that the principal is guilty of abusing agent i after observing all agents' reports, and the conviction decision is made according to:[22]

$$s \begin{cases} = 1 & \text{if} \quad \max_{i\in\{1,2\}} \pi_i > \pi^* \\ \in \{0,1\} & \text{if} \quad \max_{i\in\{1,2\}} \pi_i = \pi^* \\ = 0 & \text{if} \quad \max_{i\in\{1,2\}} \pi_i < \pi^*. \end{cases} \tag{4.1}$$

We remind the reader of the notation used in the the single-agent benchmark of section 3.2: ω_s^* and ω_s^{**} are the reporting cutoffs, q_s is the principal's conviction probability in case he is accused, and π_s is the equilibrium probability of crime. We have the following result:

Proposition 3. *Under conviction rule (4.1) there is for L large enough an equilibrium s.t.*

1. *The principal abuses each agent with probability π_s, and the probability with which he abuses agent i is independent of whether he abuses agent j or not.*
2. *The principal is convicted with probability mq_s where $m \in \{0,1,2\}$ is the number of accusations filed against him.*
3. *Agent i accuses the principal if and only if $\omega_i \leq \omega_s^*$ and $\theta_i = 1$, or $\omega_i \leq \omega_s^{**}$ and $\theta_i = 0$.*
4. *The equilibrium probability of crime is $1-(1-\pi_s)^2$.*

In the limit where $\lim_{L\to\infty}\lim_{\delta\to 1}$, the equilibrium probability of crime vanishes to 0.

This proposition provides one of the main insights of our analysis: using the maximum probability with which the principal is guilty of a *specific crime*, instead of the probability with which he is guilty of *at least one crime*, as the basis for conviction eliminates the negative correlation in the agents' private observations and restores the credibility of their reports. The result may be generalized to any finite number of agents, by changing the probability of crime from $1-(1-\pi_s)^2$ to $1-(1-\pi_s)^n$.

Note that Proposition 3 is not driven by the assumption that the principal can be convicted with any arbitrary probability at the conviction threshold π^*. In general, under any increasing function, that maps $\max_{i\in\{1,2\}} \pi_i$ to a level of punishment, one can show that agents' observations of crime cannot be negatively corrected in any equilibrium that respects *presumption of innocence* (Axiom 1) and a *monotonicity axiom* (Axiom 2 in Online Appendix A) which says that the expected punishment is nondecreasing when a larger set of agents report.

[22] A more straightforward solution is to punish the principal by mL when $\left|\{i|\pi_i \geq \pi^*\}\right| = m$. However, this is not feasible when punishment is by nature discrete and not easily scalable with respect to the number of crimes. This is the case when the loss from conviction mainly comes from its side effects, such as the loss of one's job, power, or influence.

The comparison between Proposition 3 and Theorems 1 and 2 provides a rationale for treating crimes against different individuals separately. Nevertheless, implementing such a conviction rule in practice requires the evaluator to have commitment power. To illustrate, let us revisit the numerical example in the introduction:

	Prob of abusing victim 1	Prob of abusing victim 2	Prob of abusing at least one victim
Defendant 1	49 %	49 %	98 %

	Prob of abusing victim 3	Prob of abusing victim 4	Prob of abusing at least one victim
Defendant 2	60 %	1 %	61 %

Under the conviction rule in the baseline model, defendant 1 is more likely to be convicted compared to defendant 2, which coincides with most people's preferences as well as those of the society's. Under conviction rule (4.1), however, defendant 1 is acquitted for sure as long as defendant 2 is acquitted despite the former is almost surely guilty of at least one crime. Therefore, our new conviction rule requires the evaluator to make conviction decisions against his own preferences or against the society's preferences after observing the agents' reports (or more generally, after collecting evidence related to the cases).

The comparison between the two conviction rules has different implications in judicial decision-making and in informal conviction decisions (such as decisions made by firms and organizations to promote, demote or fire their members or employees). First, the required commitment power to implement the second conviction rule is enshrined in our current legal framework. Under procedural law, judges are required to treat different crimes separately, and a defendant can only be convicted when it is beyond reasonable doubt that he has committed a well-defined crime (e.g., abusing a particular agent). In another word, judges are not allowed to convict people based on the high likelihood that they are guilty of at least some crime. Our results provide a novel rationale for this aspect of the criminal justice system: the conviction rule in the baseline model (one that is based on crime entanglement) induces potential criminals to negatively correlate his criminal behaviors across different potential victims. In equilibrium, this strategic response of the potential criminal weakens the credibility of the potential victims' reports and increases the chances with which crimes taking place.

Second, in context of informal conviction decisions, namely, those carried out by firms, organizations, communities, and public opinions, it is reasonable to assume that decision makers face lack-of-commitment problems. For instance, firms face social pressure to fire a manager whose probabilities of abusing his subordinates are given by the first row, compared to a manager whose probabilities of committing abuses are given by the second row. Political parties may face public pressure to ostracize a party member with a bad reputation (e.g., an individual that is believed to have committed at least some offenses with high probability), even when the probability of him committing any of the specific offenses was not be proved to be high. In these scenarios, our results demonstrate the importance of insulating evaluators from public pressure.

Comparison with Single-Agent Benchmark: The single-agent benchmark in section 3.2 has an alternative interpretation: the evaluator is required to treat crimes against different agents separately, and furthermore, he *cannot* bring evidence from other criminal cases. Formally, the evaluator makes his conviction decisions based on $\mathbb{E}[\theta_i|a_i]$ instead of $\mathbb{E}[\theta_i|\boldsymbol{a}]$. The comparison between Proposition 3 and Proposition 1 suggests that allowing judges to bring evidence from other cases does not hurt the informativeness of potential victims' testimonies.

4.2 Moderate Punishment

We show that lowering the punishment to the convicted principal improves the informativeness of the agents' reports and reduces the probability of crime.

Proposition 4. *For every $c > 0$, there exists an interval $[\underline{L}(c), \overline{L}(c)]$ such that an equilibrium exists when $L \in [\underline{L}(c), \overline{L}(c)]$. In every equilibrium, the conviction probabilities satisfy:*

$$q(1,1) + q(0,0) - q(1,0) - q(0,1) < 0.$$

For every $\epsilon > 0$, there exists $c_\epsilon > 0$ such that when $c > c_\epsilon$ and $L \in [\underline{L}(c), \overline{L}(c)]$, there exists an equilibrium in which the probability of crime is less than ϵ.

The proof is in Online Appendix D. Proposition 4 shows that there exists an intermediate range of L, under which the conviction probabilities are *concave* in the number of reports in all equilibria. As a result, the principal's decisions are strategic complements. In equilibrium, he either abuses no agent or abuses both agents. Moreover, abusing a single agent is strictly suboptimal.

The principal's equilibrium strategy induces a *positive correlation* between θ_1 and θ_2. In contrast to the case in which L is large, each agent's coordination motive now encourages him to report when he has witnessed a crime and vice versa. This increases the distance between his two reporting thresholds, making it strictly larger than b. As a result, the informativeness of reports increases and the probability of crime decreases. As each agent's loss from miscoordination becomes arbitrarily large, each individual report becomes arbitrarily informative and the equilibrium probability of crime converges to zero.

Proposition 4 implies that in order to minimize the probability of crime, the optimal magnitude of punishment is *interior* when there are multiple potential witnesses who are vulnerable to retaliation. This finding differs from Becker's (1968) seminal analysis of criminal justice and law enforcement, which suggests that increasing the magnitude of punishment helps reduce crime. From this perspective, our finding provides a novel rationale for being lenient to the convicted. Our logic applies to settings in which smoking-gun evidence is scarce and the potential victims' claims are hard to verify.

Importantly, reducing L comes at the cost of increasing the number of crimes conditional on the principal being guilty. This reveals a tradeoff between reducing the probability of crime and reducing the number of crimes.

4.3 Monetary Transfers

In this section, we maintain the assumption that L is large and explore the use of monetary transfers to mitigate the coordination problem across agents. The transfers are contingent on the vector of reports: we let $t_i(\boldsymbol{a}) \in \mathbb{R}$ denote the transfer to agent i under reporting profile $\boldsymbol{a}$. We begin by constructing a transfer scheme that eliminates coordination inefficiencies. Let

$$t_1^*(a_1, a_2) = \begin{cases} c & \text{if } (a_1, a_2) = (1, 0) \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad t_2^*(a_1, a_2) = \begin{cases} c & \text{if } (a_1, a_2) = (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

Under transfer scheme (t_1^*, t_2^*), the next proposition shows that agents' reports become arbitrarily informative and the probability of crime vanishes to zero as $L \to \infty$.

Proposition 5. *For every $\epsilon > 0$, there exists $\overline{L}_\epsilon > 0$ such that for all $L > \overline{L}_\epsilon$ and under transfer scheme (t_1^*, t_2^*), the informativeness ratio of each agent's report exceeds $1/\epsilon$ and the probability of crime is less than ϵ in all equilibria.*

The proof is in Online Appendix E.1. Proposition 5 implies that when it is hard to scale down punishment (e.g., conviction has other negative consequences on one's career that is beyond the control of the judge), one can restore the informativeness of reports by compensating lone accusers. The amount to be transferred exactly offsets an agent's loss from retaliation, which eliminates the agents' incentives to coordinate. As a result, the distance between the two reporting cutoffs equals b. However, the equilibrium outcome does not coincide with that of the single-agent benchmark: the principal's incentives are different, and the cutoffs under the two-agent setting with transfer scheme (t_1^*, t_2^*) are strictly higher than those arising in the single-agent benchmark. For any given L, the informativeness of reports is strictly lower and the probability of crime is strictly higher than in the single-agent setting.

The above transfer scheme presents two weaknesses. First, the designer needs to incur a budget deficit with positive probability. This deficit is increasing in the agents' losses from retaliation. Second, this scheme encourages collusion between the principal and the agents. For example, the principal and agents can agree that only agent 1 reports and the principal does not retaliate. After agent 1 obtains the transfer c, he will share it with agent 2 and the principal.

Motivated by these concerns—as well as by appropriately positioning our work relative to Crémer and McLean (1988)—we explore the possibility of restoring reporting informativeness via budget -balanced transfer schemes, namely, transfers such that $\sum_{i=1}^{2} t_i(\boldsymbol{a}) = 0$ for every $\boldsymbol{a} \in \{0,1\}^2$. In Online Appendix E.2, we show that the informativeness of agents' reports is uniformly bounded from above when L is large. This is true even when we allow for asymmetric equilibria and $t(\cdot)$ to be asymmetric across agents. Intuitively, this is driven by the tension between mitigating the adverse effect of coordination motives and deterring false accusations.[23]

5 Extensions

Section 5.1 extends our theorems to three or more agents. Section 5.2 allows heterogeneous propensities to commit crimes and shows the robustness of our insights against such heterogeneity. This extension also accounts for the existence of people who commit multiple crimes. Section 5.3 explores several variants of the baseline model, which include the arrival of ex post evidence that exposes false accusations, the agents and the evaluator facing uncertainty about the number of potential victims, alternative payoff functions for the principal and agents, and the mechanical types' reports being

[23]The results in Crémer and McLean (1988) do not apply in our setting since the agents' payoff shocks are independent. Therefore, their convex independence condition fails.

informative about the true state.

5.1 Arbitrary Number of Agents

We say that an equilibrium is *unanimous* if $q(\boldsymbol{a}) = 0$ unless $\boldsymbol{a} = (1, 1, ..., 1)$, namely, no conviction unless agents unanimously report. According Theorem 1, when there are two agents, every symmetric equilibrium that survives the presumption of innocence refinement must be unanimous. Proposition 6 generalizes this insight to three or more agents:

Proposition 6. *For every $n \geq 2$, there exists $\overline{L}_n \in \mathbb{R}_+$ such that when $L > \overline{L}_n$,*

1. *There exists a symmetric unanimous equilibrium.*
2. *Every symmetric equilibrium that satisfies Axiom 1 must be an unanimous equilibrium. In this equilibrium, the principal abuses at most one agent.*

For every $\epsilon > 0$, there exists $\overline{L}_{n,\epsilon} \geq \overline{L}_n$ such that when $L > \overline{L}_{n,\epsilon}$, in every symmetric unanimous equilibrium, each agent's reporting cutoffs are lower than $-1/\epsilon$, the aggregate informativeness of reports lower than $1 + \epsilon$, and the equilibrium probability of crime is greater than $\pi^* - \epsilon$.

The proof is in Online Appendix F. The intuition is similar to the two agent case. If the punishment in case of a conviction is large enough and the principal is convicted for sure when all agents unanimously report, then he has a strict incentive not to commit any crimes which contradicts the conclusion of Lemma 3.1. As a result, the principal is convicted with positive probability only when agents unanimously report. As long as the conviction probabilities possess this property, the principal's marginal cost of abusing an additional agent is increasing in the number of agents that have already been abused. Consequently, either no agent is abused or only one agent is abused in equilibrium. This property of the principal's equilibrium strategy induces a negative correlation in the agents' private observations of crime. This, together with the agents' incentive to coordinate their report, discourages them from accusing the principal when they have been abused and vice versa.

Previous sections have established that moving from a single agent to two agents reduces informativeness and increases the probability of crime under a large punishment. We now extend this result by studying the comparative statics with respect to the number of agents, focusing on symmetric unanimous equilibria. In an environment with $n \in \mathbb{N}$ agents, let ω_n^* and ω_n^{**} be an agent's reporting cutoffs, let π_n

be the equilibrium probability of crime, and let

$$\mathcal{I}_n \equiv \frac{\Pr(n \text{ agents report} |\, \overline{\theta} = 1)}{\Pr(n \text{ agents report} |\, \overline{\theta} = 0)}$$

be the aggregate informativeness of the agents' reports. According to Proposition 6, at most one agent is abused in any unanimous equilibrium. Therefore, the aggregate informativeness of reports coincides with the informativeness ratio of an individual report, namely,

$$\mathcal{I}_n = \frac{\Pr(a_i = 1|\theta_i = 1)}{\Pr(a_i = 1|\theta_i = 0)} = \frac{\delta\Phi(\omega_n^*) + (1-\delta)\alpha}{\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha}.$$

Recall the definition of $\overline{L}_n$ in Proposition 6, we have the following result:

Proposition 7. *For every k and n with $k > n$ and $L > \max\{\overline{L}_n, \overline{L}_k\}$, the following inequalities hold once we compare any symmetric unanimous equilibrium when there are k agents with any symmetric unanimous equilibrium when there are n agents: (1) $\omega_k^* - \omega_k^{**} < \omega_n^* - \omega_n^{**}$; (2) $\mathcal{I}_n > \mathcal{I}_k$; (3) $\pi_k > \pi_n$; (4) $\omega_k^* > \omega_n^*$ and $\omega_k^{**} > \omega_n^{**}$.*

The proof is in Online Appendix F.1. Proposition 7 shows that as the number of potential victims increases, the distance between the reporting cutoffs decreases, the informativeness of reports decreases, both at the individual and at the aggregate level. As a result, the probability of crime increases. More interestingly, each potential victim is *more likely* to report when there are more potential victims, no matter whether he has been abused or not. The driving force behind such comparative statics is still the interaction between the coordination motives among agents and the negative correlation in their private observations of crime.

5.2 Heterogeneous Principals

We now consider two types of principal: a *virtuous* type, who has no benefit from committing crimes (and never commits crimes) and an *opportunistic* one with the same payoff function as in the baseline model. The principal's type is private and the prior probability of the virtuous type is π^v. We consider two cases, based on the relative values of π^v and $1 - \pi^*$.

Low Probability of a Virtuous Type: When $\pi^v < 1 - \pi^*$, the equilibrium outcome remains the same as in the baseline model. In particular, there exists $\overline{L} \in \mathbb{R}_+$ such that when $L > \overline{L}$, every equilibrium is characterized by a quadruple $(\omega^*, \omega^{**}, q, \pi) \in \mathbb{R} \times \mathbb{R} \times (0,1) \times (0, \pi^*)$ such that:

1. Agent i accuses the principal when $\{\omega_i \leq \omega^* \text{ and } \theta_i = 1\}$ or when $\{\omega_i \leq \omega^{**} \text{ and } \theta_i = 0\}$.

2. An opportunistic principal chooses $(\theta_1, \theta_2) = (0,0)$ with probability $\frac{1-\pi}{1-\pi^v}$, $(\theta_1, \theta_2) = (1,0)$ with probability $\frac{\pi}{2(1-\pi^v)}$, and $(\theta_1, \theta_2) = (0,1)$ with probability $\frac{\pi}{2(1-\pi^v)}$.

3. The conviction probabilities satisfy $q(0,0) = q(0,1) = q(1,0) = 0$ and $q(1,1) = q$.

When L is large and the fraction of mechanical agent vanishes ($\delta \to 1$), the agents' reports become arbitrarily uninformative and the equilibrium probability of crime converges to π^*.

High Probability of a Virtuous Type: When $\pi^v > 1 - \pi^*$ and L is large enough, the following holds: First, two reports are required to convict the principal. As a result, the principal's decisions to commit crimes against the agents are strategic substitutes and the agents' decisions to report crimes are strategic complements. Second, the equilibrium features a high probability of crime, in the sense that an opportunistic principal commits crime for sure and, moreover, commits two crimes with strictly positive probability. Third, agents' private signals are negatively correlated, as in the benchmark setting with a single type of principal. Unlike in the single-agent benchmark ad the baseline model with two agents, the informativeness of agents' reports, given by

$$\mathcal{I} \equiv \frac{\Pr(a_1 = a_2 = 1|\overline{\theta} = 1)}{\Pr(a_1 = a_2 = 1|\overline{\theta} = 0)},$$

is constant for large enough L, equal to $\frac{1-\pi^v}{\pi^v} \Big/ \frac{\pi^*}{1-\pi^*}$. We formalize this in Proposition 8:

Proposition 8. *When $\pi^v > 1 - \pi^*$, there exists $\overline{L} \in \mathbb{R}_+$ such that for every $L \geq \overline{L}$:*

1. $q(0,0) = q(1,0) = q(0,1) = 0$ *and* $q(1,1) \in (0,1)$.
2. *The equilibrium prob of crime is $1 - \pi^v$ and the informativeness ratio $\mathcal{I}$ equals $\frac{1-\pi^v}{\pi^v} \Big/ \frac{\pi^*}{1-\pi^*}$.*
3. *An opportunistic principal commits two crimes with positive probability.*
4. $\Pr(\theta_i = 1|\theta_j = 1) < \Pr(\theta_i = 1|\theta_j = 0)$.

The proof is in Online Appendix G. Proposition 8 shows that in a society with a significant fraction of virtuous people, using the overall probability of guilt (taken into account all the reports and crimes against all agents) as a criterion for conviction leads to a high probability of crime in equilibrium.[24]

[24]Since the virtuous principal never commits any crime, $1 - \pi^v$ is the highest possible probability of crime.

As before, a conviction rule that entangles crimes in this way generates an endogenous negative correlation in the agents' private observations and creates an endogenous coordination motive across agents in reporting crimes. The addition of virtuous types conveys the robustness of our insights when there is heterogeneity in the propensity to commit crimes, and also explains the presence of criminals that commit multiple crimes.[25]

The intuition for the first statement is the same as in the baseline model: If conviction happens for sure following two reports and the punishment in case of conviction is large enough, the opportunistic principal has a strict incentive not to commit crimes, which cannot occur in equilibrium. The conviction probabilities computed in Statement 1 then follow, and imply that the principal's decisions to commit crimes are substitutes and the agents' incentives to report crimes are complements. Since the ex ante probability of crime cannot exceed $1-\pi^v$, the informativeness ratio $\mathcal{I}$ must exceed $\frac{1-\pi^v}{\pi^v} \Big/ \frac{\pi^*}{1-\pi^*}$, which implies that an opportunistic principal has a strict incentive to commit crime.[26] The agents' private signals must be negatively correlated: a positive correlation would imply that $\omega^* - \omega^{**} \geq b$, causing $\mathcal{I}$ to explode as $L \to \infty$. This would contradict the property of conviction probabilities stated in Statement 1 and the previous conclusion that an opportunistic type principal surely commits crime.

5.3 Other Extensions

Decreasing Marginal Benefits from Crime: Our result remains robust when the principal faces decreasing marginal returns from committing multiple crimes (Becker 1968) or receives a punishment larger than L when he is believed to have committed multiple crimes. These changes motivate the principal to commit fewer crimes and induce, as in the baseline model, negative correlation in the agents' private information. As in the baseline model, the agents' coordination motives undermine the informativeness of their reports and increase the probability of crime. In Online Appendix H.3, we study an extension of the baseline model that formalizes these arguments. In particular, the principal is convicted of a minor crime and receives punishment L if the probability with which he is guilty of at least one crime exceeds π^*, and is convicted of a felony and receives punishment $L'(> L)$ if the probability with which he is guilty of two crimes exceeds π^{**}. When L is large (e.g., the principal can lose his lucrative position when convicted with a minor offense), the principal commits at most one

[25] As a corollary of Proposition 8, the aforementioned insights of ours are also robust when the principal can be a vicious type with small but positive probability, namely, a type that mechanically commits crime against two agents.

[26] Otherwise, the equations derived in section 3.4 still apply, and the informativeness ratio converges to 1 as $L \to \infty$.

crime in every symmetric equilibrium that satisfies presumption of innocence (Axiom 1).

Agent's Reporting Cost: Our results remain the same when agents' reporting costs are positive when the principal is convicted, as long as their reporting costs are strictly higher when the principal is acquitted. Our results also extend when each agent's suffers a lower retaliation cost when there are more reports filed against the principal, as long as the retaliation cost is strictly positive whenever the principal is acquitted. This variation strengthens the coordination motives among agents without affecting the negative correlation between their private information.

Agent's Interdependent Preferences: As mentioned in section 2, agents may directly care about crimes committed against or witnessed by other agents. Recall that in general, strategic agent i's payoff function is given by:

$$\Big\{ \underbrace{\omega_i - b\Big((1-\gamma)\theta_i + \gamma\overline{\theta}\Big) - ca_i}_{\text{payoff when the principal is acquitted}} \Big\} \qquad \underbrace{(1-s)}_{\text{event that principal is acquitted}} \qquad . \tag{5.1}$$

According to (5.1), agent i's payoff when the principal is convicted is normalized to 0. His payoff when the principal is acquitted depends on whether he has been abused and on whether the principal has committed crimes against other agents. The parameter $\gamma \in [0,1]$ measures his social preference, namely, the weight attached to whether the principal has committed a crime or not.

We show in Online Appendix H.1 that when L is large, agents' reports become arbitrarily uninformative and the equilibrium probability of crime approaches π^*. Interestingly, the agents' social preferences reduce the informativeness of their reports. This is because agent i's report a_i becomes more responsive to his belief about θ_j. When θ_i and θ_j are negatively correlated, a_i becomes less responsive to θ_i.

Agents' Preferences for Truth-telling: Suppose each agent receives a direct benefit d (> 0) from filing an accusation when he has witnessed a crime, regardless of the conviction decision. This can arise, for example, when agents have intrinsic preferences for telling the truth. We show in the Online Appendix H.1 that in environments with two agents and given that $d < \frac{l^*}{l^*+2}c$, the informativeness ratio of each agent's report, measured by (3.13), is bounded from above by:

$$\frac{cl^*}{cl^* - (l^*+2)d}. \tag{5.2}$$

Therefore, our findings are robust when agents receive small benefits from telling the truth.

Ex Post Evidence & Punishing False Accusations: We consider the possibility that evidence may arrive ex post, which exposes false accusations. For example, suppose that when an innocent principal is convicted, hard evidence arrives with probability p^* that reveals his innocence, causing every false accuser to be penalized by some constant $\ell \geq 0$. Our analysis is essentially unchanged, because such punishments are equivalent to an increase in the added benefit b from reporting after witnessing a crime. This extension is formally considered in Online Appendix H.1.

Uncertainty about Number of Potential Victims: In applications such as workplace bullying, physical assaults and discrimination, the number of potential victims is usually not observed by the judge and the victims. Motivated by this concern, we consider an extension of our model, in which nature randomly a subset $\widetilde{N}$ of $\{1, 2, ..., n\}$, interpreted as the set of agents the principal has opportunities to abuse. We assume that only agents in $\widetilde{N}$ can be assaulted and file reports, the latter assumption being interpreted as follows: if an agent outside of $\widetilde{N}$ files an accusation, this accusation can be easily refuted by the defendant (e.g., using an alibi). Only the principal observes $\widetilde{N}$. Agent i privately observes whether $i \in \widetilde{N}$ or not in addition to his private information in the baseline model.

We informally argue that the logic behind our results are stronger when the judge and the agents face this extra layer of uncertainty. Since the judge does not observe the size of $\widetilde{N}$, whether the principal is convicted or not depends only on the number of reports, but not on the number of potential victims. Since the principal is convicted with weakly higher probability when there are more reports, he has stronger incentives to commit assaults when fewer agents can report (that is, $|\widetilde{N}|$ is smaller). Conditional on an agent being assaulted, the agent infers that $|\widetilde{N}|$ is more likely to be small and, hence, the expected number of reports filed by other agents is also likely to be small. This effect dampens the abused agents' incentive to report and lowers the agent's credibility in equilibrium, by the same logic as in the baseline model.

Mechanical Types' Strategies: Suppose, in contrast to the baseline model, that the mechanical type agent accuses the principal with probability $\overline{\alpha}$ when he has observed a crime, and accuses the principal with probability $\underline{\alpha}$ otherwise, with $1 > \overline{\alpha} \geq \underline{\alpha} > 0$. This formulation of mechanical types incorporates strategic types who are immune to the principal's retaliation. This is because without retaliation, a strategic agent maximizes the expected value of $(\omega_i - b\theta_i)(1-s)$. In equilibrium, his reporting cutoffs are b and 0, depending on whether he has been abused or not. The conditional probabilities with which he reports are $\overline{\alpha} = \Phi(b)$ and $\underline{\alpha} = \Phi(0)$, respectively. Therefore, he behaves *as if* he is a mechanical type that plays an informative cutoff-strategy.

We extend our results to this environment in Online Appendix H.2. We show that no matter how informative the mechanical types' reports are, it will be overturned by the strategic type agent's coordination motives when δ is close to 1 and L is large. As a result, the agents' reports become arbitrarily uninformative as $L \to \infty$.

6 Conclusion

Our analysis relies on key assumptions that must be put in a broader perspective.

Equilibrium Analysis *vs.* Nonequilibrium Adjustments: Our results are derived from an equilibrium analysis. This methodology, which is ubiquitous in economics, is best suited when each player understands the rules of the game, including the payoff consequences of their actions, and other players' strategy.[27] When social rules change, as in the case of a sudden crackdown on specific crimes, the introduction of new laws and regulations, drastic shifts in social norms, or the emergence of new social media that change the social consequences of one's actions, equilibrium analysis may be viewed as a potential harbinger of issues that will emerge as economic and social actors learn to interact under these new rules.[28]

Tradeoffs in Designing Conviction Rules: Our analysis suggests several trade-offs concerning the design of conviction rules for nonverifiable crimes. First, we characterize a stark tradeoff between deterrence and fairness: while the probability of crime decreases as the conviction threshold π^* decreases, this benefit comes at the cost of increasing the fraction of innocent defendants who are convicted. This trade-off arises because when there are multiple potential victims and the conviction punishment being large, the posterior probability that a convicted principal is guilty is approximately

[27] Foundations of Nash equilibrium based on players learning one another's strategies have a long history in economics. See for example Fudenberg and Levine (1995) for an in-depth discussion.

[28] This distinction seems particularly relevant in the context of the recent *me too* movement, for which abusers before the emergence of the movement are likely to underestimate the legal and professional consequences of their abusive behavior.

π^* and the fraction of innocent people among those that are convicted is approximately $1 - \pi^*$.[29]

Importantly, we have also shown that this tradeoff is more subtle than it would seem on first thought: Proposition 4 shows that using a more lenient sentence in case of conviction can be an effective way to deter crimes, without increasing the probability of convicting the innocent. Thus, the tradeoff between deterrence and fairness depends on which instrument is considered: the conviction threshold or the sentence.

Second, suppose that the judge can *commit* to convicting the principal if at least one report is filed. Such a commitment eliminates the agents' coordination motive. When the punishment L is large, the principal has a strict incentive not to commit any crime. However, in order to fulfill his promise, the judge must convict the principal after receiving any report from the agent, even though the principal is known to be innocent with probability 1. This leads to an undesirable outcome since all convicted individuals are innocent and the probability of convicting the innocent is significant.[30]

Shielding Accusers from Stigma through Secret Accusations: To address the potential pressure that is sometimes experienced by lone accusers, institutions have been developed under which reports are submitted to a third party and are only released when enough of them have been filed.[31]

It must also be noted that, taken at face value, such institutions may protect wrongful accusers from stigma. Indeed, an agent holding a grudge against the principal has an opportunity to secretly file a report against the principal in the hope that other agents, rightfully or not, will also accuse the principal. While these institutions are clearly well intentioned and worth considering, it is also important to evaluate their long-term reliability.

In some cases, accusations may be leaked to the principal due to corruption, imperfect institutions, and other reasons. This risk is especially high if the principal is powerful and well-connected. Our results apply when such leakages occur with strictly positive probability: the acquitted principal can retaliate against the reporting agents once the information is leaked. This is because our results do

[29] This trade-off has been discussed in reduced-form by Harris (1970) and Miceli (1991). Becker (1968) and Landes (1970) ignore wrongful convictions. In Kaplow (2011), punishments are expressed in terms of fines, i.e., zero sum transfers that do not affect the social surplus. Kaplow considers the "chilling" effect of punishment on behavior. Siegel and Strulovici (2018) provide a framework that includes both deterrence and fairness considerations. Klement and Neeman (2005) study the design and cost of settlement procedures for civil cases, taking into account their effect on deterrence.

[30] This paradox induced by commitment commonly arises in plea bargaining models, in which agents who reject pleas and are found guilty at trial are known to be innocent. See Grossman and Katz (1983), Reinganum (1988), and Siegel and Strulovici (2018).

[31] In particular, the nonprofit organization Callisto has a "match" feature, whereby a report is made official only if at least two victims name the same perpetrator. See www.projectcallisto.org.

not require any conditions on the magnitude of c. They only require that first, an agent's expected loss from retaliation is strictly positive, and second, the loss from retaliation is strictly larger when the principal is acquitted compared to the case in which the principal is convicted.

Sequential Reporting: The forces that underlie our results are also present in dynamic versions of our model, in which reports may be filed sequentially. First, the negative correlation between the agents' private information (θ_i) continues to arise endogenously whenever a strategic principal is concerned about having too many reports made against him. Second, an individual agent has an incentive to coordinate with other agents whenever he is unsure about whether his report is pivotal or not. In a dynamic setting, this incentive can materialize after a *cold start* (i.e., where very few people have reported before and no agent wants to be the first accuser). It can also occur when an agent has observed many reports and is unsure of the number of reports needed to convict the principal (for example, if he faces uncertainty about the conviction standard π^* used by the judge). The inefficiencies and lack of credibility caused by the agents' coordination motives thus still arise in a dynamic environment.[32]

A Omitted Proofs

A.1 Proof of Lemma 3.1

To start with, $\alpha \in (0,1)$ and $\delta \in (0,1)$ imply that every reporting profile occurs with positive probability. For statement 1, suppose towards a contradiction that $\overline{\theta} = 0$ occurs with probability 1. Then the posterior probability of $\overline{\theta} = 1$ is 0 under every reporting profile. Since $\pi^* \in (0,1)$, the principal is acquitted for sure under every reporting profile, which gives him a strict incentive to abuse all agents. This leads to a contradiction. Next, suppose $\overline{\theta} = 1$ occurs with probability 1. Then the posterior probability of $\overline{\theta} = 1$ is 1 under every reporting profile. The evaluator's optimal strategy, given by (2.2), implies that $q(0,...,0) = 1$. This contradicts Axiom 1. For statement 2, suppose that there exists i, such that $q(1, a_{-i}) = q(0, a_{-i})$ for all $a_{-i} \in \{0,1\}^{n-1}$. Then the principal's marginal cost of abusing agent i is 0. As a result, he has a strict incentive to choose $\theta_i = 1$. This means that the prior probability that $\overline{\theta} = 1$ is 1, and the principal is convicted under every reporting profile, including $(0,0,...,0)$. This contradicts Axiom 1.

A.2 Proof of Lemma 3.2

In every symmetric equilibrium, the value of $\frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)}$ is the same for all i. We begin by showing that this ratio is strictly greater than 1. First, suppose that the ratio equals 1. Then, the principal's

[32] See Lee and Suen (2019) for a model of strategic accusation in which the timing of accusation plays a major role.

marginal cost of abusing each agent is 0, which provides him a strict incentive to abuse all agents. This contradicts Axiom 1. Next, suppose the ratio is strictly less than 1. Lemma 3.1 implies that each agent is abused with strictly positive probability. As a result, for every $\boldsymbol{a} \neq (0, 0..., 0)$, we have:

$$\Pr(\overline{\theta} = 1|\boldsymbol{a}) < \underbrace{\Pr(\overline{\theta} = 1|(0, 0, ..., 0)) \leq \pi^*}_{\text{since } q(0,...,0)=0}.$$

The above inequality implies that $q(\boldsymbol{a}) = 0$ for all $\boldsymbol{a} \in \{0, 1\}^n$, which gives the principal a strict incentive to abuse all agents. This leads to a contradiction. Since $\frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=0|\theta_i=0)} > 1$ for all i and each agent is abused with positive probability, we know that $\Pr(\overline{\theta} = 1|\boldsymbol{a}) > \Pr(\overline{\theta} = 1|\boldsymbol{a'})$, for every $\boldsymbol{a}$ and $\boldsymbol{a'} \in \{0, 1\}^n$ with $\boldsymbol{a} \succ \boldsymbol{a'}$. The evaluator's best response function implies that $q(\boldsymbol{a}) \geq q(\boldsymbol{a'})$.

A.3 Proof of Theorem 1, Statement 4

First, we show that $\omega_m^* > \omega_s^*$. Suppose towards a contradiction that $\omega_m^* \leq \omega_s^*$, then the comparison between (3.1) and (3.8) implies that

$$q_m\Big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\Big) \leq q_s.$$

Therefore,

$$q_m\Big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\Big)\Big(\Phi(\omega_s^*) - \Phi(\omega_s^{**})\Big) \leq q_s\Big(\Phi(\omega_s^*) - \Phi(\omega_s^{**})\Big)$$

$$= 1/\delta L = q_m\Big(\Phi(\omega_m^*) - \Phi(\omega_m^{**})\Big)\Big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\Big). \tag{A.1}$$

Since $\omega_m^* - \omega_m^{**} < b = \omega_s^* - \omega_s^{**}$ and $\omega_m^* \leq \omega_s^*$, we have:

$$\Phi(\omega_m^*) - \Phi(\omega_m^{**}) < \Phi(\omega_s^*) - \Phi(\omega_s^{**}). \tag{A.2}$$

Inequality (A.2) implies that

$$\Big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\Big)\Big(\Phi(\omega_s^*) - \Phi(\omega_s^{**})\Big) > \Big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\Big)\Big(\Phi(\omega_m^*) - \Phi(\omega_m^{**})\Big), \tag{A.3}$$

which contradicts (A.1). This implies that $\omega_m^* > \omega_s^*$. Moreover, according to Lemma 3.3,

$$0 < \omega_m^* - \omega_m^{**} < b = \omega_s^* - \omega_s^{**},$$

we know that $\omega_m^* > \omega_s^*$ implies $\omega_m^{**} > \omega_s^{**}$. The comparison between $\mathcal{I}_s$ and $\mathcal{I}_m$ immediately follows. This is because $\omega_m^{**} > \omega_s^{**}$ and $\omega_m^* - \omega_m^{**} < \omega_s^* - \omega_s^{**}$ imply that $\mathcal{I}_s > \mathcal{I}_m$. The comparison between π_s and π_m can then be obtained by comparing (3.4) to (3.14), which yields $\pi_s < \pi_m$.

Next, we show $q_m > q_s$. Given that $\omega_m^* > \omega_s^*$, then the comparison between (3.1) and (3.8) implies that $q_m Q_1 > q_s$. That is $1 \geq Q_1 > q_s/q_m$, which implies that $q_m > q_s$.

A.4 Principal's Incentives: Strategic Substitutes or Strategic Complements

We establish the complementarity or substitutability between the principal's choices of θ_1 and θ_2.

Lemma A.1. *In every equilibrium that satisfies Axiom 1, the principal's choices of θ_1 and θ_2 are strategic substitutes if (3.6) is positive and are strategic complements if (3.6) is negative.*

Proof. For $i \in \{1,2\}$, let $\Psi_i^* \equiv \delta\Phi(\omega_i^*) + (1-\delta)\alpha$ and let $\Psi_i^{**} \equiv \delta\Phi(\omega_i^{**}) + (1-\delta)\alpha$. Let us fix $\theta_2 = 0$, by changing θ_1 from 0 to 1, the principal increases the probability of conviction by:

$$(\Psi_1^* - \Psi_1^{**})\Big((1-\Psi_2^{**})\big(q(1,0)-q(0,0)\big) + \Psi_2^{**}\big(q(1,1)-q(0,1)\big)\Big).$$

Similarly, fix $\theta_2 = 1$, by changing θ_1 from 0 to 1, the principal increases the probability of conviction by:

$$(\Psi_1^* - \Psi_1^{**})\Big((1-\Psi_2^{*})\big(q(1,0)-q(0,0)\big) + \Psi_2^{*}\big(q(1,1)-q(0,1)\big)\Big).$$

The first expression is greater than the second one, or equivalently, the principal's choices of θ_1 and θ_2 are strategic complements, if and only if:

$$(\Psi_1^* - \Psi_1^{**})(\Psi_2^* - \Psi_2^{**})\Big(q(1,0)+q(0,1)-q(0,0)-q(1,1)\Big) > 0.$$

Since $\omega_i^* > \omega_i^{**}$ for $i \in \{1,2\}$, we know that $(\Psi_1^* - \Psi_1^{**})(\Psi_2^* - \Psi_2^{**}) > 0$. Therefore, the above inequality is equivalent to $q(1,0)+q(0,1)-q(0,0)-q(1,1) > 0$, which concludes the proof. $\square$

A.5 Proof of Theorem 2

According to Lemma 3.3 and Lemma 3.4, we only need to show that $\omega_m^* \to -\infty$ or $\omega_m^{**} \to -\infty$ as $L \to \infty$. Recall that the principal's indifference condition is given by:

$$(\delta L)^{-1} = q_m\Big(\delta\Phi(\omega_m^{**}) + (1-\delta)\alpha\Big)\Big(\Phi(\omega_m^*) - \Phi(\omega_m^{**})\Big) \tag{A.4}$$

Suppose there exist a sequence $\{L(n)\}_{n=1}^{\infty}$ and $\big\{\omega_m^*(n), \omega_m^{**}(n), q_m(n), \pi_m(n)\big\}_{n=1}^{\infty}$ such that:

1. $L(n) \geq \overline{L}$ for every $n \in \mathbb{N}$, and $\lim_{n\to\infty} L(n) = \infty$;
2. $(\omega_m^*(n), \omega_m^{**}(n), q_m(n), \pi_m(n))$ is an equilibrium when $L = L(n)$, for every $n \in \mathbb{N}$;
3. $\lim_{n\to\infty} \omega_m^{**}(n) = \omega^{**}$ for some finite $\omega^{**} \in \mathbb{R}$.

Since $\delta\Phi(\omega_m^{**}(n)) + (1-\delta)\alpha$ is bounded away from 0, (A.4) implies:

1. *either* there exists a subsequence $\{k_n\}_{n=1}^{\infty} \subset \mathbb{N}$ such that: $\lim_{n\to\infty} q_m(k_n) = 0$.
2. *and/or* there exists a subsequence $\{k_n\}_{n=1}^{\infty} \subset \mathbb{N}$ such that: $\lim_{n\to\infty}\big(\Phi(\omega_m^*(k_n)) - \Phi(\omega_m^{**}(k_n))\big) = 0$. According to requirement 3, this is equivalent to $\lim_{n\to\infty}\big(\omega_m^*(k_n) - \omega_m^{**}(k_n)\big) = 0$.

First, suppose that $\lim_{n\to\infty} q_m(k_n) = 0$ for some subsequence $\{k_n\}_{n=1}^{\infty}$. Then (3.8) and (3.9) imply that both $\omega_m^*(k_n)$ and $\omega_m^{**}(k_n)$ converge to $-\infty$. This contradicts the third requirement that the sequence $\omega_m^{**}(n)$ converges to some finite number ω^{**}.

Next, suppose that $\lim_{n\to\infty}(\omega_m^*(k_n) - \omega_m^{**}(k_n)) = 0$ for some subsequence $\{k_n\}_{n=1}^{\infty}$. Since $\omega_m^{**}(k_n)$ converges to an interior number, both $Q_0(k_n)$ and $Q_1(k_n)$ are bounded away from 0, which suggests that $Q_0(k_n)/Q_1(k_n)$ converges to 1 as $n \to \infty$. From the previous step, we know that there exists no subsequence of $\{k_n\}_{n=1}^{\infty}$ such that $q_m(k_n)$ converges to 0. That is to say, there exists $\eta > 0$ such that $q_m(k_n) \geq \eta$ for every $n \in \mathbb{N}$. Expression (3.12) then suggests that $\omega_m^*(k_n) - \omega_m^{**}(k_n)$ converges to b. This contradicts the hypothesis that $\lim_{n\to\infty} \omega_m^*(k_n) - \omega_m^{**}(k_n) = 0$.

A.6 Proof of Proposition 1

Apply the expressions of ω_s^* and ω_s^{**} as functions of q, the principal's expected cost of committing a crime is

$$q\Big(\Phi(b - c\frac{1-q}{q}) - \Phi(-c\frac{1-q}{q})\Big),$$

which is continuous in q. For every $\underline{q} \in (0,1)$, the above expression is bounded away from 0 if $q \geq \underline{q}$. As $q \to 0$, it converges to 0. In equilibrium, the value of this expression is $1/\delta L$. This implies the existence of equilibrium that satisfies Axiom 1 when

$$\max_{q\in[0,1]} q\big(\Phi(b - c\frac{1-q}{q}) - \Phi(-c\frac{1-q}{q})\big) \geq \frac{1}{\delta L}. \tag{A.5}$$

When L is large enough, $\omega_s^*, \omega_s^{**} < \mu$, the mean of the normal distribution, and the equilibrium is unique.

When $L \to \infty$ while holding c constant, $1/\delta L$ converges to 0. Suppose towards a contradiction that q_s converges to some strictly positive number $\underline{q}$ along some sequence $\{L_n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} L_n = \infty$. Then ω_s^* and ω_s^{**} converge to $b - c(1-\underline{q})/\underline{q}$ and $-c(1-\underline{q})/\underline{q}$, respectively. The LHS of (3.4) converges to

$$\delta\underline{q}\Big(\Phi(b - c\frac{1-\underline{q}}{\underline{q}}) - \Phi(-c\frac{1-\underline{q}}{\underline{q}})\Big) \tag{A.6}$$

which is strictly bounded away from 0. This leads to a contradiction and establishes that $q_s \to 0$. The expressions for ω_s^* and ω_s^{**} in (3.1) and (3.2) imply that both cutoffs converge to $-\infty$. In the limiting economy in which $\delta \to 1$, we have:

$$\lim_{\omega_s^*\to-\infty} \lim_{\delta\to 1} \frac{\delta\Phi(\omega_s^*) + (1-\delta)\alpha}{\delta\Phi(\omega_s^* - b) + (1-\delta)\alpha} = \infty. \tag{A.7}$$

The above equation makes use of the observation that the tail events of normal distributions are arbitrarily informative, or formally, $\lim_{\omega\to-\infty} \Phi(\omega)/\Phi(\omega - b) \to \infty$ for every $b > 0$. That is to say, it extends to all distributions of ω with thin left tails. As a result, the equilibrium probability of crime π_s vanishes to 0.

References

[1] Ali, S. Nageeb, Maximilian Mihm and Lucas Siga (2018) "Adverse Selection in Distributive Politics," Working Paper.

[2] Ambrus, Attila, and Satoru Takahashi (2008) "Multi-sender Cheap Talk with Restricted State Spaces,"*Theoretical Economics*, 3, 1-27.

[3] Austen-Smith, David and Jeffrey Banks (1996) "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *American Political Science Review*, 90(1), 34-45.

[4] Ba, Bocar (2018) "Going the Extra Mile: The Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago," Working paper

[5] Ba, Bocar and Roman Rivera (2019) "The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago," Working paper.

[6] Baliga, Sandeep, Ethan Bueno de Mesquita and Alexander Wolitzky (2019) "Deterrence with Imperfect Attribution," Working Paper.

[7] Banerjee, Abhijit (1992) "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107(3), 797-817.

[8] Battaglini, Marco (2002) "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica*, 70(4), 1379-1401.

[9] Battaglini, Marco (2017) "Public Protests and Policy Making," *Quarterly Journal of Economics*, 132(1), 485-549.

[10] Becker, Gary (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2), 169-217.

[11] Bhattacharya, Sourav (2013) "Preference Monotonicity and Information Aggregation in Elections," *Econometrica*, 81(3), 1229-1247.

[12] Chamberlin, John (1974) "Provision of Collective Goods As a Function of Group Size," *The American Political Science Review*, 68(2), 707-716.

[13] Chassang, Sylvain and Gerard Padró i Miquel (2018) "Corruption, Intimidation and Whistle-Blowing: A Theory of Inference from Unverifiable Reports," NBER working paper.

[14] CIUS (1996) "Crime Index Offenses Reported."

[15] Crémer, Jacques and Richard McLean (1988) "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56(6), 1247-1257.

[16] Dresher, Melvin (1962) "A Sampling Inspection Problem in Arms Control Agreements – A Game-Theoretic Analysis," Memorandum, The RAND Corporation, Santa Monica, California.

[17] Ekmekci, Mehmet and Stephan Lauermann (2019) "Informal Elections with Dispersed Information," Working Paper.

[18] Feddersen, Timothy and Wolfgang Pesendorfer (1998) "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting," *American Political Science Review*, 92(1), 23-35.

[19] Fudenberg, Drew and David Levine (1995) "The Theory of Learning in Games," MIT Press.

[20] Grossman, Gene and Michael Katz (1983) "Plea Bargaining and Social Welfare," *American Economic Review*, 73(4), 749-757.

[21] Harris, John (1970) "On the Economics of Law and Order," *Journal of Political Economy*, 78(1), 165-174.

[22] Kaplow, Louis (2011) "On the Optimal Burden of Proof," *Journal of Political Economy*, 119(6), 1104-1140.

[23] Klement, Alon, and Zvika Neeman (2005) "Against Compromise: A Mechanism Design Approach," *Journal of Law, Economics, and Organization*, 21(2), 285-314.

[24] Landes, William (1971) "An Economic Analysis of the Courts," *Journal of Law and Economics*, 14(1), 61-107.

[25] Lee, Frances Xu and Wing Suen (2019) "Credibility of Crime Allegations," *American Economic Journal-Microeconomics*, forthcoming.

[26] Miceli, Thomas (1991) "Optimal Criminal Procedure: Fairness and Deterrence," *International Review of Law and Economics* 11(1), 3-10.

[27] Morgan, John and Phillip Stocken (2008) "Information Aggregation in Polls," *American Economic Review*, 98(3), 864-896.

[28] Morgan, Rachel and Grace Kena (2016) "Criminal Victimization, 2016," Bulletin, Bureau of Justice Statistics.

[29] Myerson, Roger (1978) "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, 7(2), 73-80.

[30] Ottaviani, Marco and Peter Norman Sørensen (2000) "Herd Behavior and Investment: Comment," *American Economic Review*, 90(3), 695-704.

[31] Persico, Nicola (2004) "Committee Design with Endogenous Information," *Review of Economic Studies*, 70(1), 1-27.

[32] RAND Cooperation (2018) "Sexual Assault and Sexual Harassment in the US Military," Technical Report.

[33] Reinganum, Jennifer (1988) "Plea Bargaining and Prosecutorial Discretion," *American Economic Review*, 78(4), 713-728.

[34] Scharfstein, David and Jeremy Stein (1990) "Herd Behavior and Investment," *Amercian Economic Review*, 80(3), 465-479.

[35] Schmitz, Patrick and Thomas Tröger (2011) "The Sub-Optimality of the Majority Rule," *Games and Economic Behavior*, 651-665.

[36] Siegel, Ron and Bruno Strulovici (2018) "Judicial Mechanism Design," Working Paper.

[37] Silva, Francesco (2018) "If We Confess Our Sins," Working Paper.

[38] Smith, Lones and Peter Norman Sørensen (2000) "Pathological Outcomes of Observational Learning," *Econometrica*, 68(2), 371-398.

[39] Stigler, George (1970) "The Optimal Enforcement of Laws," *Journal of Political Economy*, 78(3), 526-536.

[40] Strulovici, Bruno (2018) "Can Society Learn from Anethical Agents? A Theory of Mediated Learning with Information Attrition," Working Paper.

[41] U.S. Equal Employment Opportunity Commission (2017) "Fiscal Year 2017 Enforcement And Litigation Data," Research Brief.

[42] USMSPB (2018) "Update on Sexual Harassment in the Federal Workplace," Research Brief.