Crime Aggregation, Deterrence, and Witness Credibility*

Harry Pei Bruno Strulovici

March 10, 2021

Abstract

We present a model of criminal behavior and information aggregation in which the incentives to commit and to report crime are endogenous. An individual has several opportunities to commit crime, each of which associated with a witness with private reporting preferences and retaliation risk. We study how the mechanism used to map witness testimonies into verdicts affects criminal behavior and witness credibility when the punishment in case of conviction is large relative to the benefit of committing crime. We show that convicting defendants based on the probability that they have committed at least one crime reduces the maximal number of crimes but increases expected crime frequency and undermines the informativeness of witness testimonies. We characterize mechanisms that minimize the expected number of crimes subject to an upper bound on the fraction of wrongful convictions. The optimum is always attained by one of two aggregation rules discussed in legal scholarship.

Keywords: Information Aggregation, Soft Evidence, Deterrence, Adjudication Rule.

JEL Codes: D82, D83, K42.

*We thank S. Nageeb Ali, Bocar Ba, Arjada Bardhi, Laura Doval, Mehmet Ekmekci, Alex Frankel, Yingni Guo, Marina Halac, Andreas Kleiner, Anton Kolotinin, Frances Xu Lee, Alessandro Pavan, Joao Ramos, Joyce Sadka, Ron Siegel, Vasiliki Skreta, Kathryn Spier, Takuo Sugaya, Satoru Takahashi, Teck Yong Tan, Matthew Thomas, Maren Vairo, Alex Wolitzky, Siyang Xiong, and Boli Xu for helpful comments, and the National Science Foundation grants SES-1151410 and SES-1947021 for financial support.

1 Introduction

When a defendant faces multiple charges, the legal norm is to consider these charges separately and to convict the defendant if there is at least one specific charge whose corresponding evidence meets the appropriate standard of proof. However, the desirability of this norm for deterrence and fairness is not a priori obvious. For example, consider a defendant to whom a judge assigns probability 0.8 of committing each of two crimes, independently of each other. If the conviction threshold for each offense is 0.9, the defendant must be acquitted on both counts, even though the probability that he is guilty of *at least one* offense is $1 - 0.2 \times 0.2 = 0.96$. By contrast, a defendant accused of a single offense may be convicted even if his probability of guilt is 0.91, and thus lower than the first defendant's.

Starting with Cohen (1977) and Bar Hillel (1984), legal scholarship has explored the possibility of aggregating charges into an overall probability of guilt instead of treating charges separately. Harel and Porat (2009) define the *Aggregate Probabilities Principle* ("APP") as a rule requiring that a defendant be convicted if the probability that he has committed some *unspecified* offense exceeds a given threshold. This principle has been used in laws concerning organized crime (Lynch 1987), pretrial detentions (Dobbie et al 2018), and is routinely used in firms and organizations, for instance, to fire an individual facing multiple accusations of wrongdoing. Compared to the commonly used *Distinct Probabilities Principle* ("DPP"), which requires that a defendant be convicted if the probability that he has committed some *specific* offense exceeds a pre-specified threshold, Harel and Porat argue that APP can reduce adjudication errors, improve deterrence, and reduce the cost of enforcement. They advocate its use to adjudicate both civil and criminal cases. Similar arguments appear in Schauer and Zeckhauser (1996), who observe that "*although sound reasons for the criminal law's refusal to cumulate multiple low-probability accusations exist, the reasons for such refusal are often inapt in other settings ...taking adverse decisions based on cumulating multiple low-probability charges is often justifiable both morally and mathematically.*"

Legal studies comparing APP to DPP have assumed that the strength of each accusation against a defendant is exogenous and that the defendant's guilt concerning each offense is independently distributed across offenses. Crucially, they ignore how different adjudication rules (such as APP vs. DPP) affect the incentives of potential offenders to commit crimes and the incentives of potential witnesses to report crimes.

We study a model in which adjudication rules affect the incentives of potential offenders and witnesses. We show that APP can undermine the informativeness of witness testimonies and provide a rationale for not using it in most civil and criminal lawsuits. We also examine the effectiveness of APP and DPP for deterring crimes. Depending on the parameters of the model, either APP or DPP minimizes the expected number of

crimes subject to an upper bound on the fraction of wrongful convictions.

In our model, a potential offender (hereafter, *principal*) has two opportunities to commit crime,[1] and trades off the benefit of committing either or both crimes with the expected punishment that this entails. Each crime is associated with a distinct witness (hereafter, *agent*) who observes whether that crime takes place. The agent could be a mere observer, who may become a whistleblower, or he could be a direct victim of the crime. For example, the principal could be a firm manager who has multiple opportunities to violate the law or abuse his subordinates, and the agents could be whistleblowers or victims who witness the violations. Each agent decides whether to accuse the principal based on three considerations: (1) a preference for punishing offenders, (2) a personal cost of accusing the principal, which is greater when the principal is acquitted than when the principal is convicted, and (3) some idiosyncratic preference for getting the principal convicted or acquitted. The accusation cost may be interpreted as retaliation by the principal against accusers or a stigma that accusers often experience, especially if accusations against the principal are not deemed credible enough to warrant a conviction.[2] Finally, a judge observes agents' reports and decides whether to convict the principal.

We focus on situations in which the magnitude of the *realized* punishment to the principal, if he is convicted, is large relative to the benefit from committing crime. This assumption, made for tractability, seems realistic for offenses whose gratification is short-lived or financially small relative to large punitive damages or to the large reputation and career damages that come with a conviction. Nevertheless, the *expected* punishment may and typically will be much smaller and commensurate with the benefit.

We begin our analysis by adopting a mechanism design perspective. A designer commits to a mechanism that maps agents' reports to a probability of conviction.[3] Theorem 1 characterizes the lowest expected number of crimes over all monotone mechanisms when the designer faces an upper bound on the fraction of wrongful convictions.[4] We show that when the designer can tolerate a high fraction of wrongful convictions, it is optimal to set of the probability of conviction to be linear in the number of accusations. Otherwise, it is optimal to convict the principal only if both agents accuse him.

[1] In a working paper (Pei and Strulovici (2020)), which this paper subsumes, we study a related model that allows for an arbitrary number of crime opportunities and heterogeneity of principal's propensity from committing crime. The main insights are unchanged, but the paper considers two specific rules rather than a general mechanism design problem.

[2] In our model, some offenses go unreported and some charges are not deemed credible enough to lead to a conviction. These patterns are consistent with the studies on police brutality and inaction by Ba (2018) and Ba and Rivera (2019). Similar patterns are documented by a survey conducted by the USMSPB, which concluded that 21% of women and 8.7% of men working as federal employees experienced at least one of 12 categorized behaviors of sexual harassment, of which only 16% led to merit resolutions.

[3] As noted, the punishment from conviction is exogenous and large relative to the benefit from the crime. The probability of conviction, however, is endogenous and is a key variable of the analysis.

[4] A mechanism is monotone if the probability of conviction is *weakly* increasing in the set of agents who accuse the principal. The qualitative features of the optimal mechanism remain unchanged if we also impose a constraint on the fraction of mistaken acquittals, as explained in Section 5.

Intuitively, when conviction against the principal requires that both agents accuse him (this is the "convex" case, as the probability of conviction is convex in the number of accusations), the principal's decisions to commit each of these crimes are *strategic substitutes*. As a result, the principal commits at most one crime in equilibrium. While desirable, this feature also implies that the agents' observations are *negatively correlated*, since one agent observes a crime only if the other does not. Given that each agent faces a lower retaliation cost when the principal is convicted, agents' decisions to accuse the principal are *strategic complements*. This combination of coordination motive and negative correlation reduces an agent's willingness to accuse the principal when he has witnessed a crime, and vice versa. This lowers the credibility of agents' reports and, in equilibrium, increases the probability that the principal commits crime.

By contrast, if the probability of conviction is linear in the number of accusations, the principal's decisions to commit crime are *uncorrelated* across crime. The principal commits multiple crimes with positive probability, but agents' reports are more informative.[5]

In summary, the designer faces a tradeoff between reducing the probability that the principal commits multiple crimes and reducing the probability that he commits at least one crime. The mechanism that optimally resolves this trade-off depends on two effects: (1) the effectiveness of linear conviction probabilities in improving the informativeness of accusations and (2) the effectiveness of this improved informativeness in reducing the probability of crime. We show that linear conviction probabilities are optimal when the fraction of agents who have an incentive to make false accusations is low and the designer has a high tolerance for wrongful convictions, and vice versa.

Next, we analyze equilibrium outcomes when conviction is based on the posterior probability that the defendant is guilty given accusations that are leveled against him. We consider two rules: (APP) the defendant is convicted if the probability that he has committed at least one unspecified crime exceeds some threshold; (DPP) the defendant is convicted if the probability that he has committed a specific crime exceeds some threshold.

Theorems 2 and 3 show that the optimal outcome in the mechanism design problem is attained in every equilibrium under APP when convex conviction probabilities are optimal, and is attained in every equilibrium under DPP when linear conviction probabilities are optimal. These findings provide a justification for using APP and DPP from the perspective of crime deterrence. APP leads to undesirable outcomes from a fairness perspective since the probability that an innocent individual being convicted is arbitrarily close to the probability that a guilty individual being convicted. This is not the case of DPP, where the difference in

[5]We also show that setting the conviction probability to be a concave function in the number of reports does not improve on the linear case. Intuitively, agents' reports in the concave case are about as informative as in the linear case, but the probability that the principal commits multiple crimes is significantly higher in the concave case.

conviction probabilities between a guilty and an innocent individual is bounded away from 0.

Our paper contributes to the law and economics literature by examining the interaction between (i) the incentives to commit crimes and (ii) the incentives to report crimes and (iii) the rules used to aggregate information and incorporate it into a judicial decision. This approach stands in contrast to and complements several recent papers that focus on witnesses' incentives to report crimes, such as Lee and Suen (2020), Cheng and Hsiaw (2020), and Naess (2020), and to Siegel and Strulovici (2020), in which a mechanism designer can modify the adjudication rule without affecting the quality of witness testimonies. Silva (2019) and Baliga, Bueno de Mesquita and Wolitzky (2020) consider the case of multiple potential offenders. These works do not consider witnesses' incentives. They assume that defendants' guilt is negatively correlated across defendants whereas this property arises endogenously in our setting.

Our results also pertain to the attainability of the optimal commitment outcome when judges adjudicate guilt based on her posterior belief about crime after observing evidence (e.g., witness testimonies). These findings bridge a gap between two existing approaches to study crime deterrence: the mechanism design approach in Silva (2019), Siegel and Strulovici (2020), and Naess (2020), and the game theoretic approach in Baliga, Bueno de Mesquita and Wolitzky (2020), Lee and Suen (2020), and Cheng and Hsiaw (2020). The details of these papers as well as the connections between our work and the existing literature on voting, communication, and coordination games are described further in Section 6.

2 Baseline Model

We study a three-stage game between a potential criminal (the *principal*), two potential witnesses or victims (the *agents*), and a judge who is either a designer who can commit to a mechanism or one who uses a predefined conviction rule based on her posterior belief. In stage 1, the principal chooses $\boldsymbol{\theta} \equiv (\theta_1, \theta_2) \in \{0,1\}^2$, where $\theta_i = 1$ means that the principal commits the crime witnessed by agent i.

In stage 2, agent $i \in \{1,2\}$ decides whether to accuse the principal ($a_i = 1$) or not ($a_i = 0$) after privately observing $\theta_i \in \{0,1\}$ and two idiosyncratic preference parameters: $c_i \in [0, \overline{c}]$, and $\omega_i \in \mathbb{R}$. The parameter c_i is agent i's cost of accusing the principal. This cost may be the result of retaliation by the principal or of social stigma. The parameter ω_i captures i's preference (such as a grudge or affinity) for or against convicting the principal. We let Φ denote the cdf of ω_1 and ω_2 and F denote the cdf of c_1 and c_2. We assume that ω_1, ω_2, c_1 and c_2 are independently distributed, that $\overline{c} > 0$, and that Φ and F both have full support on their respective domains and have continuous density functions ϕ and f. We also assume that the density f is large enough at 0. This technical assumption will be used to establish the existence of a

nontrivial equilibrium in Proposition 1:[6]

$$\sup_{\alpha\in\mathbb{R}}\Big\{\alpha\Phi(-\alpha)\Big\}f(0)\geq 1. \tag{2.1}$$

In stage 3, the designer or judge observes $\boldsymbol{a}\equiv(a_1,a_2)\in\{0,1\}^2$ and chooses $s\in\{0,1\}$, where $s=1$ stands for convicting the principal and $s=0$ stands for acquitting him.

The principal's payoff is $y\sum_{i=1}^{2}\theta_i-s$. Thus, the principal receives a benefit $y>0$ from committing each crime and a punishment normalized to 1 when he is convicted. Agent i's payoff is

$$s(\theta_i-\gamma c_i a_i-\omega_i)+(1-s)(-c_i a_i) \tag{2.2}$$

where $\gamma\in(0,1)$.[7] Thus, agent i's payoff is $\theta_i-\gamma c_i a_i-\omega_i$ when the principal is convicted (i.e., $s=1$) and is $-c_i a_i$ when the principal is acquitted (i.e., $s=0$). This implies that (1) each agent has a stronger incentive to convict the principal when he has witnessed a crime (i.e., $\theta_i=1$) and when his payoff shock ω_i is lower;[8] (2) since $\gamma<1$, an agent's cost of accusing the principal is higher if the principal is acquitted than if he is convicted. This is consistent with our interpretation that c_i is a retaliation cost at the hand of the principal or a stigma experienced from making an accusation.

Section 3 analyzes a setting in which a designer commits to a mapping from agents' reports to the probability of conviction $q:\{0,1\}^2\to[0,1]$.[9] Her objective is to minimize the expected number of crimes $\mathbb{E}[\theta_1+\theta_2]$ subject to the constraint that the fraction of wrongful convictions, $\Pr(\boldsymbol{\theta}=(0,0)|s=1)$, does not exceed $\overline{\pi}\in(0,1)$, where $\overline{\pi}$ measures the designer's tolerance toward wrongful convictions.[10]

Section 4 analyzes a setting in which the judge makes conviction decisions based on one of two adjudication rules (APP or DPP) and on her posterior belief about the principal's actions. The first rule (APP) requires that the judge convict the principal when the probability that he is guilty of *at least one unspecified crime* exceeds some threshold. The second rule, DPP, requires that the judge convict the principal when the probability that he is guilty of a *particular* crime exceeds some threshold.

[6] Starting from any arbitrary distribution F, (2.1) can be achieved by reallocating an arbitrarily small mass of the distribution to a right neighborhood of 0. Since $\sup_\alpha \alpha\Phi(-\alpha)$ is strictly positive for any Φ that puts positive weight on negative realizations, it suffices to "pinch" the distribution F a bit near 0 to guarantee that $f(0)$ exceeds $1/\sup_\alpha \alpha\Phi(-\alpha)$.

[7] When $\gamma=1$, an agent's cost from accusing the principal is independent of whether the principal is convicted, and agents have no incentive to coordinate their reports. When $\gamma=0$, the agent may not use a cutoff strategy in equilibrium, see inequality (B.2) on page 26.

[8] The extent to which agent i's preference depends on his private observation of crime, θ_i, is not stochastic. Our main insights extend when the coefficient in front of θ_i, denoted by b_i, is also stochastic, as long as (b_1,c_1,ω_1) is independent of (b_2,c_2,ω_2).

[9] As noted in the Introduction, the punishment from a conviction is taken as exogenous, although the probability of conviction, and hence the principal's expected punishment, is endogenous.

[10] Section 5 considers the case of mistaken acquittals, as mentioned below.

In Section 5, we argue that the designer cannot improve upon the optimal mechanism of Section 3 by eliciting agents' private information about $(\omega_1, \omega_2, c_1, c_2, \theta_1, \theta_2)$. We also consider settings in which the designer also faces a constraint on the fraction of mistaken acquittals $\Pr(\boldsymbol{\theta} \neq (0,0)|s=0)$ and in which agents' preferences depend on the crimes committed against (or observed by) other agents.

3 Optimal Commitment Outcomes

We restrict our attention to mechanisms for which accusations *weakly increase* the probability of conviction.

Definition. $q : \{0,1\}^2 \to [0,1]$ *is monotone if* $q(1, a_{-i}) \geq q(0, a_{-i})$ *for every* i *and* a_{-i}.

Monotonicity implies that each accusation is a move against the principal. This property is consistent with the assumption that accusers incur a retaliation cost from the principal, unlike non-accusers.

A Bayes Nash equilibrium is a strategy profile $\big(\sigma_p, \sigma_1, \sigma_2\big)$, in which the principal's strategy $\sigma_p \in \Delta\big(\{0,1\}^2\big)$ is a distribution of (θ_1, θ_2), and agent $i \in \{1,2\}$'s strategy $\sigma_i : \mathbb{R} \times [0, \overline{c}] \times \{0,1\} \to [0,1]$ maps ω_i, c_i, and θ_i to the probability of accusing the principal, namely, choosing $a_i = 1$.

A mechanism and a strategy profile induce a distribution of $(\boldsymbol{\theta}, \boldsymbol{a}, s)$, which we call an *outcome*. Let $\boldsymbol{\psi}(\boldsymbol{a}) \in \Delta\big(\{0,1\}^2\big)$ denote the distribution of (θ_1, θ_2) conditional on $\boldsymbol{a} \equiv (a_1, a_2)$. This distribution is uniquely defined by Bayes rule for every $\boldsymbol{a}$ that occurs with positive probability under that outcome. If $\boldsymbol{a}$ occurs with zero probability, we allow $\boldsymbol{\psi}(\boldsymbol{a})$ to take any arbitrary value.

Definition. *An outcome is* $\overline{\pi}$***-valid*** *if*

1. ***Upper Bound on the Fraction of Wrongful Convictions:*** $\Pr(\boldsymbol{\theta} = (0,0)|s=1) \leq \overline{\pi}$.
2. ***Independence of Uninformative Messages:*** *For every* $\boldsymbol{a}, \boldsymbol{a'}$*, if* $\boldsymbol{\psi}(\boldsymbol{a}) = \boldsymbol{\psi}(\boldsymbol{a'})$*, then* $q(\boldsymbol{a}) = q(\boldsymbol{a'})$.

The first part of this definition imposes a constraint on the fraction of wrongful convictions. The second part requires that whenever two message profiles lead to the same posterior belief about the defendant's actions, they must imply the same probability of conviction. This second requirement is motivated by a legal and normative perspective: conviction decisions should be independent of signals that are orthogonal to the defendant's guilt. It rules out equilibria in which the principal never commits a crime observed by agent i, agent i files an accusation with positive probability, and i's report, while completely uninformative of the principal's guilt concerning any crime, affects the principal's probability of conviction.

Let

$$R \equiv \begin{cases} \frac{\int_{-\infty}^{1} \Phi(\omega) d\omega}{\int_{-\infty}^{0} \Phi(\omega) d\omega} & \text{if} \quad \int_{-\infty}^{0} \Phi(\omega) d\omega < +\infty \\ 1 & \text{if} \quad \int_{-\infty}^{0} \Phi(\omega) d\omega = +\infty, \end{cases} \tag{3.1}$$

The parameter R, which plays an important role in the analysis, may be viewed as a structural truthfulness index, based on the primitives of the model. Intuitively, an agent has an incentive to falsely accuse the principal only if $\omega \leq 0$. According to (3.1), R is greater when the distribution of ω has a thinner left tail, i.e., when agents are less likely to make false accusations. Let

$$\pi_{\min}(\overline{\pi}) \equiv \frac{1}{1+\bar{l}}\Big(2\bar{l} + R + 1 - \sqrt{(R+1)^2 + 4R\bar{l}}\Big) \quad \text{where} \quad \bar{l} \equiv \frac{1-\overline{\pi}}{\overline{\pi}}. \tag{3.2}$$

It is readily checked that

$$\pi_{\min}(\overline{\pi}) < 1 - \overline{\pi} \quad \text{if and only if} \quad R > \frac{\bar{l}}{2} + 1. \tag{3.3}$$

Theorem 1 characterizes the lowest expected number of crimes among all $\overline{\pi}$-valid outcomes that can be implemented by monotone mechanisms when the benefit from committing crime y is small relative to the punishment from conviction.

Theorem 1. *For every $\varepsilon > 0$, there exists $\overline{y}_\varepsilon > 0$ such that when $y \in (0, \overline{y}_\varepsilon)$, $\mathbb{E}[\theta_1 + \theta_2] \geq \min\{1 - \overline{\pi}, \pi_{\min}(\overline{\pi})\} - \varepsilon$ for every $\overline{\pi}$-valid outcome that can be implemented by monotone mechanisms.*

1. *If $R \leq \frac{\bar{l}}{2} + 1$, there exists a $\overline{\pi}$-valid outcome with $\mathbb{E}[\theta_1 + \theta_2] \leq 1 - \overline{\pi}$ that can be implemented by a monotone mechanism satisfying $q(1,1) \in (0,1)$ and $q(1,0) = q(0,1) = q(0,0) = 0$.*

2. *If $R > \frac{\bar{l}}{2} + 1$, there exists a $\overline{\pi}$-valid outcome with $\mathbb{E}[\theta_1 + \theta_2] \leq \pi_{\min}(\overline{\pi})$ that can be implemented by a monotone mechanism satisfying $q(1,1) = 2q(1,0) = 2q(0,1) > 0$ and $q(0,0) = 0$.*

Theorem 1 implies that when the principal's benefit from committing crime, y, is small, the expected number of crimes is close to $\min\{1 - \overline{\pi}, \pi_{\min}(\overline{\pi})\}$ under the optimal mechanism. Moreover, this outcome can be achieved by a mechanism that lies in one of two classes: (1) *unanimous mechanisms*, which convict the principal with positive probability only if both agents accuse him, (2) *linear mechanisms*, for which the probability of conviction is linear in the number of accusations.

According to Theorem 1, it is optimal to use linear mechanisms when agents are unlikely to make false accusations and the constraint on wrongful convictions is less stringent, i.e., when R and $\overline{\pi}$ are large. Otherwise, it is optimal to use unanimous mechanisms.

The intuition for this result is as follows. Compared to linear mechanisms, unanimous mechanisms eliminate the possibility that the principal commits multiple crimes but result in a higher probability that he commits at least one crime. The type of mechanism that minimizes the expected number of crimes depends on the extent to which linear mechanisms can reduce the probability that the principal commits at least one

crime. This, in turn, depends on the mechanism's effectiveness in improving the informativeness of agents' accusations and the extent to which improved informativeness can lower the probability of crime. A larger R improves the informativeness of accusations under linear mechanisms.[11] When $\overline{\pi}$ increases, the designer can set a lower standard of proof, and improved informativeness will have a larger impact on the equilibrium probability of crime.

The proof of Theorem 1 is in Appendix D. We provide an intuitive explanation in three steps. First, let

$$Q \equiv q(1,1) + q(0,0) - q(1,0) - q(0,1). \tag{3.4}$$

When $Q > 0$, the principal's decisions are strategic substitutes. Since the principal commits no crime with positive probability, he commits two crimes with zero probability. When $Q \leq 0$, the principal's decisions to commit different crimes are strategic complements and he may commit two crimes with positive probability.

Next, we consider equilibrium outcomes when $Q > 0$. Since the principal commits multiple crimes with zero probability, agents' private observations of crimes are *negatively correlated*. Since q must be monotone and each agent faces a lower retaliation cost when the principal is convicted, agents' decisions to accuse the principal are *strategic complements*. Given that each agent accuses the principal with higher probability when he has witnessed a crime, his incentive to coordinate with the other agent discourages him to accuse the principal when he has witnessed a crime and vice versa. This reduces the value of

$$\mathcal{I}_i \equiv \frac{\Pr(a_i = 1|\theta_i = 1)}{\Pr(a_i = 1|\theta_i = 0)}. \tag{3.5}$$

The above likelihood ratio measures the *informativeness* of agent i's accusation since according to Bayes rule,

$$\frac{\Pr(\theta_i = 1)}{1 - \Pr(\theta_i = 1)}\mathcal{I}_i = \frac{\Pr(\theta_i = 1|a_i = 1)}{1 - \Pr(\theta_i = 1|a_i = 1)},$$

where $\frac{\Pr(\theta_i=1)}{1-\Pr(\theta_i=1)}$ is the prior likelihood ratio of crime against agent i, $\frac{\Pr(\theta_i=1|a_i=1)}{1-\Pr(\theta_i=1|a_i=1)}$ is the posterior likelihood ratio after observing agent i's accusation, and hence, $\mathcal{I}_i$ measures the change in the likelihood ratio of crime after observing agent i's accusation.

In fact, we show that $\mathcal{I}_i$ is close to 1 when y is close to 0, meaning that the informativeness of each agent's accusation is arbitrarily low. Since the fraction of wrongful convictions cannot exceed $\overline{\pi}$, the probability of crime is close to $1 - \overline{\pi}$ when $\mathcal{I}_1$ and $\mathcal{I}_2$ are close to 1.[12]

[11] By contrast, the informativeness of agents' reports is low under unanimous mechanisms regardless of R, as explained below.

[12] Similarly, one can show that in any $\overline{\pi}$-valid outcome where θ_1 and θ_2 are negatively correlated, the probability that the principal commits at least one crime is close to $1 - \overline{\pi}$ when y is close to 0.

Next, we consider $\overline{\pi}$-valid outcomes where θ_1 and θ_2 are *uncorrelated*, which can only be implemented by mechanisms with $Q = 0$, i.e., the principal's actions are neither complements nor substitutes. In this case, the principal commits multiple crimes with positive probability, but the probability that he commits at least one crime is lower than in the negatively correlated case since agents' coordination motive no longer undermines the informativeness of their accusations. In particular, the informativeness ratio $\mathcal{I}_i$ (defined in (3.5)) is approximately R, which in general, is strictly greater than 1.

We explain why $\mathcal{I}_i$ is close to R when θ_1 and θ_2 are uncorrelated, y is close to 0, and Φ has a thin left tail in the sense that $\int_{-\infty}^{0} \Phi(x)dx < +\infty$. According to (2.2), each agent's equilibrium strategy can be characterized by two linear functions, $\omega_i^*(c)$ and $\omega_i^{**}(c)$, such that when $c_i = c$, agent i accuses the principal if $\omega_i \leq \omega_i^*(c)$ and $\theta_i = 1$, or if $\omega_i \leq \omega_i^{**}(c)$ and $\theta_i = 0$. The intercepts of $\omega_i^*(\cdot)$ and $\omega_i^{**}(\cdot)$ are 1 and 0, respectively. Let $\omega_i^*(c_i) = 1 - c_i K_i^*$ and $\omega_i^{**}(c_i) = -c_i K_i^{**}$. When θ_1 and θ_2 are uncorrelated, $K_i^* = K_i^{**}$. The definition of $\mathcal{I}_i$ in (3.5) implies that

$$\mathcal{I}_i \equiv \frac{\Pr(a_i = 1|\theta_i = 1)}{\Pr(a_i = 1|\theta_i = 0)} = \frac{\int_0^{\overline{c}} \Phi(\omega_i^*(c))dF(c)}{\int_0^{\overline{c}} \Phi(\omega_i^{**}(c))dF(c)} = \frac{K_i^{**}}{K_i^*} \cdot \frac{\int_{-\infty}^{1} f(\frac{1-x}{K_i^*})\Phi(x)dx}{\int_{-\infty}^{0} f(\frac{-x}{K_i^{**}})\Phi(x)dx} = \frac{\int_{-\infty}^{1} f(\frac{1-x}{K_i^*})\Phi(x)dx}{\int_{-\infty}^{0} f(\frac{-x}{K_i^{**}})\Phi(x)dx}.$$

Since the principal commits crime with positive probability for all values of y, the probability of conviction converges to 0 as y goes to 0. When each agent's accusation has an arbitrarily small effect on the probability of conviction, the probability that he files an accusation converges to 0 which means that K_i^* and K_i^{**} diverge to $+\infty$. When $\int_{-\infty}^{0} \Phi(x)dx$ is finite, the dominated convergence theorem implies that

$$\lim_{K_i^* \to +\infty} \int_{-\infty}^{1} f\left(\frac{1-x}{K_i^*}\right)\Phi(x)dx = \int_{-\infty}^{1} \lim_{K_i^* \to +\infty} f\left(\frac{1-x}{K_i^*}\right)\Phi(x)dx = \lim_{c \downarrow 0} f(c) \int_{-\infty}^{1} \Phi(x)dx,$$

$$\lim_{K_i^{**} \to +\infty} \int_{-\infty}^{0} f\left(\frac{-x}{K_i^{**}}\right)\Phi(x)dx = \int_{-\infty}^{0} \lim_{K_i^{**} \to +\infty} f\left(\frac{-x}{K_i^{**}}\right)\Phi(x)dx = \lim_{c \downarrow 0} f(c) \int_{-\infty}^{0} \Phi(x)dx.$$

These two equations and the expression of $\mathcal{I}_i$ imply that both $\mathcal{I}_1$ and $\mathcal{I}_2$ converge to R when $y \to 0$.

To compute the expected number of crimes and the fraction of wrongful convictions, we observe that any q that (i) satisfies $Q = 0$ and (ii) implements a $\overline{\pi}$-valid outcome, must be symmetric in the sense that $q(1,0) - q(0,0) = q(0,1) - q(0,0)$. Suppose by way of contradiction that $q(1,0) - q(0,0) > q(0,1) - q(0,0)$, $Q = 0$ implies that $q(1,1) - q(0,1) = q(1,0) - q(0,0)$ and $q(1,1) - q(1,0) = q(0,1) - q(0,0)$, so the principal's cost of committing crime against agent 1 is strictly greater than that against agent 2. In equilibrium, the principal commits crime against agent 1 with zero probability, yet the conviction probabilities are responsive to agent 1's report. This violates our requirement that the conviction probabilities must be independent of

uninformative messages. One can also use this requirement to show that the principal commits crime against each agent with the same probability.

Let $\widehat{\pi}$ be the probability that the principal commits crime against each individual agent. According to Bayes rule, the fraction of wrongful convictions equals

$$\frac{(1-\widehat{\pi})^2 \Pr(a_i=1|\theta_i=0)}{\widehat{\pi}(1-\widehat{\pi})(\Pr(a_i=1|\theta_i=0)+\Pr(a_i=1|\theta_i=1))+\widehat{\pi}^2 \Pr(a_i=1|\theta_i=1)+(1-\widehat{\pi})^2 \Pr(a_i=1|\theta_i=0)}.$$

When the above expression equals $\overline{\pi}$ and $\mathcal{I}_i \equiv \frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)}$ is close to R, some algebra in the online appendix reveals that the expected number of crimes $2\widehat{\pi}$ is close to $\pi_{\min}(\overline{\pi})$.

In the last step, we argue that mechanisms with $Q<0$ (i.e., *concave mechanisms*) cannot improve upon the optimal linear mechanism since the probability that the principal commits multiple crimes increases while the improvement in the informativeness of accusation $\mathcal{I}_i$ is negligible.

Intuitively, the principal's decisions to commit different crimes are strict complements when $Q<0$, so he commits either no crime or two crimes. This leads to a significant increase in the probability that he commits multiple crimes compared to the case in which θ_1 and θ_2 are uncorrelated. Let

$$W_1(\theta_1) \equiv \Big(q(1,0)-q(0,0)\Big)\Pr(a_2=0|\theta_1)+\Big(q(1,1)-q(0,1)\Big)\Pr(a_2=1|\theta_1) \tag{3.6}$$

which is the expected increase in the probability of conviction when agent 1 accuses the principal. When θ_1 and θ_2 are uncorrelated, we have $\frac{W_1(\theta_1=1)}{W_1(\theta_1=0)}=1$. As a result, the informativeness of agent 1's accusation increases compared to the uncorrelated case only if $\frac{W_1(\theta_1=1)}{W_1(\theta_1=0)}>1$.

However, when y is small enough, $\Pr(a_2=1|\theta_1)\approx 0$ for every $\theta_1\in\{0,1\}$. Since $Q<0$, we have $q(1,0)-q(0,0)>q(1,1)-q(0,1)$ and (3.6) is approximately $q(1,0)-q(0,0)$. As a result, $\frac{W_1(\theta_1=1)}{W_1(\theta_1=0)}\approx 1$, which explains why concave mechanisms have negligible impact on the informativeness of accusations.

4 Equilibrium Outcomes under Bayesian Conviction Rules

We now analyze equilibrium outcomes when convictions are decided by a Bayesian judge based on her posterior belief regarding the defendant's guilt. We consider two principles to adjudicate guilt: the aggregate probabilities principle (APP) and the distinct probabilities principle (DPP). As noted in the Introduction, APP has been applied to pretrial detention, laws against organized crimes such as RICO, and in firms and organizations. DPP is widely used in criminal and civil lawsuits.

1. **Aggregate Probabilities Principle (APP):** The judge convicts the principal when the probability that

he is guilty of at least one crime exceeds some threshold $\pi^* \in (0,1)$. Let $\overline{\theta} \equiv \max_{i\in\{1,2\}} \theta_i$ which stands for whether the principal is guilty of at least one crime ($\overline{\theta}=1$) or not ($\overline{\theta}=0$). The judge's decision satisfies:

$$s \begin{cases} = 1 & \text{if} \quad \Pr\left(\overline{\theta}=1 \middle| \boldsymbol{a}\right) > \pi^* \\ \in \{0,1\} & \text{if} \quad \Pr\left(\overline{\theta}=1 \middle| \boldsymbol{a}\right) = \pi^* \\ = 0 & \text{if} \quad \Pr\left(\overline{\theta}=1 \middle| \boldsymbol{a}\right) < \pi^*. \end{cases} \tag{4.1}$$

2. **Distinct Probabilities Principle (DPP):** The judge convicts the principal when the probability that he is guilty of a particular criminal behavior exceeds some threshold $\pi^{**} \in (0,1)$. The judge's decision satisfies:

$$s \begin{cases} = 1 & \text{if} \quad \max_{i\in\{1,2\}} \Pr\left(\theta_i=1 \middle| \boldsymbol{a}\right) > \pi^{**} \\ \in \{0,1\} & \text{if} \quad \max_{i\in\{1,2\}} \Pr\left(\theta_i=1 \middle| \boldsymbol{a}\right) = \pi^{**} \\ = 0 & \text{if} \quad \max_{i\in\{1,2\}} \Pr\left(\theta_i=1 \middle| \boldsymbol{a}\right) < \pi^{**}. \end{cases} \tag{4.2}$$

APP and DPP coincide when there is only one agent and $\pi^* = \pi^{**}$. When there are two agents, APP and DPP can lead to different decisions, as illustrated by the example in the beginning of Section 1.

An equilibrium is a tuple $\left(\sigma_p, \sigma_1, \sigma_2, q\right)$ where the principal's strategy σ_p and agent $i \in \{1,2\}$'s strategy σ_i are defined in the same way as in Section 3. The judge's strategy $q : \{0,1\}^2 \to [0,1]$ maps the agents' messages to the conviction probabilities. Recall that $\boldsymbol{\psi}(\boldsymbol{a}) \in \Delta(\{0,1\}^2)$ is the distribution of (θ_1, θ_2) conditional on $\boldsymbol{a}$, or equivalently, the judge's posterior belief about (θ_1, θ_2) after observing $\boldsymbol{a}$. We examine the common properties of *all* Bayes Nash equilibria that satisfy three refinements.

Refinement 1. $q(1, a_{-i}) \geq q(0, a_{-i})$ *for every* $i \in \{1,2\}$ *and* $a_{-i} \in \{0,1\}$.

Refinement 2. *For every* $\boldsymbol{a}$ *and* $\boldsymbol{a'}$, $\boldsymbol{\psi}(\boldsymbol{a}) = \boldsymbol{\psi}(\boldsymbol{a'})$ *implies* $q(\boldsymbol{a}) = q(\boldsymbol{a'})$.

Refinement 1 is the monotonicity constraint on q. Refinement 2 requires that the judicial outcome to be independent of uninformative messages. These refinements were introduced and discussed in Section 3. We also require that the principal be acquitted unless at least one agent accuses him.

Refinement 3 (No Conviction Unless Accused)**.** $q(0,0) = 0$.

One interpretation of Refinement 3 is that the principal need not even be arrested if nobody accuses him, and a fortiori cannot be convicted in this case. This refinement has no bearing on the optimal mechanism analyzed in Section 3. However, for a Bayesian rule, it rules out equilibria in which the principal commits crime against both agents with probability one and is convicted regardless of agents' reports. Such equilibria are unappealing from a legal standpoint since the principal is convicted based on the judge's prior belief about his guilt rather than informative witness testimonies.

4.1 Equilibrium Outcomes under APP

Theorem 2 characterizes the equilibrium outcomes when the judge uses APP to adjudicate guilt.

Theorem 2. *Suppose that conviction is based on criterion (4.1). For every $\varepsilon > 0$, there exists $\overline{y}_\varepsilon > 0$, such that when $y \in (0, \overline{y}_\varepsilon)$, in every equilibrium that satisfies Refinements 1, 2, and 3,*

1. ***Probability of Crime:*** $\Pr(\theta_1 = 1) = \Pr(\theta_2 = 1) \in (\frac{\pi^* - \varepsilon}{2}, \frac{\pi^*}{2})$ *and* $\Pr(\boldsymbol{\theta} = (1,1)) = 0$.

2. ***Fraction of Wrongful Convictions:*** $\Pr(\boldsymbol{\theta} = (0,0)|s = 1) = 1 - \pi^*$.

3. ***Equilibrium Conviction Probabilities:*** $q(1,1) + q(0,0) - q(1,0) - q(0,1) > 0$.

According to Theorem 2, when the principal's benefit from committing crime is relatively small, he commits at most one crime in every equilibrium but does so with probability close to the conviction cutoff π^*. Therefore, the expected number of crimes $\mathbb{E}[\theta_1 + \theta_2]$ is close to π^*. The conviction probability is strictly convex in the number of accusations and the fraction of wrongful convictions equals $1 - \pi^*$.

Theorem 2 unveils a tension between reducing the probability of crime and reducing the fraction of wrongful convictions. For example, if the judge sets π^* to 10%, then the probability that the potential criminal commits crime is below 10%, but 90% of the convicted people will be innocent.

The proof of Theorem 2 is in Appendix B, An intuitive explanation is provided in the remainder of this section. First, we observe that the principal commits each crime with positive probability. Suppose by way of contradiction that the principal never commits the crime corresponding to agent i. In this case, the judge's posterior belief about (θ_1, θ_2) is independent of agent i's report a_i, and so is the probability of conviction, by Refinement 2. This implies that the principal's expected cost of committing crime against agent i is 0, yet his benefit from doing so is strictly positive. Therefore, he has a strict incentive to choose $\theta_i = 1$, a contradiction.

Next, if the principal's benefit from committing crime is low, he must be convicted with positive probability only if both agents accuse him. To see this, suppose by way of contradiction that a single accusation suffices to convict the principal, which means that either $\Pr(\overline{\theta} = 1|\boldsymbol{a} = (1,0)) \geq \pi^*$ or $\Pr(\overline{\theta} = 1|\boldsymbol{a} = (0,1)) \geq \pi^*$. Since each agent is more likely to accuse the principal when he has witnessed a crime, each additional accusation increases the probability that the principal has committed *at least one crime*. As a result, $\Pr(\overline{\theta} = 1|\boldsymbol{a} = (1,1)) > \pi^*$, so the judge surely convicts the principal when he is accused by both agents: $q(1,1) = 1$. When the benefit from committing crime y is small relative to the loss from conviction, however, the principal strictly prefers not to commit any crime when he is convicted for sure under two accusations, which contradicts our conclusion that the probability of crime is positive.

These arguments suggest that $Q > 0$ in every equilibrium. Similarly to our analysis of Section 3 on unanimous mechanisms, the principal's decisions to commit different crimes are therefore strategic substitutes. This implies that the principal never commits multiple crimes and that his equilibrium strategy induces *negative correlation* in agents' private observations of crimes. In addition, agents' decisions to accuse the principal are strategic complements since each agent's loss from retaliation is lower when the principal is convicted than when he is acquitted. Taken together, these observations imply that an agent is less likely to trigger a conviction if he observed a crime than if he didn't and, hence, that an agent's incentive to accuse the principal when he witnessed a crime is reduced by this negative correlation. Similarly, an agent's incentive to accuse the principal when he did *not* witness a crime is increased by this negative correlation. In summary, the negative correlation affecting agents' observations combined with agents' coordination motive thus reduces the informativeness of agents' reports. This reduces the principal's expected cost of committing crime, which leads to a higher probability of crime.

4.2 Equilibrium Outcomes under DPP

Theorem 3 characterizes the equilibrium outcomes when the judge uses DPP to adjudicate guilt.

Theorem 3. *Suppose that conviction is based on criterion (4.2). For every $\varepsilon > 0$, there exists $\overline{y}_\varepsilon > 0$, such that when $y \in (0, \overline{y}_\varepsilon)$, in every equilibrium that satisfies Refinements 1, 2 and 3,*

- ***Correlation Between Crimes & Equilibrium Probability of Crime:*** *θ_1 and θ_2 are uncorrelated and* $\mathbb{E}[\theta_1] = \mathbb{E}[\theta_2] \in \Big(\frac{\pi^{**}}{(1-\pi^{**})R+\pi^{**}} - \varepsilon, \frac{\pi^{**}}{(1-\pi^{**})R+\pi^{**}} + \varepsilon\Big)$.

- ***Fraction of Wrongful Convictions:***

$$\Pr(\boldsymbol{\theta} = (0,0)|s=1) \in \Big(\frac{(1-\pi^{**})^2 R}{(1-\pi^{**})R+\pi^{**}} - \varepsilon, \frac{(1-\pi^{**})^2 R}{(1-\pi^{**})R+\pi^{**}} + \varepsilon\Big). \tag{4.3}$$

- ***Linear Conviction Probabilities:*** *$q(0,0) = 0$ and $q(1,1) = 2q(1,0) = 2q(0,1) > 0$.*

According to Theorem 3, when y is small relative to 1, different crimes are uncorrelated and the probability of conviction is linear in the number of accusations. The expected number of crimes is

$$\mathbb{E}[\theta_1 + \theta_2] \approx \frac{2\pi^{**}}{(1-\pi^{**})R+\pi^{**}}. \tag{4.4}$$

When R is large, i.e., when the fraction of agents who are prone to make false accusations is small, Theorems 2 and 3 unveil the relative power of APP and DPP to deter crime, fixing the fraction of wrongful

convictions. Compared to APP, DPP introduces the possibility that potential criminals committing multiple crimes but reduces the probability that they commit at least one crime. Whether APP or DPP minimizes the *expected number of crimes* depends on the standards of proof (π^*, π^{**}) that determine the extent to which improved informativeness can deter crimes, and R that determines the extent to which DPP can improve the informativeness of accusations. More details on this comparison are discussed in Section 4.3.

The proof of Theorem 3 is in Appendix C. The key distinction between APP and DPP is that, while a larger number of accusations against the principal always increases the probability that the principal is guilty of at least one crime, it does not necessarily increase the probability that the principal is guilty of a specific crime. When θ_1 and θ_2 are negatively correlated (as in APP), or uncorrelated (as in DPP), an accusation by agent 2 undermines or leaves unchanged the credibility of an accusation by agent 1. As a result, the value of the criterion $\max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a})$, used in DPP, need not increase as a result of agent 2's accusation, and in particular, it can stay at the conviction threshold π^{**}. This explains why interior conviction probabilities that are linear in the number of accusations arise under DPP but not under APP.

Next, we explain why θ_1 and θ_2 must be uncorrelated. When y is small, both $q(1,0)$ and $q(0,1)$ are close to 0. Intuitively, when $q(1,0)$ or $q(0,1)$ were bounded away from 0, the principal would have a strict incentive not to commit any crime when y is small enough and, hence, would never be convicted given that the ex ante probability of crime is zero. This, however, gives the principal a strict incentive to commit crime and yields a contradiction.

Suppose first that θ_1 and θ_2 are *negatively correlated*, then $\Pr(\theta_1 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_1 = 1|\boldsymbol{a} = (1,0))$ and $\Pr(\theta_2 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_2 = 1|\boldsymbol{a} = (0,1))$. When conviction is based on criterion (4.2), the above inequalities imply that $q(1,0) \geq q(1,1)$ and $q(0,1) \geq q(1,1)$, and the monotonicity requirement on q implies that $q(1,1) = q(1,0) = q(0,1) = 1$. This contradicts our previous conclusion that $q(1,0)$ and $q(1,0)$ are close to 0, which rules out negative correlation.

Suppose next that θ_1 and θ_2 are *positively correlated*, in which case $\boldsymbol{a} = (1,1)$ is the unique maximizer of $\Pr\left(\theta_1 = 1\middle|\boldsymbol{a}\right)$ and $\Pr\left(\theta_2 = 1\middle|\boldsymbol{a}\right)$. According to (4.2), $q(1,1) \geq \max\{q(1,0), q(0,1)\} \geq q(0,0) = 0$. Given our previous conclusion that $q(1,0)$ and $q(0,1)$ are close to 0, either $q(1,1) = 1$, or $q(1,1) \in (0,1)$ and $q(1,0) = q(0,1) = 0$. In both cases, we have $Q > 0$. Therefore, the principal's decisions to commit different crimes are strategic substitutes. This contradicts the presumption that θ_1 and θ_2 are positively correlated.

Since we have shown that θ_1 and θ_2 are uncorrelated, the principal's decisions to commit different crimes are neither substitutes nor complements. This implies that $Q = 0$. Since the probability that the principal commits crime against each agent is strictly between 0 and 1, his expected cost of committing each crime

must be the same. This implies that $q(1,0) = q(0,1)$. Refinement 3 requires that $q(0,0) = 0$. This together with $Q = 0$ and $q(1,0) = q(0,1)$ implies that q is a linear function in the number of accusations. The derivation that $\mathcal{I}_i \approx R$ follows from the one under linear mechanisms in Section 3, which we omit in order to avoid repetition.

4.3 Attainability of Optimal Commitment Outcomes

Taken together, Theorems 1, 2, and 3 imply that the optimal commitment outcome can be approximately attained by any equilibrium that satisfies Refinements 1, 2 and 3 in a game without commitment in which the judge makes conviction decisions based on APP or DPP using her posterior belief about (θ_1, θ_2).

1. When $R \leq \frac{\bar{l}}{2} + 1$, or equivalently, $1 - \overline{\pi} \leq \pi_{\min}(\overline{\pi})$, the optimal commitment outcome can be approximately attained in every equilibrium when the judge uses (4.1) to adjudicate guilt with $\pi^* \equiv 1 - \overline{\pi}$.

2. When $R > \frac{\bar{l}}{2} + 1$, or equivalently, $\pi_{\min}(\overline{\pi}) < 1 - \overline{\pi}$, the optimal commitment outcome can be approximately attained in every equilibrium when the judge uses (4.2) to adjudicate guilt with π^{**} satisfying
$$\overline{\pi} = \frac{(1-\pi^{**})^2 R}{(1-\pi^{**})R + \pi^{**}}. \tag{4.5}$$
Intuitively, one can use (4.3) to verify that when π^{**} satisfies (4.5) and the judge uses DPP to adjudicate guilt, the fraction of wrongful convictions is approximately $\overline{\pi}$.

4.4 Discussion

Equilibrium Existence: Theorems 2 and 3 establish the common properties of all equilibria that satisfy our refinements. Proposition 1 complements these theorems by establishing the existence of such equilibria.

Proposition 1. *Suppose the environment satisfies inequality (2.1). There exists $\overline{y} > 0$ such that for every $y \in (0, \overline{y})$, there always exists an equilibrium that satisfies Refinements 1, 2, and 3 when conviction is based on criterion (4.1) or on criterion (4.2).*

The proof, in Appendix A, uses Brouwer's fixed point theorem to construct equilibria that satisfy our refinements. The existence of equilibrium requires y to be small. For example, if $y \geq 1$, the principal's benefit from committing crime exceeds his loss from conviction. As a result, the principal commits crime with probability 1 in equilibrium, and if the judge uses APP, the principal is convicted regardless of agents'

accusations in all Bayes Nash equilibria. This means that there exists no equilibrium that satisfies Refinement 3.

Single-Agent Benchmark: A takeaway from our analysis is that APP leads to a high probability of crime due to an *endogenous negative correlation* in agents' private observations of crime and an *endogenous coordination motive* among agents. An alternative explanation for the high frequency of crime under APP is that due to the cost of reporting crime, agents free-ride on each other's accusations when there are multiple agents. We rule out this alternative explanation by examining a benchmark scenario in which there is only one agent and the judge uses (4.1) to adjudicate guilt. Proposition 2 compares the equilibrium outcomes in this single-agent benchmark with those in the two-agent scenario.

Proposition 2. *If there is only one agent, the judge uses (4.1) to adjudicate guilt, and there exists $\overline{\omega} \in \mathbb{R}$ such that $\phi(\omega)$ is strictly increasing when $\omega < \overline{\omega}$, then there exists $\overline{y} \in (0,1)$ such that when $y \in (0, \overline{y})$,*[13]

- *Compared to any equilibrium that satisfies Refinements 1, 2 and 3 in the two-agent environment, the probability that each agent files an accusation conditional on any $\theta \in \{0,1\}$ is strictly lower in any equilibrium that satisfies Refinements 1, 2 and 3 in the single-agent benchmark.*

Proposition 2, whose proof is in Appendix F, shows that each agent accuses the principal with strictly higher probability in the two-agent scenario than in the single-agent benchmark. In Pei and Strulovici (2020), we generalize this result to any finite number of agents in a closely related model: an increase in the number of agents (or equivalently, an increase in the number of opportunities to commit crimes) increases the probability that each potential witness files an accusation.

Proposition 2 shows that the inefficiency predicted by Theorem 2 is not driven by the free-riding logic that underlies results on inefficient public good provision. In our model, accusing the principal is costly but each agent accuses the principal with *strictly higher* probability when there are more agents. The increased probability of crime with more agents is not caused by agents' incentives to free-ride on one another's accusations, but rather by the low informativeness of accusations that results from the endogenous negative correlation in agents' private observations of crimes.

Independence of Uninformative Messages: The refinement requiring that uninformative messages do not affect the judicial outcome imposes restrictions on agents' beliefs after they observe the principal taking off-path actions. For example, suppose that APP is used to adjudicate guilt and that the principal commits

[13] The above condition on the distribution of ω is satisfied for all normal distributions and exponential distributions.

crime against agent 2 with zero probability. To deter crime against agent 2, the probability of conviction and the probability that agent 2 accuses the principal must be strictly higher when agent 2 has witnessed a crime. However, agent 2's belief about $\theta_1 = 1$ is independent of his observation of θ_2. Therefore, any equilibrium with these features must violate the independence refinement, since agent 2's message does not affect the judge's belief about (θ_1, θ_2) and yet the judge's decision varies with his message.

More generally, for mechanisms such that $Q > 0$, equilibria that satisfy our refinement 2 are proper equilibria (Myerson 1978). By contrast, equilibria in which the principal never commits crime against one of the agents (e.g., equilibria that satisfy some passive belief refinement) are not proper equilibria: In such equilibria, the principal chooses $(\theta_1, \theta_2) = (0, 0)$ and $(1, 0)$ with positive probability (say), and other actions with zero probability. Since $Q > 0$, the principal's crimes are strategic substitutes. Therefore, the principal's loss from deviating to $(\theta_1, \theta_2) = (1, 1)$ is strictly greater than that from deviating to $(\theta_1, \theta_2) = (0, 1)$. Proper equilibrium requires that under each tremble along an infinite sequence, the probability that the principal chooses $(\theta_1, \theta_2) = (1, 1)$ be arbitrarily small compared to the probability of choosing $(\theta_1, \theta_2) = (0, 1)$. Therefore, agent 2 should believe that $\theta_1 = 0$ occurs with probability 1 after observing $\theta_2 = 1$. An argument similar to the proof of Lemma B.4 rules out such equilibria by showing that under the above off-path beliefs, the expected cost of committing crime against agent 2 is strictly lower than that against agent 1. Hence, whenever the principal has a weak incentive to commit crime against agent 1, he has a strict incentive to commit crime against agent 2.

5 Extensions

General Mechanisms: The mechanism design analysis of Section 3 focuses on mechanisms that map agents' reports to a probability of conviction. One could a priori consider more general *direct mechanisms* in which agents report not only their private observations of crime but also their type (ω_i, c_i).

As it turns out, eliciting private information about ω_i and c_i cannot lower the expected number of crimes or the fraction of wrongful convictions as long as (ω_1, c_1) is independent of (ω_2, c_2). Intuitively, each agent has a strict preference for either convicting or acquitting the principal almost surely. Regardless of the realization of (ω_i, c_i), agent i will report the pair $(\widehat{\omega}_i, \widehat{c}_i)$ that maximizes the probability of conviction if the agent strictly prefers to convict the principal, and minimizes this probability if he strictly prefers to acquit the principal. All the information that can be inferred from the agent's richer report is therefore reduced to whether the agent wishes to accuse the principal or not, which is precisely what is considered in Section 3.

Another restriction is that the designer can only commit to the probability of conviction but cannot adjust

the size of punishment. This assumption, made for tractability, is consistent with some applications, in which the principal's disutility from a conviction includes large reputation and career losses that are beyond the control of the judge and come in addition to any fine or jail time that can be set by the judge.

Objective Functions & Constraints: The design problem considered in Section 3 aims to minimize the expected number of crimes subject to an upper bound on the fraction of wrongful convictions.

The qualitative features of our Theorem 1 extend when the designer minimizes an objective function that puts general weights on having one crime and having multiple crimes. For example, the designer might aim to minimize

$$\Pr\Big((\theta_1,\theta_2)=(1,0)\Big)+\Pr\Big((\theta_1,\theta_2)=(0,1)\Big)+\beta\Pr\Big((\theta_1,\theta_2)=(1,1)\Big), \tag{5.1}$$

where $\beta > 1$ is a parameter. When $\beta = 2$, the objective function in (5.1) coincides with the one in our baseline model. When β is larger, the designer is more averse to having multiple crimes.

The value of β affects the conditions on R and $\overline{\pi}$ that determine whether linear mechanisms or unanimous mechanisms are optimal, but it does not affect our result qualitatively. In particular, linear mechanisms are optimal when R and $\overline{\pi}$ are large, and unanimous mechanisms are optimal otherwise. When β increases, the designer puts more weight on reducing the probability that the principal commits multiple crimes, and increases the desirability of unanimous mechanisms.

The designer may also face an upper bound on the fraction of mistaken acquittals, of the form $\Pr(\boldsymbol{\theta} \neq (0,0)|s=0) \leq \pi'$ for some $\pi' \in (0,1)$. When the benefit from crime y is small enough, the probability of conviction is close to 0 under every $\overline{\pi}$-valid outcome. Therefore, the probability of mistaken acquittals is arbitrarily close to the unconditional probability that the principal commits at least one crime, which can also be characterized by our analysis. For example, under unanimous mechanisms, $\Pr(\boldsymbol{\theta} \neq (0,0)|s=0)$ is approximately $1-\overline{\pi}$, and under linear mechanisms,

$$\Pr(\boldsymbol{\theta} \neq (0,0)|s=0) \approx \pi_{\min}(\overline{\pi}) - \frac{\pi^2_{\min}(\overline{\pi})}{4}. \tag{5.2}$$

Our analysis can then characterize the set of $(\overline{\pi},\pi')$ such that there exists at least one $\overline{\pi}$-valid outcome that satisfies $\Pr(\boldsymbol{\theta} \neq (0,0)|s=0) \leq \pi'$ and $\Pr(\boldsymbol{\theta} = (0,0)|s=1) \leq \overline{\pi}$ when y is small. For every $(\overline{\pi},\pi')$ such that the set of outcomes satisfying both constraints is nonempty, we can compute the lowest expected number of crimes, or more generally, the lowest value of the general objective function (5.1).

Heterogenous Propensity to Commit Crimes: A realistic consideration in the study of criminal behavior, omitted from our analysis, is that the principal's benefit from the commission of crimes may be privately known. This consideration complicates the analysis of the problem but does not fundamentally change the forces identified in the paper.

In an early version of this project (Pei and Strulovici 2020), we examine situations in which the principal is either a *virtuous type*, who experiences no benefit from committing crime, or an *opportunistic type*, whose benefit from committing crime is strictly positive. When the probability of the virtuous type is below some cutoff, the equilibrium outcomes under DPP and APP coincide with those in the baseline model. When the probability of the virtuous type is above that cutoff, the opportunistic type principal commits multiple crimes with positive probability under APP, but different crimes remain *negatively correlated* as in the analysis of the present paper. Compared to the equilibrium outcomes under DPP, APP reduces the probability that the principal committing multiple crimes, but increases the probability that he commits at least one crime when R is above some cutoff.

Altruistic Agents & Communication Between Agents: In our baseline model, a witness's utility depends only on the offense that he may be observe, not on the offense associated with the other witness. Realistically, agents may also care about these other offenses. This possibility can be modeled by modifying agent i's payoff as follows:

$$s(\theta_i + \alpha\theta_j - \gamma c_i a_i - \omega_i) + (1-s)(-c_i a_i), \tag{5.3}$$

where $\alpha \geq 0$ measures the extent to which agent i's payoff internalizes crime committed against agent j.

Somewhat counter-intuitively, adding this altruistic component to agents' utility worsens the coordination and credibility problems emphasized in earlier sections when APP is used to adjudicate the case. The principal's incentives to commit different crimes are still strategic substitutes under APP, which means that different crimes are negatively correlated. The informativeness of agents' reports is even lower, since agent i's incentive to convict the principal depends not only on his belief about a_j, but also on his belief about θ_j given that θ_j directly affects his payoff. Given that θ_i and θ_j are negatively correlated, the dependence of agent i's payoff on θ_j weakens his incentive to accuse the principal when $\theta_i = 1$ and vice versa. This lowers the informativeness of each agent's accusation. By contrast, different crimes remain uncorrelated under DPP and the comparative advantages of APP and DPP remain the same as in the baseline model.

When agents can **communicate** with each other before deciding whether to accuse the principal, the collusion between agents alleviates the coordination failures, but lowers the informativeness of their accusations, especially when both agents accuse (or do not accuse) at the same time.

Simultaneous vs Sequential Reporting: A more realistic model of criminal behavior would allow crimes to be committed at different times and witnesses to report crimes sequentially, possibly observing earlier reports.[14] The forces underlying our results are also present in dynamic versions of our model, in which reports are filed sequentially. First, the negative correlation between witnesses' observations θ_1 and θ_2 still occurs when the probability of conviction is strictly convex in the number of accusations, which is the case under APP when y is small enough. Second, each agent has an incentive to coordinate with the other agent 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 negative correlation and agents' coordination motive thus still arise in a dynamic environment.

6 Concluding Remarks

We discuss the applicability of our analysis and its connections to the existing literature on voting, communication and coordination games, and law and economics.

Rule Change and Equilibrium Analysis: Equilibrium analysis, like the one in this paper, assumes that players have correct expectations about the consequences of their actions and other players' strategies. When social rules change, as in the case of a sudden crackdown on a specific type of offense, the introduction of new regulation, a drastic shift 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 or norms. This distinction seems particularly relevant in the context of the recent *me too* movement, since abusers before the emergence of the movement likely underestimated the legal and professional consequences of their abusive behavior.

Related Literature: The game studied in Section 4 may be viewed as a voting model in which agents are voters who have endogenous, correlated signals about the, also endogenous, state of the world, and votes are aggregated to form a conviction decision. It may also be viewed as a model of strategic information transmission with multiple senders, in which the state is endogenously determined by the principal.

[14] Such a model is developed by Lee and Suen (2020) and discussed in the next section.

Our framework may be distinguished from the models of voting and communication by Feddersen and Pesendorfer (1998), Battaglini (2002), Ambrus and Takahashi (2008), and Ekmekci and Lauermann (2019), in which the facts of interest are exogenous; models with endogenous information acquisition such as Persico (2004), Pei (2015), Argenziano, Severinov and Squintani (2016), and Deimen and Szalay (2019), in which there is either only one player who can acquire private information, or players' private signals are assumed to be conditionally independent; voting models with negatively correlated private signals or payoffs such as Schmitz and Tröger (2012) and Ali, Mihm, and Siga (2018) in which voters' information structures are exogenous; and dynamic voting models in which information acquisition may induce negative correlation in voters' continuation values (Strulovici 2010).

Our finding that agents' reports become arbitrarily uninformative when the principal's decisions are strategic substitutes provides a new mechanism for failures of information aggregation. In Banerjee (1992), Bikhchandani, Hirshleifer and Welch (1992), and Smith and Sørensen (2000), agents fail to act on their private information because they can observe the actions taken by their predecessors. By contrast, agents move simultaneously in our model and information aggregation fails as a result of the combination of the negative correlation in agents' private information and of agents' incentives to coordinate their reports.[15]

The coordination motive among witnesses is reminiscent of the literature on global games.[16] In Carlsson and Van Damme (1993) and Morris and Shin (1998), agents receive conditionally independent private signals about the state. In Baliga and Sjöström (2004), each player privately observes his value for a decision, which is independent of the signal received by the other player. In our model, the coordination arises endogenously, agents' private signals are correlated, and this correlation is endogenous.

Our paper contributes to the literature on law and economics by studying how to optimally aggregate the probabilities of offenses in environments where the incentives to commit and to report crimes are both endogenous. We study mechanisms that minimize the expected number of crimes subject to an upper bound on the fraction of wrongful convictions, and show that the optimal commitment outcome can be attained even judges cannot commit to a mechanism and make decisions based on posterior beliefs. This approach stands in contrast to several contemporaneous papers that focus on witnesses' incentives to report crimes. Lee and Suen (2020) study the timing of reports by victims and libelers when a defendant commits crimes against two potential victims with exogenous probability and focus attention on equilibria in which the defendant is convicted only if both agents accuse him. They provide an explanation for the well-documented observation

[15] Strulovici (2020) studies a sequential learning model in which an agent is less likely to have an informative signal, other things equal, if another agent has found such a signal. This induces negative correlation in the *informativeness* of agents' signals rather than in the *direction* of these signals, as in the present paper.

[16] This approach to crime reporting is explicitly studied by Cheng and Hsiaw (2020), discussed in the next paragraph.

that victims sometimes delay their accusations. Cheng and Hsiaw (2020) adopt a global game perspective to study the reporting incentives of a continuum of agents who observe conditionally independent signals of the state of the world. Naess (2020) also considers witnesses' reporting incentives and, among other results, finds that making reporting costly may improve social welfare.

Siegel and Strulovici (2020) apply mechanism design to a judicial setting in which the mechanism designer can elicit information from the defendant and the designer can modify the sentencing scheme without affecting the evidence available. In the present paper, the designer solicits information from potential witnesses or victims. The quality of the evidence is endogenous and depends on the mechanism as well as the defendant's incentives to commit crimes.[17]

[17] Mechanism design has been applied to study other areas of law. In particular, Spier (1994), Klement and Neeman (2005), and Demougin and Fluet (2006) analyze settlement and fee-shifting rules between plaintiffs and defendants.

A Proof of Proposition 1

A.1 Existence of Equilibrium under APP

We establish the existence of an equilibrium that satisfies Refinements 1, 2, and 3, such that $q(0,0) = q(1,0) = q(0,1) = 0$, $q(1,1) \in (0,1)$, and the principal chooses $(0,0)$, $(1,0)$, and $(0,1)$ with positive probability.

Lemma A.1. *Suppose that F and Φ satisfy (2.1). Then, there exists a constant $\overline{y} > 0$ such that the following holds. For any $(\Phi^*, \Phi^{**}) \in (0,1)^2$ that satisfies*

$$\omega^*(c) = 1 + c(1-\gamma) - \frac{c}{\Phi^{**}}, \tag{A.1}$$

$$\omega^{**}(c) = c(1-\gamma) - \frac{2}{2+l^*} \cdot \frac{c}{\Phi^{**}} - \frac{l^*}{2+l^*} \cdot \frac{c}{\Phi^*}, \tag{A.2}$$

$\Phi^* \equiv \int \Phi(\omega^*(c))dF(c)$, and $\Phi^{**} \equiv \int \Phi(\omega^{**}(c))dF(c)$, we have

$$\Phi^{**}(\Phi^* - \Phi^{**}) \geq \overline{y}.$$

Proof. First, we bound Φ^{**} from below. We have

$$\Phi^{**} = \int \Phi\Big(c(1-\gamma) - \frac{2}{2+l^*} \cdot \frac{c}{\Phi^{**}} - \frac{l^*}{2+l^*} \cdot \frac{c}{\Phi^*}\Big)dF(c) \geq \int \Phi\Big(c(1-\gamma) - \frac{c}{\Phi^{**}}\Big)dF(c). \tag{A.3}$$

Let $g(\Phi^{**}) = \Phi^{**}$ and $h(\Phi^{**}) = \int \Phi\Big(c(1-\gamma) - \frac{c}{\Phi^{**}}\Big)dF(c)$. We have $g(0) = h(0)$, $g(1) > h(1)$, and

$$h(\Phi^{**}) \geq \int_0^{\alpha\Phi^{**}} \Phi\Big(\alpha\Phi^{**}(1-\gamma) - \alpha\Big)dF(c) = F(\alpha\Phi^{**})\Phi(\alpha\Phi^{**}(1-\gamma) - \alpha) \text{ for every } \alpha > 0.$$

The derivative of the RHS with respect to Φ^{**} is $\alpha f(0)\Phi(-\alpha)$ at $\Phi^{**} = 0$, and the derivative of $g(\Phi^{**})$ is 1. When F and Φ satisfy (2.1), there exists $\varepsilon > 0$ such that $h(\Phi^{**}) > g(\Phi^{**})$ for every $\Phi^{**} \in (0, \varepsilon)$. Moreover, there exists a fixed point with $\Phi^{**} > \varepsilon$ according to the intermediate value theorem.

Next, we bound $\Phi^* - \Phi^{**}$ from below. First, $\Phi^* > \Phi^{**}$.[18] Equations (A.1) and (A.2) imply that

$$\omega^*(c) - \omega^{**}(c) = 1 - c \cdot \frac{l^*}{2+l^*} \cdot \frac{\Phi^* - \Phi^{**}}{\Phi^* \cdot \Phi^{**}}. \tag{A.4}$$

[18] Suppose not: then, $\omega^*(c) > \omega^{**}(c)$ from (A.1) and (A.2), which implies that $\Phi^* > \Phi^{**}$ from the definitions of Φ^* and Φ^{**} following (A.2).

Since $\Phi^* > \Phi^{**} > \varepsilon$, for every $c^* > 0$, there exists $\eta > 0$ such that when $\Phi^* - \Phi^{**} < \eta$, we have $\omega^*(c) - \omega^{**}(c) > \frac{1}{2}$ for every $c \in [0, c^*]$. Suppose that $\Phi^* - \Phi^{**} < \eta$ for some c^* to be chosen shortly. Then,

$$\Phi^* - \Phi^{**} \geq \int_0^{c^*} \Big(\Phi(\omega^{**}(c) + \frac{1}{2}) - \Phi(\omega^{**}(c))\Big) dF(c) - \int_{c^*}^{+\infty} \Phi(\omega^{**}(c)) dF(c). \tag{A.5}$$

Let $\Psi \equiv \min_{\omega \in [1-\gamma-\frac{1}{\varepsilon}, 0]} \Big\{\Phi(\omega + \frac{1}{2}) - \Phi(\omega)\Big\}$. The RHS of (A.5) is at least $F(1)\Psi - (1 - F(c^*))$ when $c^* > 1$. Pick $c^* > 1$ large enough such that $\frac{1}{2}F(1)\Psi > 1 - F(c^*)$. We have $\Phi^* - \Phi^{**} \geq \frac{1}{2}F(1)\Psi$. Hence, we have$\Phi^* - \Phi^{**} \geq \min\{\frac{1}{2}F(1)\Psi, \eta\}$. Combining this with the first part of the proof yields a uniform lower bound on $\Phi^{**}(\Phi^* - \Phi^{**})$. □

We use the constant ε derived in the proof of Lemma A.1 and fix some $y \in (0, 1)$. For every $(\Phi^*, \Phi^{**}, q) \in [\varepsilon, 1] \times [\varepsilon, 1] \times [y, 1]$, let $f \equiv (f_1, f_2, f_3) : [\varepsilon, 1] \times [\varepsilon, 1] \times [y, 1] \to [\varepsilon, 1] \times [\varepsilon, 1] \times [y, 1]$ be defined as:

$$f_1(\Phi^*, \Phi^{**}, q) = \max\Big\{\varepsilon, \int \Phi\Big(1 + c(1-\gamma) - \frac{c}{q\Phi^{**}}\Big) dF(c)\Big\}, \tag{A.6}$$

$$f_2(\Phi^*, \Phi^{**}, q) = \max\Big\{\varepsilon, \int \Phi\Big(c(1-\gamma) - \frac{2}{2+l^*} \cdot \frac{c}{q\Phi^{**}} - \frac{l^*}{2+l^*} \cdot \frac{c}{q\Phi^*}\Big) dF(c)\Big\}, \tag{A.7}$$

$$f_3(\Phi^*, \Phi^{**}, q) = \min\Big\{1, \frac{y}{\Phi^{**}(\Phi^* - \Phi^{**})}\Big\}. \tag{A.8}$$

Since f is continuous, Brouwer's fixed-point theorem implies the existence of a fixed point. The construction of ε in Lemma A.1 implies that when $y < \varepsilon$, we must have $\Phi^* > \varepsilon$ and $\Phi^{**} > \varepsilon$ at every fixed point. Suppose not: we must have $\Phi^{**} = \varepsilon$ for some fixed point of f. Equation (A.8) together with $y < \varepsilon$ implies that $q = 1$. Equation (A.7) then implies that

$$\Phi^{**} = \int \Phi\Big(c(1-\gamma) - \frac{2}{2+l^*} \cdot \frac{c}{q\Phi^{**}} - \frac{l^*}{2+l^*} \cdot \frac{c}{\Phi^*}\Big) dF(c) \geq \int \Phi\Big(c(1-\gamma) - \frac{c}{\varepsilon}\Big) dF(c) > \varepsilon,$$

a contradiction. Similarly, $\Phi^* > \varepsilon$ since $\Phi^* \geq \Phi^{**}$. Lemma A.1 implies that when the distribution satisfies (2.1) and $y < \overline{y}$, every fixed point of f has $q < 1$ when $y < \overline{y}$. This implies the existence of an equilibrium that satisfies Refinements 1, 2 and 3.

A.2 Existence of Equilibrium under DPP

We establish the existence of an equilibrium that satisfies Refinements 1, 2, and 3 when

$$4y \leq \min_{K \in [-(1+\gamma), 1-\gamma]} \int_0^{\overline{c}} \Big(\Phi(1 - cK) - \Phi(-cK)\Big) dF(c). \tag{A.9}$$

In particular, the conviction probabilities are linear in the number of accusations, the agent's decisions to commit crimes are independently distributed, and the principal is indifferent between his four actions. First, Brouwer's fixed point theorem implies that there exists $(\Phi^*, \Phi^{**}, q^*, r^*) \in [0,1] \times [0,1] \times [0, \frac{1}{2}] \times [0,1]$ that is a fixed point of:

$$q^* = \min\Big\{\frac{1}{2}, \frac{y}{\Phi^* - \Phi^{**}}\Big\}, \tag{A.10}$$

$$\frac{\Phi^*}{\Phi^{**}} \cdot \frac{r^*}{1-r^*} = \frac{\pi^{**}}{1-\pi^{**}}. \tag{A.11}$$

$$\Phi^* = \int_0^{\overline{c}} \Phi\Big(1 + c(1-\gamma) - c\frac{1-(1-\gamma)q^*(r^*\Phi^* + (1-r^*)\Phi^{**})}{q^*}\Big) dF(c), \tag{A.12}$$

$$\Phi^{**} = \int_0^{\overline{c}} \Phi\Big(c(1-\gamma) - c\frac{1-(1-\gamma)q^*(r^*\Phi^* + (1-r^*)\Phi^{**})}{q^*}\Big) dF(c). \tag{A.13}$$

This fixed point $(\Phi^*, \Phi^{**}, q^*, r^*)$ is an equilibrium of the game under DPP if $q^* < 1/2$. Suppose toward a contradiction that there exists a fixed point with $q^* = 1/2$. Then

$$\frac{1-(1-\gamma)q^*(r^*\Phi^* + (1-r^*)\Phi^{**})}{q^*} \in [0,2],$$

and the value of $\Phi^* - \Phi^{**}$ is greater than the RHS of (A.9). As a result, $\frac{y}{\Phi^* - \Phi^{**}} < 1/4$, which contradicts the hypothesis that $q^* = 1/2$ is a fixed point.

B Proof of Theorem 2

The proof is decomposed into a sequence of lemmas.

Lemma B.1. *In every equilibrium that satisfies Refinements 1, 2, and 3,* $\Pr(\theta_i = 1) \in (0,1)$ *for every* $i \in \{1,2\}$.

Recall the definition of Q in (3.4). The next lemma shows that when the benefit from committing crime is small, the value of Q is strictly positive for every equilibrium.

Lemma B.2. *There exists* $\overline{y} \in (0,1)$ *such that for every* $y \in (0, \overline{y})$ *and every equilibrium that satisfies Refinements 1, 2 and 3, we have* $Q > 0$ *and* $q(0,1) = q(1,0) = 0$.

The next lemma provides a sufficient statistic that determines whether the principal's choices of θ_1 and θ_2 are strategic substitutes or strategic complements.

Lemma B.3. *The principal's choices of* θ_1 *and* θ_2 *are strategic substitutes if and only if* $Q > 0$*, and strategic complements if and only if* $Q < 0$.

We show that when y is small, all equilibria that satisfy our refinements are symmetric across agents.

Lemma B.4. *There exists $\overline{y} \in (0,1)$ such that for every $y \in (0,\overline{y})$, $\sigma_1 = \sigma_2$ and the principal chooses $\boldsymbol{a} = (0,1)$ and $\boldsymbol{a} = (1,0)$ with equal probability in every equilibrium that satisfies Refinements 1, 2, and 3.*

In every equilibrium that satisfies Refinements 1, 2, and 3, Lemma B.1 implies that the principal chooses $(\theta_1, \theta_2) = (0,0)$ with positive probability. Lemma B.2 implies that $q(1,1) + q(0,0) - q(1,0) - q(0,1) > 0$. Lemma B.3 implies that the principal chooses $(\theta_1, \theta_2) = (1,1)$ with zero probability. Lemma B.4 implies that in every equilibrium, the principal chooses $(\theta_1, \theta_2) = (1,0), (0,1)$ and $(0,0)$ with positive probability. The last lemma uses these conclusions to show that the informativeness ratio converges to 1 in all equilibria.

Lemma B.5. *For every $\varepsilon > 0$, there exists $\overline{y} \in (0,1)$ such that when $y \in (0,\overline{y})$, in every equilibrium that satisfies Refinements 1, 2, and 3,*

$$\mathcal{I} \equiv \frac{\Pr(\boldsymbol{a} = (1,1) | \overline{\theta} = 1)}{\Pr(\boldsymbol{a} = (1,1) | \overline{\theta} = 0)} < 1 + \varepsilon. \tag{B.1}$$

According to Bayes rule, $\mathcal{I} \frac{\Pr(\overline{\theta}=1)}{1-\Pr(\overline{\theta}=1)} = \frac{\Pr(\overline{\theta}=1|\boldsymbol{a}=(1,1))}{1-\Pr(\overline{\theta}=1|\boldsymbol{a}=(1,1))}$. Since $q(1,1) \in (0,1)$, we have $\Pr(\overline{\theta} = 1 | \boldsymbol{a} = (1,1)) = \pi^*$ and $\Pr(\overline{\theta} = 1) < \pi^*$. Therefore, π converges to π^* as $y \to 0$, which concludes the proof. We prove Lemma B.1 in Section B.1, Lemma B.3 in Section B.2, Lemma B.4 in Section B.3, and Lemma B.5 in Section B.4. The proof of Lemma B.2 is similar to that of Lemmas C.1 and D.2, which receive a unified treatment in Appendix E.

B.1 Proof of Lemma B.1

Equation (2.2) implies that agent i strictly prefers $a_i = 0$ if

$$\underbrace{\mathbb{E}[q(0,a_{-i}) - q(1,a_{-i})]}_{\leq 0}(\omega_i - \theta_i) < c_i \underbrace{\mathbb{E}[1 - (1-\gamma)q(1,a_{-i})]}_{>0}. \tag{B.2}$$

and strictly prefers $a_i = 1$ if $\mathbb{E}[q(0,a_{-i}) - q(1,a_{-i})](\omega_i - \theta_i) > c_i \mathbb{E}[1 - (1-\gamma)q(1,a_{-i})]$. Since $\mathbb{E}[1 - (1-\gamma)q(1,a_{-i})] > 0$, inequality (B.2) is satisfied when $\omega_i > 1$ and $c_i > 0$. This together with the full support assumption implies that for every $i \in \{1,2\}$, agent i chooses $a_i = 0$ with positive probability.

Suppose toward a contradiction that $\theta_i = 0$ with probability 1 for some $i \in \{0,1\}$. The principal weakly prefers $\theta_i = 0$ to $\theta_i = 1$. Since his benefit from choosing $\theta_i = 1$ is strictly positive, its cost must also be strictly positive, which implies that $\mathbb{E}[q(0,a_{-i}) - q(1,a_{-i})] < 0$. We consider two cases separately. First, if both $a_i = 1$ and $a_i = 0$ occur with positive probability, then for every a_{-i} that occurs

with positive probability, both $\boldsymbol{\pi}(1, a_{-i})$ and $\boldsymbol{\pi}(0, a_{-i})$ are pinned down by Bayes rule and are equal to each other. Refinement 2 implies that $q(1, a_{-i}) = q(0, a_{-i})$. This contradicts the previous conclusion that $\mathbb{E}[q(0, a_{-i}) - q(1, a_{-i})] < 0$. Second, if $a_i = 0$ with probability 1, since $\mathbb{E}[q(0, a_{-i}) - q(1, a_{-i})] < 0$ and the distributions of ω_i and c_i have full support, there exist ω_i and c_i such that $\mathbb{E}[q(0, a_{-i}) - q(1, a_{-i})](\omega_i - \theta_i) > c_i \mathbb{E}[1 - (1 - \gamma)q(1, a_{-i})]$. As a result, agent i chooses both $a_i = 1$ and $a_i = 0$ with positive probability, which contradicts $a_i = 0$ with probability 1.

When the equilibrium satisfies Refinement 3, suppose toward a contradiction that there exists $i \in \{1, 2\}$ such that $\theta_i = 1$ with probability 1. Since we have shown that $\boldsymbol{a} = (0, 0)$ with strictly positive probability, $\Pr(\overline{\theta} = 1 | \boldsymbol{a} = (0, 0)) \geq \Pr(\theta_i = 1 | \boldsymbol{a} = (0, 0)) = 1$. Therefore, $q(0, 0) = 1$, which violates Refinement 3.

B.2 Proof of Lemma B.3

Let $\Phi_i^* \equiv \Pr(a_i = 1 | \theta_i = 1)$ and let $\Phi_i^{**} \equiv \Pr(a_i = 1 | \theta_i = 0)$. Lemma B.1 implies that $\mathbb{E}[q(0, a_{-i}) - q(1, a_{-i})] < 0$ in every equilibrium that satisfies Refinements 1, 2 and 3. Inequality (B.2) implies that $\Phi_i^* > \Phi_i^{**}$ for every $i \in \{1, 2\}$. According to Lemma B.2, the difference in the probability of conviction conditional on $(\theta_1, \theta_2) = (0, 0)$ and conditional on $(\theta_1, \theta_2) = (1, 0)$ is

$$(\Phi_1^* - \Phi_1^{**}) \Big((1 - \Phi_2^{**}) \big(q(1, 0) - q(0, 0) \big) + \Phi_2^{**} \big(q(1, 1) - q(0, 1) \big) \Big), \tag{B.3}$$

while the difference in the probability of conviction conditional on $(\theta_1, \theta_2) = (0, 1)$ and on $(\theta_1, \theta_2) = (1, 1)$ is

$$(\Phi_1^* - \Phi_1^{**}) \Big((1 - \Phi_2^{*}) \big(q(1, 0) - q(0, 0) \big) + \Phi_2^{*} \big(q(1, 1) - q(0, 1) \big) \Big). \tag{B.4}$$

Offenses are strategic substitutes if and only if (B.3) is less than (B.4) or, equivalently, if

$$(\Phi_1^* - \Phi_1^{**})(\Phi_2^* - \Phi_2^{**}) \Big(q(1, 0) + q(0, 1) - q(0, 0) - q(1, 1) \Big) < 0.$$

This inequality is equivalent to $q(1, 0) + q(0, 1) - q(0, 0) - q(1, 1) < 0$.

B.3 Proof of Lemma B.4

Agent $i \in \{1, 2\}$'s strategy is characterized by two functions $\omega_i^*(c)$ and $\omega_i^{**}(c)$, such that (1) when $\theta_i = 1$ and the realized cost is c_i, agent i chooses $a_i = 1$ when $\omega_i \leq \omega_i^*(c_i)$, and (2) when $\theta_i = 1$ and the realized

cost is c_i, agent i chooses $a_i = 1$ when $\omega_i \leq \omega_i^{**}(c_i)$. Let $q^* \equiv q(1,1) \in (0,1)$, we have

$$\omega_i^*(c) = 1 + c(1-\gamma) - \frac{c}{q^*\Phi_{-i}^{**}} \text{ and } \omega_i^{**}(c) = c(1-\gamma) - \frac{c}{q^*(\beta_i\Phi_{-i}^{**} + (1-\beta_i)\Phi_{-i}^*)}, \tag{B.5}$$

where $\beta_i \equiv \Pr(\theta_{-i} = 0|\theta_i = 0)$. Equation (B.5) implies that for every $c, c' > 0$, $\omega_i^*(c) > \omega_j^*(c)$ if and only if $\omega_i^*(c') > \omega_j^*(c')$, and $\omega_i^{**}(c) > \omega_j^{**}(c)$ if and only if $\omega_i^{**}(c') > \omega_j^{**}(c')$. By definition, $\Phi_i^* \equiv \int \Phi(\omega_i^*(c))dF(c)$ and $\Phi_i^{**} \equiv \int \Phi(\omega_i^{**}(c))dF(c)$.

Suppose by way of contradiction that the principal chooses $\boldsymbol{\theta} = (1,0)$ and $\boldsymbol{\theta} = (0,1)$ with different probabilities. Then $\beta_1 \neq \beta_2$. Lemma B.1 implies that the principal chooses $\boldsymbol{\theta} = (0,1)$ with positive probability. The principal's indifference condition implies that $\Phi_1^*/\Phi_1^{**} = \Phi_2^*/\Phi_2^{**}$. If $\omega_1^*(c) > \omega_2^*(c)$ for some $c > 0$, the first part of (B.5) implies that $\omega_1^{**}(c) < \omega_2^{**}(c)$ for every $c > 0$, and therefore, $\Phi_1^* > \Phi_2^*$ and $\Phi_1^{**} < \Phi_2^{**}$. Similarly, if $\omega_1^*(c) < \omega_2^*(c)$ for some $c > 0$, then $\Phi_1^* < \Phi_2^*$ and $\Phi_1^{**} > \Phi_2^{**}$. The conclusions of both cases contradict the hypothesis that $\Phi_1^*/\Phi_1^{**} = \Phi_2^*/\Phi_2^{**}$. If $\omega_1^*(c) = \omega_2^*(c)$ for some $c > 0$, the first part of (B.5) implies that $\omega_1^{**}(c) = \omega_2^{**}(c)$, and therefore, $\Phi_1^* = \Phi_2^*$ and $\Phi_1^{**} = \Phi_2^{**}$. Since $\Phi_1^* > \Phi_1^{**}$ and $\beta_1 \neq \beta_2$, this contradicts the second part of (B.5).

Given that $\boldsymbol{\theta} = (1,0)$ and $\boldsymbol{\theta} = (0,1)$ occurs with the same probability, we show that $\omega_1^*(c) = \omega_2^*(c)$ for every $c \geq 0$, which in turn implies that $\omega_1^{**}(\cdot) = \omega_2^{**}(\cdot)$. Suppose toward a contradiction that $\omega_1^*(c) > \omega_2^*(c)$ for some $c > 0$, then we have $\Phi_2^{**} > \Phi_1^{**}$, which implies that $\omega_2^{**}(c) > \omega_1^{**}(c)$ for every $c > 0$. As a result, $\Phi_1^*/\Phi_1^{**} > \Phi_2^*/\Phi_2^{**}$, which contradicts the principal's indifference condition $\Phi_1^*/\Phi_1^{**} = \Phi_2^*/\Phi_2^{**}$.

B.4 Proof of Lemma B.5

Since all equilibria that satisfy Refinements 1, 2, and 3 are symmetric, we omit footnotes i and $-i$ in order to simplify notation. Let $q^* \equiv q(1,1)$. We show that for every $c > 0$, $\omega^*(c) \to -\infty$ as $y \to 0$. The principal being indifferent between $\boldsymbol{\theta} = (0,0)$ and $\boldsymbol{\theta} = (1,0)$ implies that

$$y = q^*\Phi^{**}(\Phi^* - \Phi^{**}). \tag{B.6}$$

Suppose that there exists a sequence $\{y_n, \omega_n^*(\cdot), \omega_n^{**}(\cdot), q_n^*, \pi_n\}_{n=1}^{\infty}$ such that $\lim_{n\to+\infty} y_n = 0$, $(\omega_n^*(\cdot), \omega_n^{**}(\cdot), q_n^*, \pi_n)$ is an equilibrium under y_n for every $n \in \mathbb{N}$, and there exists $c > 0$ and $\omega^{**} \in \mathbb{R}$ such that $\limsup_{n\to\infty} \omega_n^{**}(c) = \omega^{**}$. Along the subsequence $\{k_n\}_{n\in\mathbb{N}}$ such that $\omega_{k_n}^{**}(c) \to \omega^{**}$, $\Phi(\omega_n^{**}(c))$ is bounded away from 0, which implies that

$$\Phi^{**} \equiv \int_0^{\overline{c}} \Phi(\omega^{**}(\widetilde{c}))dF(\widetilde{c})$$

is bounded away from 0. The principal's indifference condition (B.6) implies that either $\lim_{n\to\infty} q^*_{k_n} = 0$ or $\lim_{n\to\infty} \left(\omega^*_{k_n}(c) - \omega^{**}_{k_n}(c)\right) = 0$ for some $c > 0$. First, suppose $\lim_{n\to\infty} q^*_{k_n} = 0$, then (B.5) implies that $\omega^{**}_{k_n}(c)$ converges to $-\infty$, which contradicts the hypothesis that $\omega^{**}_{k_n}(c)$ converges to ω^{**}. Next, suppose $\lim_{n\to\infty} \left(\omega^*_{k_n}(c) - \omega^{**}_{k_n}(c)\right) = 0$ for some $c > 0$. Since $\omega^{**}_{k_n}(c)$ converges to ω^{**}, Φ^* and Φ^{**} are bounded away from 0 for every k_n. This implies that Φ^*/Φ^{**} converges to 1. Since $q^*_{k_n}$ cannot converge to 0 according to the previous step, equation (B.5) implies that $\lim_{n\to\infty} \left(\omega^*_{k_n}(c) - \omega^{**}_{k_n}(c)\right) = 1 > 0$, which leads to a contradiction.

Let π be the ex ante probability of crime. According to Lemma B.4, the principal chooses $\boldsymbol{\theta} = (1, 0)$ and $\boldsymbol{\theta} = (0, 1)$ with the same probability. Therefore, $\beta = \frac{1-\pi}{1-\pi/2}$. Recall the definition of $\mathcal{I}$. Lemma B.2 implies that $\Pr(\overline{\theta} = 1 | \boldsymbol{a} = (1, 1)) = \pi^*$, which implies that

$$\beta = \frac{2\mathcal{I}}{l^* + 2\mathcal{I}} \text{ and } 1 - \beta = \frac{l^*}{l^* + 2\mathcal{I}}. \tag{B.7}$$

Moreover, $\mathcal{I} = \frac{\Phi^* \cdot \Phi^{**}}{\Phi^{**} \cdot \Phi^{**}} = \frac{\Phi^*}{\Phi^{**}}$. Equation (B.2) implies that for every $c > 0$,

$$\left| \frac{\omega^*(c) - 1 - c(1-\gamma)}{\omega^{**}(c) - c(1-\gamma)} \right| = \frac{\beta\Phi^{**} + (1-\beta)\Phi^*}{\Phi^{**}} = \frac{(l^* + 2)\mathcal{I}}{l^* + 2\mathcal{I}}. \tag{B.8}$$

Since both $\omega^*(c)$ and $\omega^{**}(c)$ converge to $-\infty$ when $y \to 0$, and the difference between $\omega^*(c) - 1 - c(1-\gamma)$ and $\omega^{**}(c) - c(1-\gamma)$ is at most 1, the LHS of (B.8) converges to 1. Since the RHS of (B.8) is a strictly increasing function of $\mathcal{I}$ and equals 1 when $\mathcal{I} = 1$, we know that in the limit, the value of $\mathcal{I}$ is 1.

C Proof of Theorem 3

The following lemma shares a similar intuition with Lemma B.2, whose proof is in Appendix E.

Lemma C.1. *For every $\varepsilon > 0$, there exists $\overline{y} \in (0, 1)$ such that for every $y \in (0, \overline{y})$ and every equilibrium that satisfies Refinements 1, 2 and 3, we have* $\max\{q(1, 0), q(0, 1)\} < \varepsilon$.

Lemma C.1 together with the argument in Section 4.2 implies that crimes are uncorrelated, the principal's strategy is symmetric across agents, and the conviction probabilities are linear in the number of accusations. We derive the equilibrium probability of crime when y is small enough. Let $p \equiv \Pr(\theta_i = 1)$, and $q \equiv q(0, 1) = q(1, 0) = \frac{1}{2}q(1, 1)$. Agent i's reporting cutoffs are $\omega^*(c) = 1 - cK$ and $\omega^{**}(c) = -cK$, where

$$K \equiv -(1-\gamma) + \frac{1 - (1-\gamma)(p\Phi^* + (1-p)\Phi^{**})q}{q}. \tag{C.1}$$

Since $K \to +\infty$ as $y \to 0$, and furthermore, $\Phi^* = \int_0^{\overline{c}} \Phi(1 - cK) dF(c)$ and $\Phi^{**} = \int_0^{\overline{c}} \Phi(-cK) dF(c)$, we have

$$\lim_{y \to 0} \frac{\Phi^*}{\Phi^{**}} = \lim_{K \to +\infty} \frac{\int_0^{\overline{c}} \Phi(\omega^*(c)) dF(c)}{\int_0^{\overline{c}} \Phi(\omega^{**}(c)) dF(c)} = \frac{\lim_{K \to +\infty} \int_{-\infty}^{1} f(\frac{1-x}{K}) \Phi(x) dx}{\lim_{K \to +\infty} \int_{-\infty}^{0} f(\frac{-x}{K}) \Phi(x) dx}. \tag{C.2}$$

If $\int_{-\infty}^{0} \Phi(x) dx$ is finite, the dominated convergence theorem implies that

$$\frac{\lim_{K \to +\infty} \int_{-\infty}^{1} f(\frac{1-x}{K}) \Phi(x) dx}{\lim_{K \to +\infty} \int_{-\infty}^{0} f(\frac{-x}{K}) \Phi(x) dx} = \frac{\int_{-\infty}^{1} \Phi(x) dx \lim_{K \to +\infty} f(\frac{1-x}{K})}{\int_{-\infty}^{0} \Phi(x) dx \lim_{K \to +\infty} f(\frac{-x}{K})} = R. \tag{C.3}$$

If $\int_{-\infty}^{0} \Phi(x) dx = +\infty$, then

$$\frac{\Phi^* - \Phi^{**}}{\Phi^{**}} = \frac{\int_0^{\overline{c}} \Phi(1 - cK) dF(c)}{\int_0^{\overline{c}} \Phi(-cK) dF(c)} = \frac{\int_{-\overline{c}K}^{0} f(-\frac{x}{K}) (\Phi(1 + x) - \Phi(x)) dx}{\int_{-\overline{c}K}^{0} f(-\frac{x}{K}) \Phi(x) dx}.$$

Since f is a continuous strictly positive function on $[0, \overline{c}]$, $\underline{f} \equiv \min f$ and $\overline{f} \equiv \max f$ exist and are both strictly greater than 0. Therefore,

$$\lim_{K \to +\infty} \int_{-\overline{c}K}^{0} f(-\frac{x}{K}) \Phi(x) dx \geq \underline{f} \lim_{K \to +\infty} \int_{-\overline{c}K}^{0} \Phi(x) dx = +\infty,$$

$$\int_{-\overline{c}K}^{0} f(-\frac{x}{K}) (\Phi(1 + x) - \Phi(x)) dx \leq \overline{f} \lim_{K \to +\infty} (\Phi(1 + x) - \Phi(x)) dx = \overline{f} \int_{0}^{1} \Phi(x) dx < +\infty.$$

This implies that the limiting value of Φ^*/Φ^{**} is 1, which equals R when $\int_{-\infty}^{0} \Phi(x) dx = +\infty$. Since $q(1, 0) \in (0, 1)$ and θ_1 and θ_2 are uncorrelated, $\Pr(\theta_1 = 1 | a_1 = 1) = \pi^*$. According to Bayes rule,

$$\frac{\Pr(a_1 = 1 | \theta_1 = 1)}{\Pr(a_1 = 1 | \theta_1 = 0)} \cdot \frac{\Pr(\theta_1 = 1)}{1 - \Pr(\theta_1 = 1)} = \frac{\Pr(\theta_1 = 1 | a_1 = 1)}{1 - \Pr(\theta_1 = 1 | a_1 = 1)}, \tag{C.4}$$

which implies that $\Pr(\theta_1 = 1)$ converges to $\frac{\pi^{**}}{(1-\pi^{**})R + \pi^{**}}$ as $y \to 0$.

D Proof of Theorem 1

We establish several properties of $\overline{\pi}$-valid outcomes and optimal $\overline{\pi}$-valid outcomes. First, we show that the principal commits crime against each agent with positive probability under every $\overline{\pi}$-valid outcome.

Lemma D.1. *For every $\overline{\pi} \in (0, 1)$, $\Pr(\theta_i = 1) > 0$ for every i and every $\overline{\pi}$-valid outcome.*

The proof is similar to that of Lemma B.1, which we omit to avoid repetition. In order to show that the expected number of crime cannot be lower than $\min\{\overline{\pi}, \pi_{\min}(\overline{\pi})\}$, it is without loss of generality to focus

on equilibria in which the principal chooses $\boldsymbol{\theta} = (0,0)$ with positive probability. This together with Lemma D.1 implies that $\Pr(\theta_1 = 1) \in (0,1)$ and $\Pr(\theta_2 = 1) \in (0,1)$.

Next, we show that it is without loss of generality to focus on mechanisms that satisfy $q(0,0) = 0$. To see this, notice that the monotonicity refinement implies that $q(1,1) \geq \max\{q(1,0), q(0,1)\} \geq q(0,0)$. If $q(0,0) \neq 0$, consider the new mechanism defined by $q^*(\boldsymbol{a}) \equiv q(\boldsymbol{a}) - q(0,0)$. The principal's and the agents' incentives are identical under q and q^*. Moreover, since $\Pr(\boldsymbol{\theta} = (0,0)|\boldsymbol{a} = (0,0)) \leq \Pr(\boldsymbol{\theta} = (0,0)|\boldsymbol{a})$ for every $\boldsymbol{a}$, the fraction $\Pr(\boldsymbol{\theta} = (0,0)|s = 1)$ of wrongful convictions is weakly lower under q^* than under q.

Let $q(1,0) = q_1$, $q(0,1) = q_2$, $q(1,1) = q$ and $q(0,0) = 0$. Agent i's equilibrium strategy is characterized by two functions $\omega_i^* : [0,\overline{c}] \to \mathbb{R}$ and $\omega_i^{**} : [0,\overline{c}] \to \mathbb{R}$ such that when $\theta_i = 1$, agent i prefers $a_i = 1$ if and only if $\omega_i \leq \omega_i^*(c_i)$, when $\theta_i = 0$, agent i prefers $a_i = 1$ if and only if $\omega_i \leq \omega_i^{**}(c_i)$. Under conviction probabilities $(q, q_1, q_2, 0)$, we have $\omega_i^*(c) = 1 - cK_i^*$ and $\omega_i^{**}(c) = -cK_i^{**}$ where

$$K_i^* \equiv -1 + \gamma + \frac{1 - (1-\gamma)q_j\mathbb{E}[\Phi_j|\theta_i = 1]}{(q - q_j)\mathbb{E}[\Phi_j|\theta_i = 1] + q_i(1 - \mathbb{E}[\Phi_j|\theta_i = 1])} \tag{D.1}$$

$$K_i^{**} \equiv -1 + \gamma + \frac{1 - (1-\gamma)q_j\mathbb{E}[\Phi_j|\theta_i = 0]}{(q - q_j)\mathbb{E}[\Phi_j|\theta_i = 0] + q_i(1 - \mathbb{E}[\Phi_j|\theta_i = 0])} \tag{D.2}$$

and Φ_j stands for the probability that $a_j = 1$ which is a convex combination of $\Phi_j^* \equiv \mathbb{E}[\Phi_j|\theta_j = 1]$ and $\Phi_j^{**} \equiv \mathbb{E}[\Phi_j|\theta_j = 0]$. By definition, $\Phi_j^* \equiv \int_0^{\overline{c}} \Phi(\omega_j^*(c))dF(c)$ and $\Phi_j^{**} \equiv \int_0^{\overline{c}} \Phi(\omega_j^{**}(c))dF(c)$. One can verify that $K_i^* < K_i^{**}$ and $\omega_i^*(c) - \omega_i^{**}(c) > 1$ when θ_1 and θ_2 are positively correlated and that $K_i^* > K_i^{**}$ and $\omega_i^*(c) - \omega_i^{**}(c) < 1$ when θ_1 and θ_2 are negatively correlated.

Lemma D.2. *For every $\varepsilon > 0$, there exists $\overline{y}_\varepsilon > 0$ such that when $y < \overline{y}_\varepsilon$, $\max\{q_1, q_2\} < \varepsilon$ in every monotone mechanism that can implement some $\overline{\pi}$-valid outcome. Moreover, for every $\overline{\pi}$-valid outcome, $\max\{\Phi_1^*, \Phi_2^*, \Phi_1^{**}, \Phi_2^{**}\} < \varepsilon$.*

The proof is in Appendix E. Recall that the principal's choices of θ_1 and θ_2 are strategic substitutes if $Q > 0$, and are strategic complements if $Q < 0$. This implies that

1. When $q(0,0) + q(1,1) - q(1,0) - q(0,1) > 0$, then the principal chooses $\boldsymbol{\theta} = (0,0), (1,0), (0,1)$ with positive probability, and $\boldsymbol{\theta} = (1,1)$ with zero probability.

2. When $q(0,0) + q(1,1) - q(1,0) - q(0,1) < 0$, then the principal chooses $\boldsymbol{\theta} = (0,0), (1,1)$ with positive probability, and chooses either $\boldsymbol{\theta} = (1,0)$ or $\boldsymbol{\theta} = (0,1)$ or both with zero probability.

We partition the set of monotone mechanisms that can implement $\overline{\pi}$-valid outcomes into two subsets.

Strategic Substitutes: When $q > q_1 + q_2$, we show that $\pi_1 + \pi_2$ is close to $\overline{\pi}$ when y is small enough.

Lemma D.3. *For every $\varepsilon > 0$, there exists $\overline{y}_\varepsilon > 0$ such that when $y < \overline{y}_\varepsilon$, we have $\pi_1 + \pi_2 \geq \overline{\pi} - \varepsilon$ for every $i \in \{1, 2\}$ under every $\overline{\pi}$-valid outcome.*

Proof. Let $X_i^{**} \equiv 1 - (1-\gamma)q_{-i}\Phi_{-i}^{**}$, $X_i^{*} \equiv 1 - (1-\gamma)q_{-i}\Phi_{-i}^{*}$, $Y_i^{**} \equiv (q - q_1 - q_2)\Phi_{-i}^{**} + q_i$, and $Y_i^{*} \equiv (q - q_1 - q_2)\Phi_{-i}^{*} + q_i$. Let $\pi_i \equiv \Pr(\theta_i = 1)$. Equations (F.1) and (F.2) imply that for every $c \in [0, \overline{c}]$,

$$\Big|\frac{\omega_i^*(c) - 1 - c(1-\gamma)}{\omega_i^{**}(c) - c(1-\gamma)}\Big| = \frac{X_i^{**}(\pi_{-i}Y_i^{*} + (1-\pi_1-\pi_2)Y_i^{**})}{Y_i^{**}(\pi_{-i}X_i^{*} + (1-\pi_1-\pi_2)X_i^{**})} = 1 + \frac{\pi_{-i}}{\pi_{-i}X_i^{*} + (1-\pi_1-\pi_2)X_i^{**}} \cdot \frac{X_i^{**}Y_i^{*} - X_i^{*}Y_i^{**}}{Y_i^{**}}, \tag{D.3}$$

with

$$\frac{X_i^{**}Y_i^{*} - X_i^{*}Y_i^{**}}{Y_i^{**}} = \frac{(\Phi_{-i}^{*} - \Phi_{-i}^{**})(q - q_1 - q_2 + (1-\gamma)q_1q_2)}{(q - q_1 - q_2)\Phi_{-i}^{**} + q_i}.$$

Since the LHS of (D.3) converges to 1 as $y \to 0$, the RHS also converges to 0, which implies that

$$\frac{\pi_{-i}}{\pi_{-i}X_i^{*} + (1-\pi_1-\pi_2)X_i^{**}} \cdot \frac{(\Phi_{-i}^{*} - \Phi_{-i}^{**})(q - q_1 - q_2 + (1-\gamma)q_1q_2)}{(q - q_1 - q_2)\Phi_{-i}^{**} + q_i}$$

converges to 0. The principal's indifference between $\boldsymbol{\theta} = (0, 1)$ and $\boldsymbol{\theta} = (1, 0)$ translates into

$$(\Phi_1^{*} - \Phi_1^{**})\Big\{(q - q_1 - q_2)\Phi_2^{**} + q_1\Big\} = (\Phi_2^{*} - \Phi_2^{**})\Big\{(q - q_1 - q_2)\Phi_1^{**} + q_2\Big\}, \tag{D.4}$$

which implies that

$$\frac{X_1^{**}Y_1^{*} - X_1^{*}Y_1^{**}}{Y_1^{**}} = \frac{X_2^{**}Y_2^{*} - X_2^{*}Y_2^{**}}{Y_2^{**}}. \tag{D.5}$$

Let $\mathcal{I}_i \equiv \frac{\Phi_1^{*}}{\Phi_1^{**}}$. Since

$$\max\Big\{\mathcal{I}_1, \mathcal{I}_2\Big\} \leq R, \quad \text{and} \quad \frac{\pi_1 + \pi_2}{1 - \pi_1 - \pi_2}\Big(\frac{\pi_1}{\pi_1 + \pi_2}\frac{\Phi_1^{*}}{\Phi_1^{**}} + \frac{\pi_2}{\pi_1 + \pi_2}\frac{\Phi_2^{*}}{\Phi_2^{**}}\Big) = \frac{\overline{\pi}}{1 - \overline{\pi}}, \tag{D.6}$$

$\pi_1 + \pi_2$ is bounded away from 0. Therefore,

$$\max\Big\{\frac{\pi_1}{\pi_1 X_2^{*} + (1-\pi_1-\pi_2)X_2^{**}}, \frac{\pi_2}{\pi_2 X_1^{*} + (1-\pi_1-\pi_2)X_1^{**}}\Big\}$$

is bounded away from 0. This implies that at least one of the expressions $\frac{(\Phi_1^{*} - \Phi_1^{**})(q - q_1 - q_2 + (1-\gamma)q_1q_2)}{(q - q_1 - q_2)\Phi_1^{**} + q_2}$ and $\frac{(\Phi_2^{*} - \Phi_2^{**})(q - q_1 - q_2 + (1-\gamma)q_1q_2)}{(q - q_1 - q_2)\Phi_2^{**} + q_1}$ converges to 0 as $y \to 0$. According to (F.5), the two expressions are equal,

and therefore, both of them converge to 0 as $y \to 0$. Since

$$\frac{(\Phi_1^* - \Phi_1^{**})(q - q_1 - q_2 + (1-\gamma)q_1q_2)}{(q - q_1 - q_2)\Phi_1^{**} + q_2} = \Big(q - q_1 - q_2 + (1-\gamma)q_1q_2\Big)\frac{\mathcal{I}_1 - 1}{(q - q_1 - q_2) + \frac{q_2}{\Phi_1^{**}}},$$

we have $\mathcal{I}_1 \to 1$ unless

$$\frac{q - q_1 - q_2 + (1-\gamma)q_1q_2}{(q - q_1 - q_2) + \frac{q_2}{\Phi_1^{**}}} \tag{D.7}$$

converges to 0, which happens if and only if $\frac{q_2}{\Phi_1^{**}}$ converges to infinity. Since $\omega_1^*(c) = 1 - cK_1^*$ with

$$K_1^* = -(1-\gamma) + \frac{1 - (1-\gamma)q_2\Phi_2^{**}}{(q - q_1 - q_2)\Phi_2^{**} + q_1},$$

this requires that $q_2/q_1 \to +\infty$ and $q_2/(q - q_1 - q_2) \to +\infty$. Plugging this into (F.4), we have $\frac{\Phi_1^* - \Phi_1^{**}}{\Phi_2^* - \Phi_2^{**}} \to +\infty$, which contradicts the requirement that $K_1^* > K_2^*$. Therefore, $\mathcal{I}_1 \to 1$ and $\mathcal{I}_2 \to 1$ as $y \to 0$. Equation (D.6) implies that $\pi_1 + \pi_2$ is close to $\overline{\pi}$ when $y \to 0$. □

Strategic Complements: When $q < q_1 + q_2$, we show that $\max\{\frac{\Phi_1^*}{\Phi_1^{**}}, \frac{\Phi_2^*}{\Phi_2^{**}}\} \le R + \varepsilon$. This together with the fact that θ_1 and θ_2 are either positively correlated or uncorrelated implies that the expected number of crimes is at least $\pi_{\min}(\overline{\pi})$. Equation (F.3) implies that when y is close to 0

$$\frac{\Phi_j^*}{\Phi_j^{**}} = \frac{\int_0^{\overline{c}} \Phi(\omega_j^*(c))dF(c)}{\int_0^{\overline{c}} \Phi(\omega_j^{**}(c))dF(c)} = \frac{K_j^{**}}{K_j^*} \cdot \frac{\int_{-\infty}^1 f(\frac{1-x}{K_j^*})\Phi(x)dx}{\int_{-\infty}^0 f(\frac{-x}{K_j^{**}})\Phi(x)dx}. \tag{D.8}$$

Since $f(\frac{1-x}{K_j^*})\Phi(x) \le \Phi(x)\sup_{c\in[0,\overline{c}]} f(c)$ and $\int_{-\infty}^0 \Phi(x)dx$ is finite, the dominated convergence theorem implies that

$$\lim_{k\to+\infty} \int_{-\infty}^1 f(\frac{1-x}{k})\Phi(x)dx = \int_{-\infty}^1 \lim_{k\to+\infty} f(\frac{1-x}{k})\Phi(x)dx = \lim_{c\downarrow 0} f(c) \int_{-\infty}^1 \Phi(x)dx, \tag{D.9}$$

$$\lim_{k\to+\infty} \int_{-\infty}^0 f(-\frac{x}{k})\Phi(x)dx = \int_{-\infty}^0 \lim_{k\to+\infty} f(-\frac{x}{k})\Phi(x)dx = \lim_{c\downarrow 0} f(c) \int_{-\infty}^0 \Phi(x)dx. \tag{D.10}$$

Therefore,

$$\lim_{k\to+\infty} \frac{\int_{-\infty}^1 f(\frac{1-x}{K_j^*})\Phi(x)dx}{\int_{-\infty}^0 f(\frac{-x}{K_j^{**}})\Phi(x)dx} = R.$$

Since Φ_j^* and Φ_j^{**} converge to 0 as $y \to 0$, and $q - q_j < q_i$, the ratio between K_i^{**} and K_i^* converge to

$$\frac{(q-q_j)\mathbb{E}[\Phi_j|\theta_i=1]+q_i(1-\mathbb{E}[\Phi_j|\theta_i=1])}{(q-q_j)\mathbb{E}[\Phi_j|\theta_i=0]+q_i(1-\mathbb{E}[\Phi_j|\theta_i=0])} \to 1. \tag{D.11}$$

Therefore, for every $\varepsilon > 0$, there exists $\overline{y} > 0$ such that $\frac{\Phi_j^*}{\Phi_j^{**}} \leq R + \varepsilon$ when $y < \overline{y}$.

Expected Number of Crimes: We compute the minimal expected number of crimes among all $\overline{\pi}$-valid outcomes. When y is small enough, the expected number of crimes is close to $1 - \overline{\pi}$ according to every $\overline{\pi}$-valid outcome with $q > q_1 + q_2$, and the expected number of crimes is close to $\frac{2\pi^{**}}{(1-\pi^{**})R+\pi^{**}}$, where π^{**} is given by (4.5). In the online appendix, we solve (4.5) and obtain:

$$\pi^{**} = \frac{2R\bar{l} + R + 1 - \sqrt{(R+1)^2 + 4R\bar{l}}}{2R(\bar{l}+1)}.$$

Plugging this expression into the expected number of crimes, we verify in the online appendix that $\frac{2\pi^{**}}{(1-\pi^{**})R+\pi^{**}} = \pi_{\min}(\overline{\pi})$.

E Proofs of Lemmas B.2, C.1 and D.2

We show that for every $\varepsilon > 0$, there exists $\overline{y} > 0$ such that for every $y \in (0, \overline{y})$, every monotone mechanism that implements some $\overline{\pi}$-valid outcome satisfies $\max\{q(1,0), q(0,1)\} < \varepsilon$.

Suppose by way of contradiction that there exists $\eta > 0$ such that for every $\overline{y} > 0$, there exists $y \in (0, \overline{y})$ and a monotone mechanism satisfying $\max\{q(1,0), q(0,1)\} \geq \eta$ and can implement some $\overline{\pi}$-valid outcome. Since the principal commits crime against each agent with positive probability, we have

$$y \geq \min(\Phi_1^* - \Phi_1^{**})\Big\{\big(q(1,0)(1-\Phi_2^{**})\big)+\big(q(1,1)-q(0,1)\big)\Phi_2^{**}, \big(q(1,0)(1-\Phi_2^*)\big)+\big(q(1,1)-q(0,1)\big)\Phi_2^*\Big\},$$

$$y \geq \min(\Phi_2^* - \Phi_2^{**})\Big\{\big(q(0,1)(1-\Phi_1^{**})\big)+\big(q(1,1)-q(1,0)\big)\Phi_1^{**}, \big(q(0,1)(1-\Phi_1^*)\big)+\big(q(1,1)-q(1,0)\big)\Phi_1^*\Big\}.$$

Suppose that $q(1,0)$ is bounded away from 0 no matter how small y is, i.e., there exists $\eta > 0$ such that $q(1,0) \geq \eta$. Since $\Phi_i^* \leq \Phi(b)$ for every $i \in \{1,2\}$, we know that both K_1^* and K_1^{**} are bounded from below. Since $\Phi_1^* = \int_0^{\overline{c}} \Phi(1 - cK_1^*)dF(c)$ and $\Phi_1^{**} = \int_0^{\overline{c}} \Phi(-cK_1^{**})dF(c)$, Φ_1^* and Φ_1^{**} are both bounded away from 0. Since $q(1,0)(1-\Phi_2^{**})$ and $q(1,0)(1-\Phi_2^*)$ are both bounded away from 0, the inequalities that characterize the principal's incentive implies that $\Phi_1^* - \Phi_1^{**}$ converges to 0 as $y \to 0$. Since Φ_1^* and Φ_1^{**}

are bounded away from 0, Φ_1^*/Φ_1^{**} converges to 1, and therefore, $K_2^* - K_2^{**} \to 0$. Hence, either both K_2^* and K_2^{**} diverge to $-\infty$, or $\Phi_2^* - \Phi_2^{**}$ is bounded away from 0.

Consider two subcases. Suppose $q(0,1)$ does not converge to 0, then for the principal's incentive constraints to hold, it must be the case that $\Phi_2^* - \Phi_2^{**}$ converges to 0. Our previous conclusion suggests that both K_2^* and K_2^{**} diverge to $-\infty$. However, the expressions for K_2^* and K_2^{**} suggest that neither of them diverge to $-\infty$ when $q(0,1)$ is bounded away from 0, which leads to a contradiction.

Suppose next that $q(0,1)$ converges to 0. We show that $q(1,1) - q(1,0) \to 0$. Suppose by way of contradiction that $q(1,1) - q(1,0)$ is bounded away from 0 along some subsequence of y. Then $(q(1,1) - q(1,0))\Phi_1^{**}$ is bounded away from 0, so both K_2^* and K_2^{**} are bounded from below. Since $K_2^* - K_2^{**} \to 0$, we have $\Phi_2^* - \Phi_2^{**}$ is bounded away from 0. The marginal incentive to commit crime against agent 2 is

$$(\Phi_2^* - \Phi_2^{**})\Big(q(0,1)(1-\Phi_1^{**}) + (q(1,1) - q(1,0))\Phi_1^{**}\Big)$$

is bounded away from 0, which leads to a contradiction.

If $q(1,1)$ and $q(1,0)$ are bounded away from 0 with $q(1,1) - q(1,0) \to 0$, while $q(0,1) \to 0$. Recalling the expressions for K_1^* and K_1^{**} in Appendix A, we know that $K_1^* - K_1^{**} \to 0$. Since K_1^* is bounded, we have

$$\Phi_1^* - \Phi_1^{**} = \int_0^{\overline{c}} \Big(\Phi(1 - cK_1^*) - \Phi(-cK_1^{**})\Big) dF(c) \tag{E.1}$$

which is bounded away from 0 when $K_1^* - K_1^{**} \to 0$. This contradicts the previous conclusion that $\Phi_1^* - \Phi_1^{**}$ converges to 0. Similarly, $q(0,1)$ cannot be bounded away from 0 when y is small enough, and that $\Phi_1^*, \Phi_2^*, \Phi_1^{**}$, and Φ_2^{**} must converge to 0 as y approaches 0. This shows Lemmas C.1 and D.2

We complete the proof of Lemma B.2 by ruling out cases in which both $q(1,0)$ and $q(0,1)$ converge to 0 as $y \to 0$, but $q(1,1) = 1$. When $q(0,1) = q(1,0) = 0$ and $q(1,1) = 1$, the argument in Lemma B.4 implies that every equilibrium that satisfies Refinments 1-3 must be symmetric, and hence,

$$\Phi_1^* \leq \int_0^{\overline{c}} \Phi\Big(1 + c(1-\gamma) - \frac{c}{\Phi_1^*}\Big) dF(c). \tag{E.2}$$

Every strictly positive fixed point of $\Phi_1^* = \int_0^{\overline{c}} \Phi\Big(1 + c(1-\gamma) - \frac{c}{\Phi_1^*}\Big) dF(c)$ has Φ_1^* bounded away from 0, which contradicts our previous conclusion that Φ_1^* converges to 0 as $y \to 0$. If $q(1,0) > 0$ or $q(0,1) > 0$ or both, then Φ_1^* increases compared to the case in which $q(1,0) = q(0,1) = 0$, which means that it is also bounded away from 0, which leads to a contradiction.

F Proof of Proposition 2

Single-Agent Benchmark: Let $q(1)$ be the probability of conviction when $a = 1$. The reporting cutoffs are:

$$\omega^*(c,1) \equiv 1 + c(1-\gamma) - \frac{c}{q(1)} \text{ and } \omega^{**}(c,1) \equiv c(1-\gamma) - \frac{c}{q(1)}, \tag{F.1}$$

which implies that first, $\omega^*(c,1) = \omega^{**}(c,1) + 1$, and second, $\omega^*(c,1)$ and $\omega^{**}(c,1)$ are both decreasing functions of c. Therefore,

$$\mathcal{I} \equiv \frac{\Pr(a=1|\theta=1)}{\Pr(a=1|\theta=0)} = \frac{\int_0^{\overline{c}} \Phi\Big(1 + c(1-\gamma) - \frac{c}{q(1)}\Big) dF(c)}{\int_0^{\overline{c}} \Phi\Big(c(1-\gamma) - \frac{c}{q(1)}\Big) dF(c)} = \frac{\int_0^{\overline{c}} \Phi(1-ck) dF(c)}{\int_0^{\overline{c}} \Phi(-ck) dF(c)} \tag{F.2}$$

where $k \equiv 1 - \gamma - \frac{1}{q(1)}$. Since the principal chooses $\theta = 1$ with positive probability, we have

$$y \geq q(1) \int \Big(\Phi\Big(1 + c(1-\gamma) - \frac{c}{q(1)}\Big) - \Phi\Big(c(1-\gamma) - \frac{c}{q(1)}\Big)\Big) dF(c).$$

When y is small enough, $q(1) \in (0,1)$, which implies that the posterior probability of $\overline{\theta} = 1$ after observing $a = 1$ is π^*. The probability of crime π satisfies

$$\frac{\pi}{1-\pi} \cdot \frac{\Pr(a=1|\theta=1)}{\Pr(a=1|\theta=0)} = \frac{\pi^*}{1-\pi^*} \quad \Rightarrow \quad \pi = \frac{\pi^*}{(1-\pi^*)\mathcal{I} + \pi^*}. \tag{F.3}$$

The value of $\mathcal{I}$ when $y \to 0$ is R following the same argument as the proof of Theorem 2.

Comparative Statics: Let $\overline{y} \in \mathbb{R}_+$ be such that for every $y < \overline{y}$, an equilibrium that satisfies Refinements 1, 2, and 3 exists both in the single-agent benchmark and in the two-agent case. When there are n agents, let $\Phi^*(n)$ be the probability that an agent chooses $a = 1$ conditional on $\theta = 1$, let $\Phi^{**}(n)$ be the probability that an agent chooses $a = 1$ conditional on $\theta = 0$, and let $\omega^*(c,n)$ and $\omega^{**}(c,n)$ be the reporting cutoffs. Equations (F.1) and (B.5) imply that for every $c > 0$, the sign of $\omega^*(c,1) - \omega^*(c,2)$ coincides with the sign of $q(1) - q(2)\Phi^{**}(2)$. As a result, $\Phi^*(1) < \Phi^*(2)$ if and only if $q(1) < q(2)\Phi^{**}(2)$.

Suppose toward a contradiction that $q(1) \geq q(2)\Phi^{**}(2)$. The principal's indifference condition implies:

$$q(2)\Phi^{**}(2)\Big(\Phi^*(1) - \Phi^{**}(1)\Big) \leq q(1)\Big(\Phi^*(1) - \Phi^{**}(1)\Big) = y = q(2)\Phi^{**}(2)\Big(\Phi^*(2) - \Phi^{**}(2)\Big). \tag{F.4}$$

As a result,

$$\Phi^*(1) - \Phi^{**}(1) \leq \Phi^*(2) - \Phi^{**}(2). \tag{F.5}$$

Recall that under Condition 1, $\phi(\omega)$ is strictly increasing when $\omega < \overline{\omega}$. For every $c > 0$ such that $\omega^*(c,1) < \overline{\omega}$, since $\omega^*(c,1) - \omega^{**}(c,1) = 1$ and $\omega^*(c,2) - \omega^{**}(c,2) < 1$, we have $\Phi(\omega^*(c,1)) - \Phi(\omega^{**}(c,1)) > \Phi(\omega^*(c,2)) - \Phi(\omega^{**}(c,2))$. Since both $q(1)$ and $q(2)$ converge to 0 as $y \to 0$, we know that for every $c^* > 0$, there exists $\overline{y}(c^*) > 0$ such that when $y < \overline{y}(c^*)$, $\omega^*(c,1) < \overline{\omega}$ and $\omega^*(c,2) - \omega^{**}(c,2) < 1/2$ for every $c \geq c^*$. Choosing c^* so that $F(c^*)$ is small enough, we have $\Phi^*(1) - \Phi^{**}(1) > \Phi^*(2) - \Phi^{**}(2)$ for every $y < \overline{y}(c^*)$, which leads to a contradiction and implies that $\Phi^*(1) < \Phi^*(2)$.

Suppose toward a contradiction that $\Phi^{**}(1) \geq \Phi^{**}(2)$, then $\Phi^*(1)/\Phi^{**}(1) < \Phi^*(2)/\Phi^{**}(2)$. When y is below some cutoff, $\frac{\Pr(a=1|\theta=1)}{\Pr(a=1|\theta=0)} = \frac{\Phi^*(1)}{\Phi^{**}(1)} \geq R > 0$ and $\frac{\Pr(\boldsymbol{a}=(1,1)|\overline{\theta}=1)}{\Pr(\boldsymbol{a}=(1,1)|\overline{\theta}=0)} = \frac{\Phi^*(2)}{\Phi^{**}(2)} \leq 1 + \varepsilon$. This leads to a contradiction and shows that $\Phi^{**}(1) < \Phi^{**}(2)$.

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] Argenziano, Rossella, Sergei Severinov and Francesco Squintani (2016) "Strategic Information Acquisition and Transmission," *American Economic Journal-Microeconomics*, 8(3), 119-155.

[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 (2020) "Deterrence with Imperfect Attribution," *American Political Science Review*, forthcoming.

[7] Baliga, Sandeep and Tomas Sjöström (2004) "Arms Races and Negotiations," *Review of Economic Studies*, 71(2), 351-369.

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

[9] Bar-Hillel, Maya (1984) "Probabilistic Analysis in Legal Factfinding," *Acta Psychologica*, 56, 267-284.

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

[11] Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992) "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades," *Journal of Political Economy*, 100, 992-1026.

[12] Carlsson, Hans and Eric van Damme (1993) "Global Games and Equilibrium Selection," *Econometrica*, 61(5), 989-1018.

[13] Cheng, Ing-Haw and Alice Hsiaw (2020) "Reporting Sexual Misconduct in the MeToo Era," Working Paper.

[14] Cohen, Jonathan (1977) "The Probable and The Provable," Oxford University Press.

[15] Deimen, Inga and Dezsö Szalay (2019) "Delegated Expertise, Authority, and Communication," *American Economic Review*, 109(4), 1349-1374.

[16] Demougin, Dominique and Claude Fluet (2016) "Preponderance of Evidence," *European Economic Review*, Vol 50(4), 963-976.

[17] Dobbie, Will, Jacob Goldin, and Crystal S. Yang (2018) "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 108(2), 201-240.

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

[19] 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.

[20] Harel, Alon and Ariel Porat (2009) "Aggregating Probabilities Across Cases: Criminal Responsibility for Unspecified Offenses," *Minnesota Law Review*, 482, 261-310.

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

[22] Lee, Frances Xu and Wing Suen (2020) "Credibility of Crime Allegations," *American Economic Journal-Microeconomics*, 12, 220-259.

[23] Lynch, Gerard (1987) "RICO: The Crime of Being a Criminal," *Columbia Law Review*, 87(4), 661-764.

[24] Morris, Stephen and Hyun Song Shin (1998) "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review*, 88(3), 587-597.

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

[26] Naess, Ole-Andreas Elvik (2020) "Under-reporting of Crime," Working Paper.

[27] Pei, Harry (2015) "Communication with Endogenous Information Acquisition," *Journal of Economic Theory*, 160, 132-149.

[28] Pei, Harry and Bruno Strulovici (2020) "Crime Aggregation, Deterrence, and Witness Credibility," arxiv preprint arXiv:2009.06470.

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

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

[31] Schauer, Federick and Richard Zeckhauser (1996) "On the Degree of Confidence for Adverse Decisions," *Journal of Legal Studies*, 25(1), 27-52.

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

[33] Siegel, Ron and Bruno Strulovici (2020) "Judicial Mechanism Design," Working Paper.

[34] Silva, Francesco (2019) "If We Confess Our Sins," *International Economic Review*, 60(3), 1389–1412.

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

[36] Spier, Kathryn (1994) "Pretrial Bargaining and the Design of Fee-Shifting Rules," *The RAND Journal of Economics*, 25(2),197-214.

[37] Strulovici, Bruno (2010) "Learning while Voting: Determinants of Collective Experimentation," *Econometrica*, 78(3), 933–971.

[38] Strulovici, Bruno (2020) "Can Society Function without Ethical Agents? An Informational Perspective," Working Paper, Northwestern University.

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

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