Crime Aggregation, Deterrence, and Witness Credibility*

Harry Pei Bruno Strulovici

January 4, 2021

Abstract

We present a model for the equilibrium frequency of crimes and the informativeness of witness testimonies when potential offenders can commit multiple crimes and witnesses have heterogeneous retaliation risk and reporting preferences. Assuming that a potential offender's benefit from committing crime is small relative to the punishment incurred if he is convicted, we characterize the mechanism that minimizes the expected number of crimes subject to an upper bound on the fraction of wrongful convictions. The optimal commitment outcome can be approximated by an equilibrium without commitment, in which a Bayesian judge uses one of the following conviction rules, depending on parameters of the model: (1) Convict the defendant when the probability that he has committed at least one, unspecified offence exceeds some threshold; (2) Convict the defendant when the probability that he has committed a *specific* offence exceeds some threshold.

Keywords: deterrence, adjudication rule, wrongful conviction, witness testimony.

JEL Codes: D82, D83, K42.

*We thank S. Nageeb Ali, Bocar Ba, Arjada Bardhi, Laura Doval, Mehmet Ekmekci, Alex Frankel, Andrei Gomberg, Yingni Guo, Marina Halac, Andreas Kleiner, Anton Kolotinin, Frances Xu Lee, Alessandro Pavan, Joao Ramos, Joyce Sadka, Ron Siegel, Vasiliki Skreta, Takuo Sugaya, Saturo Takahashi, Teck Yong Tan, Alex Wolizky, Siyang Xiong, Boli Xu, and our seminar participants for helpful comments, and 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. While this separation of charges is standard, its desirability for deterrence and fairness is by no means obvious. Consider a defendant who may have committed two offenses with probability 0.8 each, independent of each other. If the conviction threshold for each offense is 0.9, then the defendant is 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), the 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"), that a defendant is convicted if the probability that he has committed *some unspecified offense* exceeds a given threshold. They compare APP with the commonly used *Distinct Probabilities Principle* ("DPP"), which requires that the defendant be convicted only if the probability that he has committed some specific criminal behavior exceeds a threshold. They argue that APP can reduce adjudication errors, improve deterrence, and reduce the cost of enforcement, and advocate using APP to varying degrees in both civil and criminal lawsuits. Similar arguments appear in Schauer and Zeckhauser (1996), who wrote: "*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...*"

These studies all assume that the strength of each piece evidence is exogenously given, irrespective of the rule used to aggregate charges, and they all assume that the defendant's guilt on each offense is *independently distributed* across charges. These studies thus ignore how aggregation rules must affect the incentives of potential offenders to commit offenses and the incentives of witnesses to report offenses.

We study a model in which adjudication rules affect the incentives and behaviors of potential offenders and witnesses. Our results shed light on effects of APP and DPP in deterring crimes. We also provide a justification for using APP or DPP in criminal justice systems by showing that the equilibrium outcome under one of these conviction rules (depending on the parameters of the model) coincides with the outcome that minimizes the expected number of crimes when the judge can commit to a mechanism in advance and

faces an upper bound on the fraction of wrongful convictions.[1]

In our model, a potential offender (hereafter, *principal*) has two opportunities to commit crimes.[2] His decision trades off the benefit from committing crimes against the punishment from conviction. Each crime is associated with a distinct witness (hereafter, *agent*) who observes whether that crime takes place. To illustrate, the principal could be a manager with multiple opportunities of violating the law or abusing his subordinates, and agents would correspondingly be whistleblowers or victims who witness these violations whenever they occur. Each agent decides whether to accuse the principal based on three considerations: (1) a preference for punishing offenders, (2) some idiosyncratic private benefits or costs of getting the principal convicted, and (3) a cost of filing accusations, which comes from the principal's retaliation and is strictly higher when his accusation fails to convict the principal.[3] Finally, a judge observes both 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 and when it occurs, is large relative to the benefit from committing crime. This assumption is arguably satisfied 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 getting convicted. 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 mapping from agents' reports to the probability of conviction (i.e., a *mechanism*) in order to minimize the expected number of crimes subject to an upper bound on the fraction of wrongful convictions.[4] Theorem 1 characterizes the lowest expected number of crimes, and shows that under the optimal mechanism, the probability of conviction is linear in the number of accusations when the designer can tolerate a high fraction of wrongful convictions, and the principal is convicted only if both agents accuse him in the complementary scenario where the designer has a low tolerance for wrongful convictions.

Intuitively, when the conviction probability is convex in the number of accusations (e.g., convict only if both agents accuse the principal), the principal's decisions to commit different crimes are *strategic*

[1] In our model, the fraction of mistaken acquittals equals the (unconditional) probability of crime. Since our objective is to minimize crime, Wrongful convictions are therefore the only type of adjudication error that we need to concern ourselves with.

[2] In our working paper Pei and Strulovici (2020), which the present paper subsumes, we consider the case of an arbitrary number of agents in a related model, and also allow the principal to have private information about his benefit from committing crime.

[3] 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 of police brutality or inaction by Ba (2018) and Ba and Rivera (2019). Similar patterns arise in a 2016 survey conducted by the USMSPB, which concluded that 21% of women and 8.7% of men experienced at least one of 12 categorized behaviors of sexual harassment, of which only 16% led to merit resolutions.

[4] We take the punishment from conviction as exogenous instead of letting the mechanism designer choosing it endogenously. This is motivated by applications where the negative consequences of conviction mostly come from its undesirable side effects that are beyond the judge's or the mechanism designer's control (e.g., ending a promising career, losing a high-paying job or a good reputation) rather than the sentencing itself.

substitutes. Therefore, agents' private observations of crimes are *negatively correlated*. This has two effects on the expected number of crimes: it eliminates the possibility of having multiple crimes but also leads to a high probability of having at least one crime. The latter is because agents' decisions to report crimes are *strategic complements*. Since agents' private observations of crimes are negatively correlated, each agent's coordination motive discourages him to accuse the principal when he has witnessed a crime and vice versa. This lowers the credibility of accusations, which in equilibrium, increases the probability that the principal commits crime. By contrast, different crimes are uncorrelated when the conviction probability is linear in the number of accusations. The principal commits multiple crimes with positive probability, but agents' coordination motives can no longer undermine the informativeness of accusations.

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. Which type of mechanism is optimal depends on the effectiveness of linear conviction probabilities in improving the informativeness of accusations and the effect of improved informativeness on the probability of crime. These in turn depend on the standard of proof and the fraction of agents who are prone to make false accusations.[5]

Next, we examine the equilibrium outcomes when there is no commitment and the judge adjudicates guilt according to APP or DPP based on her posterior belief about crime. We show that the optimal commitment outcome is attained in every equilibrium under APP when the conviction probability is convex under the optimal mechanism, and is attained in every equilibrium under DPP otherwise. Intuitively, more agents accusing the principal strictly increases the probability that the principal is guilty of at least one crime but does not necessarily increase the probability that he is guilty of a specific crime. For example, when the two crimes are negatively correlated, an agent's accusation decreases the likelihood that the principal has committed crime against the other agent. We use this observation to show that when the benefit from committing crime is small enough, the conviction probability is convex under APP and is linear under DPP. As a result, APP induces a negative correlation between different crimes, which compares to DPP, eliminates the possibility of having multiple crimes but may increase the probability of having at least one crime.

Our paper contributes to the law and economics literature by examining situations where both the incentives to commit crimes and the incentives to report crimes are endogenous. This is motivated by civil and criminal lawsuits where facts are generated by individuals (e.g., potential criminals and witnesses) whose incentives interact with how information is aggregated, communicated, and ultimately incorporated into judicial decisions. This stands in contrast to several recent papers that focus on potential witnesses'

[5]We also show that strictly concave conviction probability cannot outperform linear conviction probability when the benefit from committing crime is low. This is because the informativeness of accusations under a concave conviction probability is close to that under a linear conviction probability, while the probability that the principal commits multiple crimes is significantly higher.

incentives to report crimes, such as Lee and Suen (2020), Cheng and Hsiaw (2020), and Naess (2020), as well as Siegel and Strulovici (2020) in which the mechanism designer elicits information only from defendants but treat the quality of witness testimonies as exogenous.

By endogenizing the distribution of potential witnesses' private signals through the strategic commission of offenses, our model stands in contrast to the existing models of information aggregation and transmission such as Banerjee (1992), Bikhchandani, et al. (1992), Smith and Sørensen (2000), Battaglini (2002), Ambrus and Takahashi (2008), and Ekmekci and Lauermann (2019) where the fact of interest is exogenous. Compared to existing works on endogenous information acquisition where either only one agent has private information (e.g., Pei 2015, Argenziano et al. 2016) or agents' private signals are conditionally independent (e.g., Persico 2004), the correlation of agents' private observations is also endogenous in our model.

Section 2 sets up the model. Section 3 characterizes the constrained optimal outcome when the judge can commit. Section 4 examines the equilibrium outcomes under APP and DPP when the judge cannot commit, and compares them with the optimal commitment outcome. Section 5 concludes and discusses the connections between our work and the existing literature.

2 Model

We study a three-stage game between one potential criminal (the principal), two potential witnesses or victims (the agents), and a mechanism designer or judge. In stage 1, the principal chooses $\boldsymbol{\theta} \equiv (\theta_1, \theta_2) \in \{0,1\}^2$, where $\theta_i = 1$ stands for the principal commits a crime witnessed by agent i and vice versa.

In stage 2, agent $i \in \{1,2\}$ decides whether to accuse the principal ($a_i = 1$) or not ($a_i = 0$) after privately observes $\theta_i \in \{0,1\}$, $\omega_i \in \mathbb{R}$, and $c_i \in [0, \overline{c}]$. We interpret ω_i as a payoff shock that affects agent i's preference toward convicting the principal, and c_i is his cost of filing an accusation which can result from social stigma or the principal's retaliation. Let Φ be the cdf of ω_1 and ω_2. Let F be the cdf of c_1 and c_2. We assume that ω_1, ω_2, c_1 and c_2 are independent, $\overline{c} \in (0, +\infty]$, both Φ and F have full support and admit continuous density functions ϕ and f. In order to establish the existence of nontrivial equilibrium, we assume throughout the paper that the density of f is large enough at 0:[6]

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

In stage 3, the mechanism designer or judge observes $\boldsymbol{a} \equiv (a_1, a_2) \in \{0,1\}^2$ and chooses $s \in \{0,1\}$,

[6]This assumption is easily achieved, starting from any arbitrary distribution F, 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 $1/\sup_\alpha \alpha\Phi(-\alpha)$.

where $s = 1$ stands for convicting the principal and $s = 0$ stands for acquitting him.

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

$$s \cdot \underbrace{(\theta_i - \gamma c_i a_i - \omega_i)}_{\text{agent } i\text{'s payoff when the principal is convicted}} + (1-s) \cdot \underbrace{(-c_i a_i)}_{\text{agent } i\text{'s payoff when the principal is acquitted}}, \tag{2.2}$$

where $\gamma \in (0, 1)$. Intuitively, agent i's benefit from convicting the principal is greater if he has witnessed a crime (i.e., $\theta_i = 1$) or when the realization of his payoff shock ω_i is lower. Agent i incurs a cost when he accuses the principal, which equals c_i if the principal is acquitted and equals γc_i if the principal is convicted.[7] Since $\gamma \in (0, 1)$, the agent's loss from retaliation is strictly greater when the principal is acquitted.

Section 3 analyzes a mechanism design problem where conviction decisions are made according to a mechanism that was committed to in advance. The designer's objective is to minimize the expected number of crimes subject to a constraint that the fraction of wrongful convictions does not exceed some cutoff.

Section 4 analyzes settings without commitment where conviction decisions are made according to the judge's preference and her posterior belief about the principal's guilt. We study two conviction rules commonly discussed in the legal scholarship, which translate into two classes of preferences for the judge (1) the judge prefers to convict the principal when the probability that he is *guilty of at least one crime* exceeds some threshold and vice versa, and (2) the judge prefers to convict the principal when the probability that he is *guilty of a particular criminal behavior* exceeds some threshold and vice versa.

3 Optimal Commitment Outcomes

Suppose the conviction decision follows a *mechanism* $q : \{0, 1\}^2 \to [0, 1]$ that was committed to in advance, which maps agents' reports to the probability of conviction.[8] The mechanism designer's objective is to minimize the expected number of crimes $\mathbb{E}[\theta_1 + \theta_2]$ subject to a constraint that the fraction of wrongful convictions, $\Pr(\boldsymbol{\theta} = (0, 0)|s = 1)$, does not exceed some threshold $\overline{\pi} \in (0, 1)$, which measures the society's tolerance toward wrongful convictions. We incorporate an upper bound on wrongful convictions instead of

[7] The coefficient in front of θ_i is deterministic and equals 1 in agent i's payoff when the principal is convicted. This is without loss of generality, since the agent's incentive depends only the ratio between that and (ω_i, c_i). Despite we assume ω_i and c_i are independent, our results extend when ω_i and c_i are correlated.

[8] Our approach takes the punishment from conviction as exogenous in which the designer only chooses the probability of conviction. This is the case when conviction entails undesirable side effects such as losing one's job or ending a promising career that have significantly larger cost to the principal compared to the sentencing. Our result also applies when the judge can endogenously design the size of the punishment given that the punishment from conviction is large enough relative to the benefit from committing crime and is independent of agents' messages.

mistaken acquittals since the fraction of mistaken acquittals $\Pr(\boldsymbol{\theta} \neq (0,0)|s=0)$ approximately equals the (unconditional) probability of crime, which can also be characterized by our analysis.

An equilibrium under mechanism q is $\big(\sigma_p, \sigma_1, \sigma_2\big)$, 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 $\sigma_i : \mathbb{R} \times [0, \bar{c}] \times \{0,1\} \to [0,1]$ maps ω_i, c_i, and θ_i to the probability of choosing $a_i = 1$. A mechanism and an equilibrium under that mechanism induce a joint distribution of $(\boldsymbol{\theta}, \boldsymbol{a}, s)$, which we call an *outcome*. Let $\boldsymbol{\pi}(\boldsymbol{a}) \in \Delta\big(\{0,1\}^2\big)$ be the posterior belief about the principal's guilt (θ_1, θ_2) after observing $\boldsymbol{a} \equiv (a_1, a_2)$. We introduce the notion of $\overline{\pi}$-valid outcome in order to formalize the constraint on wrongful convictions as well as our refinements.

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

1. $\Pr(\boldsymbol{\theta} = (0,0)|s=1) \leq \overline{\pi}$.
2. *If $\boldsymbol{\pi}(\boldsymbol{a}) = \boldsymbol{\pi}(\boldsymbol{a'})$, then $q(\boldsymbol{a}) = q(\boldsymbol{a'})$.*
3. *There exists $q : \{0,1\}^2 \to [0,1]$ with $q(1, a_{-i}) \geq q(0, a_{-i})$ for every $i \in \{1,2\}$ and $a_{-i} \in \{0,1\}$ under which there exists an equilibrium that attains this outcome.*

Intuitively, the mechanism designer can only use *monotone mechanisms* where each accusation *weakly increases* the probability of conviction and there exists at least one equilibrium that satisfies a Markovian refinement and the fraction of wrongful convictions is no more than $\overline{\pi}$.

Our *monotonicity* requirement implies that each accusation is a move against the principal. The motivation comes from our interpretation of c as the agent's loss from retaliation, since the principal would only retaliate against messages that increase the probability of conviction and will not retaliate against other messages.[9]

Our *Markovian refinement* requires that the probability of conviction under two message profiles are the same if they lead to the same posterior belief about the principal's guilt. This is motivated from a legal standpoint that conviction decisions should be independent of factors that are orthogonal to the defendant's guilt. It rules out situations where the principal never commits any crime that can be witnessed by agent i, agent i files an accusation with probability strictly between 0 and 1, yet the judicial decision is responsive to agent i's uninformative message. We discuss this refinement in Section 4.4, as well as its connections to the literature on multi-lateral contracting and other refinements such as proper equilibrium (Myerson 1978).

Focusing on the case in which the benefit from committing crime y is small relative to the punishment from conviction, Theorem 1 characterizes for every $\overline{\pi} \in (0,1)$, the lowest expected number of crimes among

[9]Suppose the principal can optimally commit to a retaliation plan (i.e., a mapping from the agent's message to his loss from retaliation) privately against each agent, as in Chassang and Padró i Miquel (2019), then he will retaliate to the maximum against the message that increases the probability of conviction and will not retaliate against the other message.

all $\overline{\pi}$-valid outcomes. We also construct mechanisms that approximately attain the optimum and study the qualitative features of the conviction probabilities under the optimal mechanisms.

Let

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

By definition, R is greater when Φ has a thinner left tail, that is, agents are more ethical and are less likely to make false accusations. One can verify that $R > 1$ when the pdf ϕ is log-concave which is the case for all Gaussian and exponential distributions, and $R = 1$ for fat-tailed distributions such as Pareto and Cauchy distributions. We demonstrate later on that R is a sufficient statistic for the informativeness of an agent's accusation when his private observation of crime is independent of the other agent's and the principal's benefit from committing crime is small relative to the punishment from conviction.

Let

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

Some algebra reveals that $\pi_{\min}(\overline{\pi}) < 1 - \overline{\pi}$ if and only if $R > \frac{\bar{l}}{2} + 1$.

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

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

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

Theorem 1 implies that when y is small, the expected number of crimes under the optimal $\overline{\pi}$-valid outcome is approximately $\min\{1 - \overline{\pi}, \pi_{\min}(\overline{\pi})\}$. Since $\min\{1 - \overline{\pi}, \pi_{\min}(\overline{\pi})\}$ is strictly decreasing in $\overline{\pi}$, there is a tradeoff between deterrence and reducing the fraction of wrongful convictions. Therefore, for every objective function that decreases with the expected number of crimes and the fraction of wrongful convictions, there exists $\overline{\pi}$ such that every mechanism that maximizes the above objective function is a constrained optimal mechanism where the upper bound on the fraction of wrongful convictions is $\overline{\pi}$.

We also construct mechanisms that can approximately attain the optimum. When the society is intolerant to wrongful convictions (i.e., $\overline{\pi}$ is small) and agents are unethical (i.e., R is small), the lowest expected number of crimes is $1 - \overline{\pi}$ and the optimal outcome can be implemented by committing to convict the principal only if he is accused by both agents. When the society is tolerant to wrongful convictions and

agents are ethical, the lowest expected number of crimes is $\pi_{\min}(\overline{\pi})$ and the optimal outcome can be implemented by a mechanism where the probability of conviction is linear in the number of accusations.

Intuitively, compared to mechanisms where the conviction probability is linear in the number of reports, mechanisms where the conviction probability is convex in the number of accusations eliminate the possibility that the principal commits multiple crimes but result in a high probability that he commits at least one crime. Which type of mechanism minimizes the expected number of crimes depends on the extent to which a linear conviction probability can reduce the probability of crime. This in turn depends on its effectiveness in improving the informativeness of agents' accusations (captured by R) and the extent to which improved informativeness can lower the probability of crime (depends on $\overline{\pi}$).

In particular, when R is larger, accusations are more informative under a linear conviction probability compared to a convex conviction probability. When $\overline{\pi}$ is higher, the mechanism designer is allowed to set a lower standard of proof, and improved informativeness has a larger impact on the equilibrium probability of crime. As a result, larger R and $\overline{\pi}$ are in favor of linear conviction probabilities and vice versa.

The proof is in Appendix D. We provide an intuitive explanation in three steps. First, whether the principal's decisions to commit different crimes are strategic complements or substitutes depends only on the sign of

$$Q \equiv q(1,1) + q(0,0) - q(1,0) - q(0,1). \tag{3.3}$$

When $Q > 0$, the conviction probability is convex in the number of accusations, and the principal's decisions to commit different crimes are strategic substitutes. Since the principal commits no crime with positive probability,[10] 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 under mechanisms where $Q > 0$. 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

$$\mathcal{I}_i \equiv \frac{\Pr(a_i = 1|\theta_i = 1)}{\Pr(a_i = 1|\theta_i = 0)}, \tag{3.4}$$

[10] If the principal commits at least one crime with probability 1, then the expected number of crime is no less than 1. Since one can construct a mechanism and an equilibrium where the expected number of crimes is $1 - \overline{\pi}$, such outcomes cannot be optimal.

which measures the informativeness of agent i's accusation since according to Bayes rule,

$$\underbrace{\frac{\Pr(\theta_i=1)}{1-\Pr(\theta_i=1)}}_{\text{likelihood ratio of crime under the judge's prior belief}} \cdot \mathcal{I}_i = \underbrace{\frac{\Pr(\theta_i=1|a_i=1)}{1-\Pr(\theta_i=1|a_i=1)}}_{\text{likelihood ratio of crime under the judge's posterior belief}},$$

i.e., $\mathcal{I}_i$ is a sufficient statistic for the responsiveness of the judge's posterior belief about θ_i after she observes 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. 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. As a result, the expected number of crimes is no less than $1-\overline{\pi}$, which is attained by a mechanism with a convex conviction probability.

Next, we consider $\overline{\pi}$-valid outcomes where θ_1 and θ_2 are uncorrelated, which can only be implemented via mechanisms where $Q=0$. A first observation is that any mechanism that satisfies $Q=0$ and implements some $\overline{\pi}$-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, which contradicts our Markovian refinement. Similarly, the Markovian refinement also implies that the principal commits crime against each agent with the same probability. In another word, every valid outcome must be symmetric across the two agents.

When θ_1 and θ_2 are uncorrelated, the principal commits multiple offenses with positive probability. This stands in contrast to mechanisms with $Q>0$, in which the principal commits at most offense. Unlike the case in which $Q>0$, agents' coordination motives no longer undermine the informativeness of their accusations. As a result, the probability that the principal commits at least one crime can be lower when θ_1 and θ_2 are uncorrelated than when $Q>0$. In order to derive the value of $\mathcal{I}_i$ when y is close to 0, note that 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 $\omega_i\leq\omega_i^{**}(c)$ and $\theta_i=0$. Since θ_1 and θ_2 are uncorrelated, both $\omega_i^*(c)$ and $\omega_i^{**}(c)$ are linear functions of c that have the same slope which we denote by $-K_i$. Since $\omega_i^*(0)=1$ and $\omega_i^{**}(0)=0$, we have $\omega_i^*(c_i)=1-c_iK_i$ and $\omega_i^{**}(c_i)=-c_iK_i$. By

definition,

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

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 the value of K_i diverges to $+\infty$. When $\int_{-\infty}^{0} \Phi(x)dx$ is finite, the dominated convergence theorem implies

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

and

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

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

Let $\widehat{\pi}$ be the probability that the principal commits crime against each individual agent. Bayes rule implies that 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 the informativeness ratio $\mathcal{I}_i \equiv \frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)}$ is close to R, some algebra reveals that the expected number of crimes $2\widehat{\pi}$ is close to $\pi_{\min}(\overline{\pi})$.

We also show that $\overline{\pi}$-valid outcomes where θ_1 and θ_2 are positively correlated cannot significantly improve upon those where θ_1 and θ_2 are uncorrelated. In particular, when y is close to 0, the informativeness ratio in (3.4) cannot improve upon the case with independent crimes. Moreover, having a positive correlation increases the probability that the principal commits multiple crimes. This implies that the constrained optimum can be either implemented via a convex conviction probability or a linear conviction probability.

4 Equilibrium Outcomes without Commitment

We now analyze equilibrium outcomes in the absence of commitment: convictions are decided by a Bayesian judge based on her preference as well as her posterior belief regarding the defendant's guilt. We consider two conviction rules discussed in the legal scholarship, which translate into two classes of preferences.

1. **Aggregate Probabilities Principle (APP):** The judge strictly prefers to convict the principal when the probability that he is guilty of at least one crime is strictly greater than some threshold $\pi^* \in (0,1)$, she strictly prefers to acquit the principal when that probability is strictly lower than π^*, and she is indifferent if that probability equals π^*. 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 best reply correspondence is:

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

   This is equivalent to the judge maximizing quadratic payoff function $-\left(s + (\pi^* - \frac{1}{2}) - \overline{\theta}\right)^2$.

2. **Distinct Probabilities Principle (DPP):** The judge strictly prefers to convict the principal when the probability that he is guilty of a particular criminal behavior is strictly greater than some threshold $\pi^{**} \in (0,1)$, she strictly prefers to acquit the principal if the probability that the principal is guilty of each criminal behavior is strictly lower than π^{**}. Otherwise, she is indifferent between acquitting and convicting the principal. The judge's best reply correspondence is:

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

The judicial decisions under APP and DPP coincide when $\pi^* = \pi^{**}$ and there is only one agent. When there are two agents, APP and DPP can lead to different decisions, which is illustrated by the example in the beginning of Section 1. In general, π^* and π^{**} may or may not be equal.

An equilibrium in the game without commitment is $\left(\sigma_p, \sigma_1, \sigma_2, q\right)$ 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. We require q to satisfy (4.1) in Section 4.1 and require q to satisfy (4.2) in Section 4.2.

We examine the common properties of *all* Bayes Nash equilibria that satisfy three refinements. The first refinement is the monotonicity requirement and the second one is the Markovian refinement, both of which have been introduced and discussed when setting up our mechanism design problem.

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

Refinement 2 (Markovian Refinement)**.** *If* $\boldsymbol{\pi}(\boldsymbol{a}) = \boldsymbol{\pi}(\boldsymbol{a'})$*, then* $q(\boldsymbol{a}) = q(\boldsymbol{a'})$.

We require in addition that the principal is acquitted for sure unless some agent accuses him.

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

Intuitively, if the principal is never accused, he is never arrested and hence is never convicted. Refinement 3 rules out equilibria where the principal commits crime against both agents with probability one and is convicted regardless of agents' reports. These equilibria are unappealing from a legal standpoint since the principal is convicted based on the judge's prior belief 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 the judge uses (4.1) to adjudicate guilt. 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. ***Convex Conviction Probabilities:*** $q(1,1) + q(0,0) - q(1,0) - q(0,1) > 0$.

According to Theorem 2, when 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 lowers the conviction threshold π^* to 10%, then the frequency of crime is below 10% but 90% of convicted people are innocent.

The proof is in Appendix B, and an intuitive explanation is provided in the remainder of this section. The first thing to note is that principal commits crime against each agent with positive probability. Suppose by way of contradiction that the probability with which the principal commits crime against agent i is zero, then the judge's posterior belief about (θ_1, θ_2) is independent of agent i's accusation a_i, so is the probability of conviction under Refinement 2. This implies that the principal's expected cost of committing crime against agent i is 0, while his benefit from doing so is strictly positive. Therefore, he has a strict incentive to choose $\theta_i = 1$. This contradicts the presumption that the principal chooses $\theta_i = 0$ with probability 1.

Next, if the principal's benefit from committing crime is low, then he is convicted with positive probability only when both agents accuse him. Intuitively, suppose that a single accusation suffices to convict the principal, then 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 strictly prefers to convict the principal when he faces two accusations, i.e., $q(1,1) = 1$. When the benefit from committing crime y is small relative to the loss from conviction, the principal strictly prefers not to commit any crime when he is convicted for sure under two accusations. This contradicts our conclusion that the equilibrium probability of crime is strictly positive, which rules out the possibility that the principal is convicted with positive probability when only one agent accuses him.

The above discussions suggest that $Q > 0$ in every equilibrium under APP. Similar to our analysis of mechanisms with a convex conviction probability, the principal's decisions to commit different crimes are strategic substitutes. This implies that the principal never commits multiple crimes and as a result, his equilibrium strategy induces a *negative correlation* in agents' private observations of crimes. In addition, the agents' decisions to accuse the principal are strategic complements since each agent's loss from retaliation is lower when the principal is convicted. The latter occurs with higher probability when the other agent accuses the principal. The endogenous negative correlation and agents' coordination motives lower the informativeness of agents' accusations, which in equilibrium, lead to a high frequency of crime.

4.2 Equilibrium Outcomes under DPP

Theorem 3 characterizes the equilibrium outcomes when the judge uses DPP to adjudicate guilt, in which different crimes are uncorrelated and the probability of conviction is linear in the number of accusations.

Theorem 3. *Suppose the judge uses (4.2) to adjudicate guilt. 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, the expected number of crimes in every equilibrium is approximately

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

In environments where R is large, the comparison between Theorems 2 and 3 unveils the comparative advantages of APP and DPP in terms of deterring crimes once we fix the fraction of wrongful convictions. In particular, DPP introduces the possibility that potential criminals committing multiple crimes, but it reduces the probability that potential criminals commits at least one crime when R is large enough. In another word, the fraction of agents who are prone to make false accusations is small.

Whether APP or DPP minimizes the *expected number of crimes* depends on the value of R, which determines the extent to which DPP improves the informativeness of accusations, as well as the standards of proof π^* and π^{**}. More details on the comparison will be discussed in Section 4.3.

The proof is in Appendix C. The key distinction between DPP and APP is driven by the observation that more accusations strictly increases the probability that the principal is guilty of at least one crime, but does not necessarily increase the probability that the principal is guilty of a particular crime. For example, when θ_1 and θ_2 are negatively correlated (or uncorrelated), agent 2's accusation undermines (or does not affect) the credibility of agent 1's, so the value of $\max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a})$ may not increase. This makes linear conviction probabilities consistent with the judge's preference under DPP but not under APP.

Next, we explain why agents' private observations of crime are uncorrelated. First, when y is small enough, $q(1,0)$ and $q(0,1)$ must also be small. The intuition is similar to why $Q > 0$ under APP, since when $q(1,0)$ or $q(0,1)$ is bounded away from 0, the principal will have a strict incentive not to commit any crime when y is small enough, and hence, will never be convicted. This cannot happen in equilibrium.

Suppose by way of contradiction 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 $\Pr(\theta_2 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_2 = 1|\boldsymbol{a} = (0,1))$. Under conviction rule (4.2), the above inequalities imply that $q(1,0) \geq q(1,1)$ and $q(0,1) \geq q(1,1)$. Since q is weakly increasing in each entry of $\boldsymbol{a}$, we have

$$q(1,1) = q(1,0) = q(0,1) = 1 \text{ and } q(0,0) = 0. \tag{4.5}$$

When y is small, the principal has a strict incentive not to commit any crime according to the conviction probability in (4.5). This contradicts the conclusion that the equilibrium probability of crime is interior.

Next, suppose by way of contradiction that θ_1 and θ_2 are *positively correlated*, in which case $\boldsymbol{a} = (1,1)$

is the unique maximizer of both $\Pr\left(\theta_1 = 1 \middle| \boldsymbol{a}\right)$ and $\Pr\left(\theta_2 = 1 \middle| \boldsymbol{a}\right)$. Under conviction rule (4.2),

$$q(1,1) \geq \max\{q(1,0), q(0,1)\} \geq q(0,0) = 0, \tag{4.6}$$

where the first inequality is strict unless $\max\{q(1,0), q(0,1)\} = 1$. When y is small, the principal has a strict incentive not to commit any crime if $q(1,0) = 1$ or $q(0,1) = 1$. This together with the presumption that θ_1 and θ_2 are positively correlated implies that

$$\pi^* = \Pr(\theta_i = 1 | \boldsymbol{a} = (1,1)) > \max\{\Pr(\theta_i = 1 | \boldsymbol{a} = (1,0)), \Pr(\theta_i = 1 | \boldsymbol{a} = (0,1))\}. \tag{4.7}$$

This suggests that $q(0,1) = q(1,0) = 0$. Since $q(1,1) > 0$, the value of Q is strictly positive so 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 θ_1 and θ_2 are uncorrelated, the principal's decisions to commit distinct offenses are neither substitutes or complements. This implies that $Q = 0$. Moreover, the marginal effect of a_i on the probability of conviction must be the same for all values of a_{-i}. 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, which means that $q(1,0) = q(0,1)$. Refinement 3 requires that $q(0,0) = 0$, and given that $Q = 0$, we have $q(1,1) = 2q(1,0) = 2q(0,1)$. The derivation of the informativeness ratio R follows from the one under mechanisms with linear conviction probabilities, which we omit to avoid repetition.

4.3 Attainability of Optimal Commitment Outcomes

Taken together, Theorems 1, 2, and 3 imply the optimal mechanism with commitment can be approximated by any equilibrium that satisfies Refinements 1, 2 and 3 in a game without commitment, in which the judge makes conviction decisions based on her posterior belief about (θ_1, θ_2) according to either APP or DPP. More precisely:

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 of the game without commitment when the judge uses APP and sets $\pi^* = 1 - \overline{\pi}$.

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 of the game without commitment when the judge uses

DPP and sets π^{**} according to

$$\overline{\pi} = \frac{(1-\pi^{**})^2 R}{(1-\pi^{**})R + \pi^{**}}. \tag{4.8}$$

Intuitively, (4.3) shows how to compute the fraction of wrongful convictions using the standard of proof under DPP. Therefore, when π^{**} satisfies (4.8) and the judge uses DPP to adjudicate guilt, the fraction of wrongful convictions in equilibrium is approximately $\overline{\pi}$.

This implication of our result provides a justification for the use of APP when the standard of proof is high (i.e., $\overline{\pi}$ is low), or in another word, the society puts a relatively high weight on reducing the fraction of wrongful convictions, or when the potential witnesses are relatively unethical and are prone to make false accusations. By contrast, DPP is optimal when the standard of proof is low and when agents are relatively ethical and the fraction of agents who are prone to make false accusations is low.

4.4 Discussion

This section establishes the existence of an equilibrium satisfying the three refinements introduced earlier, discusses our Markovian refinement in more detail, and sheds light on the forces driving our results by comparing them the equilibrium outcomes with two agents to the single-agent benchmark.

Equilibrium Existence: Theorems 2 and 3 establish the common properties of all equilibria that satisfy Refinements 1, 2, and 3 but do not establish 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 exists an equilibrium that satisfies Refinements 1, 2, and 3 under APP, and there exists an equilibrium that satisfies Refinements 1,2, and 3 under DPP.*

The proof is in Appendix A, in which we use the Brouwer's fixed point theorem to construct equilibria that satisfy our refinements. The existence of equilibrium requires y to be small. Intuitively, when y is large, for example $y \geq 1$, the principal's benefit from committing crime exceeds his loss from conviction. As a result, he commits crime with probability 1 in all equilibria, which violates Refinement 3.

Markovian Refinement: Our Markovian refinement imposes restrictions on each agent's belief after he observes the principal taking an off-path action. Similar issues arise in the multi-lateral contracting models of Segal (1999) and Gavazza and Lizzeri (2009). They introduce a *passive belief refinement*, which requires that when the principal makes an off-path offer to agent i, agent i's belief about the principal's offers to other agents remains unchanged.

Equilibria that satisfy their passive belief refinement may not satisfy our Markovian refinement. To illustrate, consider the principal's strategy of committing crime against agent 1 with probability $\pi \in (0,1)$ and committing crime against agent 2 with probability zero. Such a strategy can be part of an equilibrium under the passive belief refinement. This is because agent 2 believes that $\theta_1 = 1$ with probability π regardless of his observation of θ_2. As a result, agent 2 accuses the principal with higher probability when he has witnessed a crime, which deters the principal from choosing $\theta_2 = 1$.

However, such a strategy can never occur in any equilibrium that satisfies our Markovian refinement. This is because the judge's belief about (θ_1, θ_2) is independent of agent 2's message, so the conviction probabilities should be independent of a_2. As a result, the cost of committing crime against agent 2 is zero, which contradicts the presumption that the principal never commits crime against agent 2.

Under mechanisms where $Q > 0$, all equilibria that satisfy our Markovian refinement are proper equilibria (Myerson 1978) while equilibria where the principal never commits crime against one of the agents (e.g., equilibria that satisfy the passive belief refinement) are not proper equilibria.

This is because in such equilibria, the principal chooses $(\theta_1, \theta_2) = (0,0)$ and $(1,0)$ with positive probability, and other actions with zero probability. Since $Q > 0$, his actions to commit different crimes are strategic substitutes. Therefore, his 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 with which the principal playing $(\theta_1, \theta_2) = (1,1)$ being arbitrarily small compared to the probability of playing $(\theta_1, \theta_2) = (0,1)$, so agent 2 believes 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, so whenever the principal has a weak incentive to commit crime against agent 1, he has a strict incentive to commit crime against agent 2.

Single-Agent Benchmark: A key 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 is that due to the cost of reporting, agents free-ride on each other's accusations when there are multiple agents.

In order to rule out such a public good provision story, we examine a benchmark scenario where there is only one agent, under which APP and DPP lead to the same equilibrium outcome as long as $\pi^* = \pi^{**}$. We compare the equilibrium outcomes in the single-agent benchmark with those in the two-agent scenario. We assume that there exists $\overline{\omega} \in \mathbb{R}$ such that when $\omega < \overline{\omega}$, $\phi(\omega)$ is a strictly increasing function. This is

satisfied for all normal distributions and exponential distributions.

Proposition 2. *If there is only one agent and ϕ is strictly increasing when ω is small enough, then 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,*

1. *The informativeness of the agent's report, measured by $\frac{\Pr(a=1|\theta=1)}{\Pr(a=1|\theta=0)}$, is at least R.*
2. *The equilibrium probability of crime is at most $\frac{\pi^*}{(1-\pi^*)R+\pi^*}$.*
3. *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.*

The proof is in Appendix F. The last statement of Proposition 2 implies that when there are two agents and the benefit from committing crime is low relative to the punishment from conviction, the probability that each agent accuses the principal is *strictly higher* compared to the single-agent benchmark.[11] This distinguishes our Theorem 2 with the results on inefficient public good provision, where each agent contributes less when there are more agents due to free-riding incentives. In our model, each agent accuses the principal with strictly higher probability when there are more agents. As a result, the high equilibrium probability of crime is not caused by agents' incentives to free-ride on others' accusations.

5 Concluding Remarks

We discuss the interpretation and robustness of our results as well as the connections between our work and those in the existing literature on voting, coordination games, and law and economics.

Equilibrium Analysis *vs.* Nonequilibrium Adjustments: Our results are derived from an equilibrium analysis, which presumes that players have correct expectations about the consequences of their actions and other players' strategies. When social rules change, as in 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.

[11] In Pei and Strulovici (2020), we generalize this comparative statics result to any finite number of agents.

Shielding Accusers from Stigma through Secret Accusations: In order to address the potential pressure that is sometimes experienced by lone accusers, institutions have been developed under which reports are submitted to a third party and are only released when enough of them have been filed.[12]

Such provisions increase the risk of wrongful accusations. Indeed, an agent holding a grudge against the principal has an opportunity to secretly accuse him in the hope that other agents, rightfully or not, may also accuse the principal. While these institutions are clearly well intentioned and worth considering, it is important to evaluate their consequences once all agents understand them.

Moreover, if accusations run the risk of being leaked, even with a small probability, this creates a strictly positive expected cost of retaliation, and brings us back to the baseline model.

Simultaneous vs Sequential Reporting: The forces that underlie our results are also present in dynamic versions of our model, in which reports may be filed sequentially. First, the negative correlation between the agents' private information (θ_i) continues to arise endogenously whenever a strategic principal is concerned about having too many reports made against him. Second, an individual agent has an incentive to coordinate with other agents whenever he is unsure about whether his report is pivotal or not. In a dynamic setting, this incentive can materialize after a *cold start* (i.e., where very few people have reported before and no agent wants to be the first accuser). It can also occur when an agent has observed many reports and is unsure of the number of reports needed to convict the principal (for example, if he faces uncertainty about the conviction standard π^* used by the judge). The inefficiencies and lack of credibility caused by the agents' coordination motive thus still arise in a dynamic environment.[13]

Related Literature: Our game without commitment can be viewed as a voting model where the voting rule and the correlation between voters' private signals are endogenous. This stands in contrast to existing works where at least one of these two ingredients are exogenous. This includes the classic voting model of Feddersen and Pesendorfer (1998), voting models with endogenous information acquisition (Persico 2004), voting with negatively correlated private information or payoffs (for example, Schmitz and Tröger 2012 and Ali, Mihm, and Siga 2018), and dynamic voting models where information acquisition may induce negative correlation in voters' continuation values (Strulovici 2010).

Our result that the agents' reports are arbitrarily uninformative when the principal's decisions are strategic substitutes suggests a new channel through which information aggregation can fail. In Banerjee (1992),

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

[13] See Lee and Suen (2020) for a model of strategic accusation in which the timing of accusation plays a major role.

Bikhchandani et al. (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 because of the negative correlation in agents' private information, combined with agents' incentives to coordinate their reports. 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 creates negative correlation in the *informativeness* of agents' signals, rather than the *direction* of these signals, which hampers social learning.

The coordination motive arising endogenously in our model is reminiscent of the literature on global games. In Carlsson and Van Damme (1993) and Morris and Shin (1998), agents receive conditionally independent private signals about a common state of the world. In Baliga and Sjöström (2004), each agent privately observes his value for a decision, and the values are independent across agents. In contrast, the agents' private information is endogenous in our model, which depends on the principal's actions.

Our paper contributes to the law and economics literature by studying decision rules that aggregate the probabilities of offenses, endogenizing the informativeness of witness testimonies, and analyzing the interplay between an individual's incentive to commit offenses and witnesses' incentives to report the truth. These features of our model stand in contrast to the recent works of Silva (2019) and Baliga, Bueno de Mesquita and Wolitzky (2020), in which there are multiple defendants and the negative correlation in their guilt is exogenously assumed.

Siegel and Strulovici (2020) apply mechanism design to a judicial setting in which the mechanism designer can elicit information from the defendant the evidence available (e.g., witness testimonies) is independent of the sentencing scheme. In the present paper, the mechanism designer solicits information from witnesses or victims, and the quality of the evidence is endogenous and depends on the mechanism.

Lee and Suen (2020) study the timing of reports by victims and libelers when a criminal commits offenses against two agents with exogenous probability. They provide an explanation for the well-documented fact that victims sometimes delay their accusations. Their analysis and ours consider complementary aspects of witnesses' reporting incentives. 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 reporting incentives and, among other results, finds that making reporting costly may improve social welfare. The principal's strategic restraint that emerges endogenously in our model and the negative correlation that it induces on the agents' private information are distinctive features of our analysis.

A Existence of Equilibrium

This section establishes the existence of equilibrium that satisfies Refinements 1, 2 and 3 under APP and DPP. It also implies the existence of equilibrium that satisfies the Markovian refinement under the two classes of mechanisms described in the statement of Theorem 1.

A.1 Existence of Equilibrium under APP or Mechanisms with Convex Conviction Probabilities

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

Lemma A.1. *When F and Φ satisfy (2.1), there exists $\overline{y} > 0$ such that if $(\Phi^*, \Phi^{**}) \in (0,1)^2$ solves*

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

and

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

where $\Phi^* \equiv \int \Phi(\omega^*(c))dF(c)$ and $\Phi^{**} \equiv \int \Phi(\omega^{**}(c))dF(c)$, then $\Phi^{**}(\Phi^* - \Phi^{**}) \geq \overline{y}$.

Proof. First, we bound Φ^{**} from below. Every solution to (A.1) and (A.2) satisfies:

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

Let $g(\Phi^{**}) \equiv \Phi^{**}$ and $h(\Phi^{**}) \equiv \int \Phi\Big(c(1-\gamma) - \frac{c}{\Phi^{**}}\Big)dF(c)$. 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, $\Phi^* > \Phi^{**}$. Equations (A.1) and (A.2) imply that

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

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 $c \in [0, c^*]$. Therefore,

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

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 $\Phi^* - \Phi^{**} \geq \frac{1}{2}F(1)\Psi$. The two parts together leads to a uniform lower bound on $\Phi^{**}(\Phi^* - \Phi^{**})$. □

Recall the construction of $\overline{y}$ and ε in the proof of Lemma A.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:

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

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

$$f_1(\Phi^*, \Phi^{**}, q) = \min\Big\{1, \frac{y}{\Phi^{**}(\Phi^* - \Phi^{**})}\Big\}. \tag{A.8}$$

Since f is continuous, the Brouwer's fixed point theorem implies the existence of a fixed point. The construction of ε implies that when $y < \varepsilon$, $\Phi^* > \varepsilon$ and $\Phi^{**} > \varepsilon$ at every fixed point. Otherwise, $\Phi^{**} = \varepsilon$ at some fixed point of f. Equation (A.8) and $y < \varepsilon$ imply that $q = 1$. According to (A.7),

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

which leads to 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 or Mechanisms with Linear Conviction Probabilities

We establish the existence of 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, different crimes are independent, and the principal is indifferent between his four actions. First, the 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 presumption that $q^* = 1/2$ is a fixed point.

B Proof of Theorem 2

We provide an overview of our proof by decomposing it 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 (3.3). 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,

$$\mathcal{I} \cdot \frac{\Pr(\overline{\theta} = 1)}{1 - \Pr(\overline{\theta} = 1)} = \frac{\Pr(\overline{\theta} = 1 | \boldsymbol{a} = (1,1))}{1 - \Pr(\overline{\theta} = 1 | \boldsymbol{a} = (1,1))}. \tag{B.2}$$

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 show 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, with a unified treatment being provided in Appendix E.

B.1 Proof of Lemma B.1

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

$$\underbrace{\mathbb{E}[q(0,a_{-i}) - q(1,a_{-i})]}_{\leq 0}(\omega_i - \theta_i) < c_i \underbrace{\mathbb{E}[1 - (1-\gamma)q(1,a_{-i})]}_{>0}. \tag{B.3}$$

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.3) 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.3) 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

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

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

$$(\Phi_1^* - \Phi_1^{**}) \Big((1 - \Phi_2^{*}) \big(q(1, 0) - q(0, 0) \big) + \Phi_2^{*} \big(q(1, 1) - q(0, 1) \big) \Big). \tag{B.5}$$

Committing different offenses are strategic substitutes if and only if (B.4) is less than (B.5) or, equivalently, if

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

The above 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

$$\omega_i^*(c) = 1 + c(1-\gamma) - \frac{c}{q^*\Phi_{-i}^{**}} \text{ and } \omega_i^{**}(c) = c(1-\gamma) - \frac{c}{q^*(\beta_i\Phi_{-i}^{**} + (1-\beta_i)\Phi_{-i}^*)}, \tag{B.6}$$

where $\beta_i \equiv \Pr(\theta_{-i} = 0|\theta_i = 0)$. Equation (B.6) 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 implies that $\Phi_1^*/\Phi_1^{**} = \Phi_2^*/\Phi_2^{**}$. If $\omega_1^*(c) > \omega_2^*(c)$ for some $c > 0$, then the first part of (B.6) 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 presumption that $\Phi_1^*/\Phi_1^{**} = \Phi_2^*/\Phi_2^{**}$. If $\omega_1^*(c) = \omega_2^*(c)$ for some $c > 0$, then the first part of (B.6) 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.6).

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 $\omega_1^{**}(c) = \omega_2^{**}(c)$ for every $c \geq 0$. 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 $c > 0$, $\omega^*(c) \to -\infty$ when $y \to 0$. The principal being indifferent between $\boldsymbol{\theta} = (0,0)$ and $\boldsymbol{\theta} = (1,0)$ implies that

$$y = q^*\Phi^{**}(\Phi^* - \Phi^{**}). \tag{B.7}$$

Suppose 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 $\{k_n\}_{n\in\mathbb{N}}$ 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.7) 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.6) implies that $\omega^{**}_{k_n}(c)$ converges to $-\infty$, which contradicts the presumption 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.6) 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 $\mathcal{I}$ in (B.2). Lemma B.2 implies that $\Pr(\overline{\theta} = 1|\boldsymbol{a} = (1, 1)) = \pi^*$, which implies that

$$\beta = \frac{2\mathcal{I}}{l^* + 2\mathcal{I}} \text{ and } 1 - \beta = \frac{l^*}{l^* + 2\mathcal{I}}. \tag{B.8}$$

Moreover, $\mathcal{I} = \frac{\Phi^* \cdot \Phi^{**}}{\Phi^{**} \cdot \Phi^{**}} = \frac{\Phi^*}{\Phi^{**}}$. Equation (B.3) implies that for every $c > 0$,

$$\left|\frac{\omega^*(c) - 1 - c(1-\gamma)}{\omega^{**}(c) - c(1-\gamma)}\right| = \frac{\beta\Phi^{**} + (1-\beta)\Phi^*}{\Phi^{**}} = \frac{(l^* + 2)\mathcal{I}}{l^* + 2\mathcal{I}}. \tag{B.9}$$

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.9) converges to 1. Since the RHS of (B.9) 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, with proof in the 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

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

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

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

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, it is without loss of generality to focus on conviction rules satisfying $q(0,0) = 0$. This is because the monotonicity requirement implies that $q(1,1) \geq \max\{q(1,0), q(0,1)\} \geq q(0,0)$, and if $q(0,0) \neq 0$, consider a new conviction rule constructed according to $q^*(\boldsymbol{a}) \equiv q(\boldsymbol{a}) - q(0,0)$. The principal's and the agent's incentives remain the same. Since $\Pr(\boldsymbol{\theta} = (0,0)|\boldsymbol{a} = (0,0)) \leq \Pr(\boldsymbol{\theta} = (0,0)|\boldsymbol{a})$ for every $\boldsymbol{a}$, the fraction of wrongful conviction $\Pr(\boldsymbol{\theta} = (0,0)|s = 1)$ is weakly lower under q^* compared to 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,

$$\Phi_j^* \equiv \int_0^{\overline{c}} \Phi(\omega_j^*(c)) dF(c) \quad \text{and} \quad \Phi_j^{**} \equiv \int_0^{\overline{c}} \Phi(\omega_j^{**}(c)) dF(c). \tag{D.3}$$

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, $K_i^* > K_i^{**}$ 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 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 $q(1,1)+q(0,0)-q(1,0)-q(0,1) > 0$, and are strategic complements if $q(1,1)+q(0,0)-q(1,0)-q(0,1) < 0$. 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 $(1,0)$ or $(0,1)$ or both with zero probability.

In what follows, we partition the set of 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 $c \in [0, \overline{c}]$,

$$\Big|\frac{\omega_i^*(c) - 1 - c(1-\gamma)}{\omega_i^{**}(c) - c(1-\gamma)}\Big| = \frac{X_i^{**}(\pi_{-i}Y_i^{*} + (1-\pi_1-\pi_2)Y_i^{**})}{Y_i^{**}(\pi_{-i}X_i^{*} + (1-\pi_1-\pi_2)X_i^{**})} = 1 + \frac{\pi_{-i}}{\pi_{-i}X_i^{*} + (1-\pi_1-\pi_2)X_i^{**}} \cdot \frac{X_i^{**}Y_i^{*} - X_i^{*}Y_i^{**}}{Y_i^{**}}, \tag{D.4}$$

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.4) 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 being indifferent between $\boldsymbol{\theta} = (0,1)$ and $\boldsymbol{\theta} = (1,0)$ translates into

$$(\Phi_1^{*} - \Phi_1^{**})\Big\{(q - q_1 - q_2)\Phi_2^{**} + q_1\Big\} = (\Phi_2^{*} - \Phi_2^{**})\Big\{(q - q_1 - q_2)\Phi_1^{**} + q_2\Big\}, \tag{D.5}$$

which implies that

$$\frac{X_1^{**}Y_1^{*} - X_1^{*}Y_1^{**}}{Y_1^{**}} = \frac{X_2^{**}Y_2^{*} - X_2^{*}Y_2^{**}}{Y_2^{**}}. \tag{D.6}$$

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

$$\max\Big\{\mathcal{I}_1, \mathcal{I}_2\Big\} \leq R, \quad \text{and} \quad \frac{\pi_1+\pi_2}{1-\pi_1-\pi_2}\Big(\frac{\pi_1}{\pi_1+\pi_2}\frac{\Phi_1^*}{\Phi_1^{**}} + \frac{\pi_2}{\pi_1+\pi_2}\frac{\Phi_2^*}{\Phi_2^{**}}\Big) = \frac{\overline{\pi}}{1-\overline{\pi}}, \tag{D.7}$$

we know that $\pi_1+\pi_2$ 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 either $\frac{(\Phi_1^*-\Phi_1^{**})(q-q_1-q_2+(1-\gamma)q_1q_2)}{(q-q_1-q_2)\Phi_1^{**}+q_2}$ 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

$$\frac{q-q_1-q_2+(1-\gamma)q_1q_2}{(q-q_1-q_2)+\frac{q_2}{\Phi_1^{**}}} \tag{D.8}$$

converges to 0, which happens if and only if $\frac{q_2}{\Phi_1^{**}}$ converges to infinity. Since $\omega_1^*(c) = b - cK_1^*$ 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.7) 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 $\max\{\frac{\Phi_1^*}{\Phi_1^{**}}, \frac{\Phi_2^*}{\Phi_2^{**}}\} \leq R+\varepsilon$. 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.

$$\frac{\Phi_j^*}{\Phi_j^{**}} = \frac{\int_0^{\overline{c}}\Phi(\omega_j^*(c))dF(c)}{\int_0^{\overline{c}}\Phi(\omega_j^{**}(c))dF(c)} = \frac{K_j^{**}}{K_j^*}\cdot\frac{\displaystyle\int_{-\infty}^{1} f(\frac{1-x}{K_j^*})\Phi(x)dx}{\displaystyle\int_{-\infty}^{0} f(\frac{-x}{K_j^{**}})\Phi(x)dx} \tag{D.9}$$

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

$$\lim_{k\to+\infty}\int_{-\infty}^{1} f(\frac{1-x}{k})\Phi(x)dx = \int_{-\infty}^{1}\lim_{k\to+\infty} f(\frac{1-x}{k})\Phi(x)dx = \lim_{c\downarrow 0} f(c)\int_{-\infty}^{1}\Phi(x)dx, \tag{D.10}$$

and

$$\lim_{k\to+\infty}\int_{-\infty}^{0} f(\frac{-x}{k})\Phi(x)dx = \int_{-\infty}^{0}\lim_{k\to+\infty} f(\frac{-x}{k})\Phi(x)dx = \lim_{c\downarrow 0} f(c)\int_{-\infty}^{0}\Phi(x)dx. \tag{D.11}$$

Therefore,

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

and we only need to show that when $y\to 0$, $K_i^{**}/K_i^* \to 1$.

Lemma D.4. *For every $\varepsilon > 0$, there exists $\overline{y}_\varepsilon > 0$ such that when $y < \overline{y}_\varepsilon$, if (q, q_1, q_2) is a $\overline{\pi}$-valid conviction rule with $q \leq q_1 + q_2$, then $\max\{q_1/q_2, q_2/q_1\} > \varepsilon$.*

Proof. Under every $\overline{\pi}$-valid conviction rule $(q, q_1, q_2, 0)$ and every equilibrium that satisfies our three requirements, $\boldsymbol{\theta} = (0, 0)$ is weakly optimal for the principal, which implies that

$$y \leq (q-q_2)(\Phi_1^* - \Phi_1^{**})\Phi_2^{**} + q_1(\Phi_1^* - \Phi_1^{**})(1-\Phi_2^{**}), \tag{D.12}$$

$$y \leq (q-q_1)(\Phi_2^* - \Phi_2^{**})\Phi_1^{**} + q_2(\Phi_2^* - \Phi_2^{**})(1-\Phi_1^{**}). \tag{D.13}$$

Since the principal commits crime against each agent with positive probability,

$$y \geq \min\Big\{(q-q_2)(\Phi_1^*-\Phi_1^{**})\Phi_2^{**}+q_1(\Phi_1^*-\Phi_1^{**})(1-\Phi_2^{**}), (q-q_2)(\Phi_1^*-\Phi_1^{**})\Phi_2^{*}+q_1(\Phi_1^*-\Phi_1^{**})(1-\Phi_2^{*})\Big\}, \tag{D.14}$$

$$y \geq \min\Big\{(q-q_1)(\Phi_2^*-\Phi_2^{**})\Phi_1^{**}+q_2(\Phi_2^*-\Phi_2^{**})(1-\Phi_1^{**}), (q-q_1)(\Phi_2^*-\Phi_2^{**})\Phi_1^{*}+q_2(\Phi_2^*-\Phi_2^{**})(1-\Phi_1^{*})\Big\}. \tag{D.15}$$

Suppose toward a contradiction that $q_2/q_1 \to 0$ as $y \to 0$. Since $q_1 + q_2 \geq q$, for every $i \in \{1, 2\}$, $K_i^* \leq K_i^{**}$ and $\omega_i^*(c) - \omega_i^{**}(c) \geq 1$. Moreover,

$$(q-q_1)(\Phi_2^* - \Phi_2^{**})\Phi_1^{**} + q_2(\Phi_2^* - \Phi_2^{**})(1-\Phi_1^{**}) \geq (q-q_2)(\Phi_1^* - \Phi_1^{**})\Phi_2^{*} + q_1(\Phi_1^* - \Phi_1^{**})(1-\Phi_2^{*}),$$

which implies that

$$\frac{q_2}{q_1} \geq \frac{\Phi_1^* - \Phi_1^{**}}{(\Phi_1^* - \Phi_1^{**})\Phi_2^* + (\Phi_2^* - \Phi_2^{**})}.$$

Since $q_2/q_1 \to 0$, it must be the case that

$$\frac{\Phi_1^* - \Phi_1^{**}}{\Phi_2^* - \Phi_2^{**}} \to 0.$$

Recall the expressions of K_i^* and K_i^{**}, which imply that as $y \to 0$, K_1^* converges to $\frac{1-q_1}{q_1}$ and K_2^* converges to $\frac{1-q_2}{q_2}$, and since $q_2/q_1 \to 0$, we have $K_1^* \ll K_2^*$. Moreover, given that $q \leq q_1 + q_2$, the principal's actions are strategic complements, so $K_i^* \leq K_i^{**}$ for every $i \in \{1, 2\}$. Therefore,

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

Since

$$\lim_{k \to +\infty} \int_0^{\overline{c}} \Phi(1 - ck) dF(c) = 0,$$

there exists $\overline{K}_2 > 0$ such that for every $K_2^* \geq \overline{K}_2$,

$$\Phi_1^* - \Phi^{**} \geq \int_0^{\overline{c}} \Big(\Phi(1 - cK_1^*) - \Phi(-cK_1^*)\Big) dF(c) \geq \int_0^{\overline{c}} \Phi(1 - cK_2^*) dF(c) = \Phi_2^* > \Phi_2^* - \Phi_2^{**}.$$

The above inequality contradicts the presumption that $\frac{\Phi_1^* - \Phi_1^{**}}{\Phi_2^* - \Phi_2^{**}} \to 0$. □

Lemma D.4 together with the expressions for K_i^* and K_i^{**} implies that $\lim_{y \to 0} \frac{K_i^{**}}{K_i^*} = 1$. This together with (D.9), (D.10) and (D.11) implies that $\frac{\Phi_j^*}{\Phi_j^{**}} \leq R + \varepsilon$.

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 (4.8). In the online appendix, we solve (4.8) and obtain:

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

Plugging this into the expected number of crimes, we obtain:

$$\frac{2\pi^{**}}{(1 - \pi^{**})R + \pi^{**}} = \pi_{\min}(\overline{\pi}).$$

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 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 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

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

and

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

Suppose $q(1,0)$ 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 $\Phi_1^{**} = \int_0^{\overline{c}} \Phi(-cK_1^{**}) dF(c)$, we know that 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 $q(0,1)\to 0$. Recall the expressions for K_1^* and K_1^{**} in Appendix A, we know that $K_1^*-K_1^{**}\to 0$. However, 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, one can show that $q(0,1)$ 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 completes the proofs of Lemmas C.1 and D.2

We complete the proof of Lemma B.2 by ruling out situations where 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,

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

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,

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

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.6) 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 conditions imply that

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

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 $c\geq c^*$. Pick c^* such 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 implies 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] Chassang, Sylvain and Gerard Padró i Miquel (2019) "Corruption, Intimidation and Whistle-Blowing: A Theory of Inference from Unverifiable Reports," *Review of Economic Studies*, forthcoming.

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

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

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

[17] 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.

[18] Gavazza, Alessandro and Alessandro Lizzeri (2009) "Transparency and Economic Policy," *Review of Economic Studies*, 76, 1023-1048.

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

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

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

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

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

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

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

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

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

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

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

[30] Segal, Ilya (1999) "Contracting with Externalities," *Quarterly Journal of Economics*, 114(2), 337-388.

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

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

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

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

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

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

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