Crime Aggregation, Deterrence, and Witness Credibility*

– Preliminary –

Harry Pei Bruno Strulovici

May 28, 2020

Abstract: We present a model for the equilibrium frequency of offenses and the informativeness of witness reports when potential offenders can commit multiple offenses and witnesses are subject to retaliation risk and idiosyncratic taste shocks. We compare several ways of handling multiple accusations discussed in legal scholarship. (i) When accusations are aggregated to determine the probability that the defendant committed at least one unspecified offense and conviction entails severe punishment, witness reports are arbitrarily uninformative and offenses are frequent in equilibrium. Offenders induce negative correlation in witnesses' private information, which causes information aggregation to fail. (ii) When accusations are treated separately to adjudicate guilt and conviction entails severe punishment, witness reports are highly informative and offenses infrequent in equilibrium.

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

1 Introduction

When a defendant faces multiple charges, the legal norm is to consider these charges separately and to convict the defendant if there are *specific* charges for which the 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. For example, 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, the defendant is acquitted on both count, 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.

This issue is most salient when defendants face multiple accusations of offenses that are hard to establish beyond a reasonable doubt, such as abuses of power, extortions, and sexual assault. In such these cases, moreover, evidence often centers on consists of witness testimonies.

Legal scholarship has explored the possibility of aggregating offenses into an overall probability of guilt (Cohen

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

1977, Bar Hillel 1984, Robertson and Vignaux 1993). Harel and Porat (2009) define the *Aggregate Probabilities Principle* ("APP") as follows: a defendant is convicted if the probability that he has committed some *unspecified* offence exceeds a given threshold. Compared to the commonly used *Distinct Probabilities Principle* ("DPP"), they argue that APP can reduce adjudication errors, improve deterrence, and reduce the cost of enforcement, and advocate its use in varying degrees in both civil and criminal cases. However, their analysis takes the *joint distribution* of the defendant's guilt across charges as given, and further assumes that guilt is *independently distributed across charges*. This ignores the strategic behavior of potential offenders and witnesses. In particular, how the introduction of APP may affect the incentives of potential offenders and the informativeness of witness testimonies.

This paper considers the effects of APP and DPP from a strategic perspective. In our model, the probability of committing offenses and the informativeness of accusations are both endogenous. Since potential offenders choose their actions strategically, distinct offenses need not be independently distributed. In fact, we show that APP induces *negative correlation* in offenses and, equivalently, witnesses' private information, which severely undermines the informativeness of their reports and weakens deterrence in equilibrium.

In our model, a potential offender (the *principal*) has several opportunities to commit offenses, each of which is associated with a distinct witness (an *agent*) who observes whether the corresponding offense takes place. For example, the principal may be an employer multiple opportunities of violating the law and agents may be employees who are victims or witnesses of these violations and become whistleblowers.

Agents simultaneously decide whether to accuse the principal on the basis of three considerations: (1) a preference for punishing criminal behavior, (2) a risk of social stigma or retaliation, which is higher when accusations fail to get the principal convicted, and (3) some (possibly small) idiosyncratic private benefits or costs of getting the principal convicted, which are independent of whether offenses have taken place.[1] A Bayesian judge observes all the accusations and decide whether to convict or acquit the principal. When contemplating the commission of offenses, the principal trades off the utility from offenses with the expected cost of punishment.

We compare the potential offender's incentive to commit crimes and the informativeness of accusations and silences when either APP or DPP is used to adjudicate guilt. APP stipulates that the principal is convicted if the probability that he has committed at least one offense (even if one cannot specify which one) given all agents' reports, exceeds some exogenous threshold. DPP stipulates that the principal is convicted if there is at least

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

one specific offense for which the probability that the principal committed this offense exceeds some exogenous threshold.

These two criteria are identical when there is only one offense under consideration but differ when the principal may have committed two or more offenses. Understanding the comparative benefits of these criteria is relevant not only for criminal law and civil law but also for corporate decisions such as whether to fire an employee facing multiple allegations of misconduct.

To highlight the main forces at play, we start in section 3 by comparing APP and DPP when (1) the principal can commit two offenses (and there are two agents), and (2) the punishment from conviction is large relative to the benefit from committing crimes. Theorem 1 shows that when APP is used as the adjudication criterion, the principal commits at most one offense in every equilibrium. This strategic restraint induces *negative correlation* in agents' private information: when an agent observes an offense, it reduces the probability that the other agent observed one. This negative correlation exposes agents to the risk of contradicting each other and face retaliation. In equilibrium, this renders agents' reports arbitrarily uninformative and, despite the promise of a large punishment in case of conviction, leads to a high probability of crime.

When DPP is used, by contrast, the principal commits offenses independently and agents' private observations of crime are uncorrelated in equilibrium (Theorem 2). As the punishment in case of conviction becomes arbitrarily large, agents' accusations become arbitrarily informative and the probability of offense vanishes to zero.[2]

The logic of Theorem 1 may be explained more explicitly in two steps. First, when conviction entails a large enough punishment, the principal must be convicted in equilibrium only if *both* agents accuse him (Lemma 3.1). Intuitively, suppose that one accusation sufficed to convict the principal with positive probability. Since, as we shall see, each accusation *strictly increases* the probability that the principal is *guilty of at least one charge*, the principal would then be surely convicted when accused by both agents. With a large punishment in case of conviction, this would give the principal a strict incentive not to commit any crime. This, in turn, would prompt a Bayesian judge never to convict the principal and lead to a contradiction (Lemma 2.1).

This first step implies that the principal's decisions to commit offenses are *strategic substitutes*, because he goes unpunished when he faces only one accusation and is convicted with positive probability when he faces two accusations (Lemma 3.2). It also implies that agents' decisions to accuse the principal are *strategic complements*. This leads us to the second step: In equilibrium, the principal mixes between committing no offense and only one offense. This behavior induces *negative correlation* in agents' private information: an agent who has witnessed an offense believes that the other agent is unlikely of witnessing one, which weakens the first agent's incentive to accuse the principal. Likewise, an agent who has not witnessed an offense assigns a higher probability to the other

[2]In the benchmark model with only one offense and one witness, APP and DPP coincide. Proposition 1 shows that agent's report becomes arbitrarily informative and the probability of offense vanishes as the punishment to the convicted principal increases.

agent having witnessed an offence, which strengthens his incentive to file accusations. This negative correlation, combined with agents' *reporting complementarity*, reduces the informativeness of agents' reports and increases the equilibrium probability of offenses.

For tractability, our argument is presented under the assumption that agents' reports are made publicly and simultaneously. It should be clear that the issues exposed here are not easily addressed by relaxing this assumption. In particular, making agents' reports public only if both agents accuse the principal would remove agents' exposure to retaliation but, by the same token, allow agents to accuse the principal with impunity, and replace false negatives with false positives.[3] In fact, the risk of retaliation can sometimes be instrumental in generating highly informative reports, as we show in Proposition 3. Likewise, if reports were made sequentially, agents' considerations along the sequence would not be very different from those exposed in the simultaneous case, as explained in section 6.[4]

By contrast, DPP restores the informativeness of agents' reports by separating a potential offender's incentives to commit distinct offenses and the correlation between agents' private information. In contrast to the probability that the principal is *guilty of at least one charge*, the probability that he is guilty of a *particular charge* may not increase when a larger set of agents accuse him (e.g., when offenses witnessed by different agents are uncorrelated or negatively correlated). When the probability of conviction is weakly increasing in the set of agents who accuse the principal (a monotonicity refinement), we show that the conviction probability is *linear* in the number of reports and the principal's decisions to commit distinct crimes are neither complements nor substitutes. This implies that the principal has an incentive to commit offenses independently and agents' private observations of crime are uncorrelated. As a result, agents' coordination motive no longer undermines the informativeness of their reports. These linear conviction probabilities respect DPP since the judge's belief about each offense reaches the conviction threshold if and only if the agent who can observe this offense accuses the principal, and is unaffected by reports from other agents.

In practice, one rationale for aggregating different charges against a given defendant is that these charges may shed light on the defendant's propensity to commit offenses, i.e., his type or "character." In section 4, we account for this heterogeneity by allowing the principal to be either a *virtuous type*, with zero or negative marginal benefit from committing crimes, or an *opportunistic type* whose marginal benefit from committing crime is strictly positive. The principal's type is his private information.

[3]The idea of hiding individual reports until multiple accusations are made against an individual has been implemented by the online platform Callisto. Keith Hiatt, director of the Technology Program at the Human Rights Center at UC Berkeley School of Law noted that such platform "may also codify an entrenched attitude that women need to have corroborating evidence to be believed." *New York Times, "The War on Campus Sexual Assault Goes Digital," Nov. 13, 2015.*

[4]Intuitively, the first-moving agent who witnessed an offense may be concerned that the second agent will not accuse the principal, which is more likely to occur if there is negative correlation in offenses. And a second-moving agent is all but certain of not accusing the principal if the first-moving agent has remained silent. Similar issues arise if agents' order of move is uncertain. See also Lee and Suen (2020).

The insights of our Theorems 1 and 2 extend to this setting. When conviction entails a sufficiently large punishment and the judge uses APP, two reports are required to convict the principal in every equilibrium. In contrast to the baseline model, an opportunistic principal *commits two offenses* with strictly positive probability. However, agents' private observations remain *negatively correlated*. As in the baseline model, this negative correlation undermines the informativeness of agents' reports and in equilibrium, leads to a high probability of crime. In contrast, when conviction decisions are made according to DPP, agents' private information is uncorrelated and the principal is convicted with a probability that is linear in the number of accusations against him. As the punishment in case of conviction increases, agents' reports become arbitrarily informative and the probability of crime vanishes to zero.

In section 5, we explore the effects of APP when the principal can commit more than two offenses and, hence, there are more than two agents. We show that for any number of agents, when the punishment to a convicted principal is large enough, the informativeness of each agent's report becomes so weak that conviction occurs only when *all* agents accuse the principal. Theorem 3 shows that, as the number of agents increases, the informativeness of all the reports pooled together *decreases* even though there are more reports available. This leads to an increase in the equilibrium probability of offense. Moreover, the probability that a given agent accuses the principal *increases* with the number of agents. This distinguishes our results from theories of public good provision, in which contributions become scarcer as the number of agents increases.

In summary, our comparisons between APP and DPP suggest a rationale for each accusation being treated independently of other offenses that the defendant may have committed. This implication provides a strategic justification for DPP and echoes some critiques of procedures that link accusations across potential victims.[5]

Rules that resemble APP are also used by firms and organizations considering such as decisions as whether to fire an employee or to sanction one of their member. Those in charge of these decisions face social pressure when a defendant is accused by multiple individuals and is likely to be guilty of at least one charge.

Proposition 3 shows that in these situations, limiting the scale of punishments can, paradoxically, deter crime.[6] In particular, there exist some intermediate punishment levels under which the probability of conviction is *concave* in the number of accusations, and a single accusation suffices to convict the principal with positive probability. The principal's decisions to commit crimes are now *strategic complements*, which induce a *positive correlation* in agents' private signals. This leads to more informative reports and a lower frequency of crime. Nevertheless, such a remedy comes at a cost of increasing the number of offenses conditional on at least one offense taking place. Our

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

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

result thus uncovers a tradeoff between reducing the *probability of crime* and reducing the *severity of crime*.

We study several extensions in section 6 in order to establish the robustness of our results to different variations of our baseline model. This includes (1) the principal facing a larger punishment when the probability that he has committed multiple crimes is above some threshold; (2) the principal faces decreasing marginal benefits from committing multiple offenses; (3) false accusations can be exposed ex post with positive probability, after which false accuser(s) will be punished; (4) agents have an intrinsic preference for reporting the truth or directly care about crimes witnessed by other agents, (5) the principal has private information about the number of opportunities to commit offenses, and (6) agents face a strictly positive cost to accuse the principal even when the latter is acquitted.

Related Literature: This paper contributes to the literatures on information aggregation, strategic communication, and law and economics.

First, our results provide a novel explanation for the failure of information aggregation. In Scharfstein and Stein (1990), Banerjee (1992), et al. (1992), Ottaviani and Sørensen (2000), and Smith and Sørensen (2000), agents fail to act on their private information because they can observe informative actions taken by their predecessors.[7] In contrast, agents cannot observe one another's reports in our model, and the failure of information aggregation is driven by the negative correlation between their private signals and their incentives to coordinate.[8]

When voters' payoffs are negatively correlated, Schmitz and Tröger (2012) show that the majority rule is dominated by other voting rules. Ali, Mihm, and Siga (2018) show that supermajority rules lead to inefficient collective decisions. Our model contrasts to those since the *voting rule* and the *correlation* between agents' private signals are both endogenous. These features of our model differ from the aforementioned models of strategic voting, and existing voting models with endogenous information acquisition (e.g., Persico 2004), in which both the voting rule and the correlation structure are exogenous.

Second, our paper is related to the literature on strategic information transmission with multiple senders (Battaglini 2002, 2017, Ambrus and Takahashi 2008, and Ekmekci and Lauermann 2019). In contrast to these works, senders in our model communicate information about the *principal's action*, and therefore, the correlation between their private signals are endogenous. Our results shed light on how this endogenous correlation structure, combined with the senders' endogenous coordination motives, affects communication informativeness.

Third, our paper contributes to the law and economics literature by (1) studying decision rules that aggregate the probabilities of crime, (2) endogenizing the informativeness and credibility of witness testimonies, and (3)

[7] Information aggregation can also fail due to individual biases (Morgan and Stocken 2008) and voters using pivotal reasoning (Austen-Smith and Banks 1996, Bhattacharya 2013).

[8] Strulovici (2020) studies a sequential learning model in which an agent is less likely to have an informative signal, other things being equal, if another agent has found such a signal. This information attrition may be viewed as a form of negative correlation across agents' signals, which hampers social learning.

analyzing the interplay between an individual's incentive to commit crimes and witnesses' incentives to report the truth.[9] Our results justify a key feature of criminal justice systems, which is to treat distinct accusations separately in conviction decisions. The finding that a lower punishment can reduce the probability of crime stands in contrast to Becker's (1968) well-known observation that maximal punishments save on law-enforcement costs.[10]

Lee and Suen (2020) study the timing of reports by victims and libelers in a model in which a criminal commits crimes against each of the 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 it induces on the agents' private information are distinctive features of our analysis.

2 Model

Overview of the Model: We consider a game between a potential offender (the "principal"), n potential witnesses or victims (the "agents"), and a Bayesian judge, which unfolds in three stages.

In the first stage, the principal privately observes his benefit from committing offenses and then chooses which offenses to commit within a given set of opportunities. For tractability, our analysis considers two principal types: "virtuous" (zero benefit from committing offense) and "opportunistic" (benefit from committing offense is strictly positive).[11] Our results are first presented when the principal is opportunistic with probability 1 (section 3). This allows us to describe as simply as possible the forces at play and show that they do not rely on type heterogeneity.[12] We then extend the analysis to allow for both principal types (section 4). The set of opportunities to commit offenses is fixed in the baseline model. We argue in section 6 that the negative correlation underlying our main result continues to hold when this set is stochastic and privately observed by the principal.

[9] Silva (2019) studies a model with multiple suspects and constructs a mechanism that elicits truthful confessions among suspects. In Baliga, Bueno de Mesquita and Wolitzky (2019), only one of the potential assailants has an opportunity to commit crime. In both papers, the negative correlation in suspects' types is exogenous.

[10] Stigler (1970) observes that several punishment levels should be used when criminals can choose between different levels of crime. The rationale is to provide marginal incentives not to commit the worst crimes. In this scenario, applying the maximal punishment to the worst crimes remains optimal from the perspective of deterring crimes. Siegel and Strulovici (2020) find that extreme punishments are optimal among judicial mechanisms when defendant's type is binary.

[11] One could consider more types; for instance, some ethical principals could have a negative benefit from committing the offense. In our model they would behave as the virtuous ones.

[12] This also addresses an ideological and legal concern about assuming heterogeneous defendant propensities to commit crime. With heterogeneous types, accusing someone of an offense affects the belief about the person's type or "character" and, consequently, the probability of guilt regarding other accusations. This may conflict with rules on forbidding the use of "character evidence", which include the Federal Rule of Evidence 404 in the United States.

In the second stage, each agent observes whether a distinct offense has taken place or not, and each offense opportunity is associated with exactly one agent. Agents independently report what they have observed, possibly lying in either direction. Agent's observations and reports are assumed to be binary throughout the paper.

In the third stage, a judge observes all agents' reports and decides on a punishment based either on the *aggregate probabilities principle* ("APP", defined in (2.4)) and the *distinct probabilities principle* ("DPP", defined in (2.5)). We consider a binary punishment in the baseline model: either the principal is punished, or he is not. In section 6, we discuss situations in which a larger punishment is applied to the principal if he is believed to have committed multiple offenses, which only strengthens our results.

In terms of players' payoffs, the principal trades off his (possibly, null) benefit from committing offenses and the punishment from conviction. Agents choose their reports based on the following considerations: (1) a benefit from convicting a guilty principal, (2) a risk of social stigma or retaliation, which is strictly higher if they accuse the principal but fail to have the latter convicted, (3) some idiosyncratic preference for seeing the principal convicted or acquitted, and, in an extension, (4) a desire for telling the truth. For tractability, all components of the principal's and the agents' payoffs are additively separable. To make sure that all message profiles occur on the equilibrium path, we also allow for an infinitesimally small fraction of behavioral agents, who always accuse or always abstain from accusing the principal. We do not explicitly model the judge's payoff. However, notice that both decision rules can arise from maximizing some quadratic payoff functions.

2.1 Baseline Model

Consider a three-stage game between a principal, n agents, and a judge. In stage 1, the principal privately observes his type $t \in \{t^v, t^o\}$, in which t^v stands for *virtuous type* and t^o stands for *opportunistic type*. He then chooses an n-dimensional vector $\boldsymbol{\theta} \equiv (\theta_1, ..., \theta_n) \in \{0,1\}^n$, where $\theta_i = 1$ (respectively, $\theta_i = 0$) means that the principal commits (does not commit) an offense witnessed by (or against) agent i.

In stage 2, each agent $i \in \{1, 2, ..., n\}$ *privately observes* θ_i, the realization of a payoff shock $\omega_i \in \mathbb{R}$, and whether he is strategic or behavioral. Each agent then chooses between accusing the principal ($a_i = 1$) or not ($a_i = 0$). If agent i is strategic, he chooses a_i to maximize his expected payoff. A fraction $\alpha \in (0, 1)$ of the behavioral agents always accuses the principal, and a fraction $1 - \alpha$ never accuses the principal.[13]

In stage 3, the judge observes $\boldsymbol{a} \equiv (a_1, ..., a_n) \in \{0,1\}^n$, and chooses $s \in \{0,1\}$, with $s = 1$ stands for convicting the principal and $s = 0$ stands for acquitting the principal.

[13]The presence of behavioral agents allows us to rule out reports that are off the equilibrium path. They help to refine away unreasonable equilibria in which the principal commits offenses against all agents with probability 1 and is surely convicted even when no agent reports. Our results continue to hold under alternative specifications of the behavioral agents' strategies. For example, we allow the behavioral type of different agents to use different strategies, the behavioral agents' reports to be responsive to their observations of θ_i, and so on.

Type Distributions: The exogenous random variables t, $\{\omega_i\}_{i=1}^n$, and whether an agent is strategic or behavioral are independently distributed. Let $\pi^o \in (0,1]$ be the probability that the principal is opportunistic. Each ω_i is drawn from a normal distribution with mean μ and variance σ^2, with $\Phi(\cdot)$ and $\phi(\cdot)$ denote their cdf and pdf, respectively. Each agent is strategic with probability $\delta \in (0,1)$. We are interested in situations in which δ is arbitrarily close to 1, which means that the fraction of behavioral agents is positive but arbitrarily small.

Payoffs: The principal's payoff is

$$\mathbf{1}\{t = t^o\} \cdot \sum_{i=1}^{n} \theta_i - sL. \tag{2.1}$$

According to (2.1), the virtuous type does not benefit from committing offenses. The opportunistic type tradeoffs the benefit from committing each offense (normalized to 1) with the cost of being convicted $L > 0$. Strategic agent i's payoff is:

$$u_i(\omega_i, \theta_i, a_i) \equiv \begin{cases} 0 & \text{if} \quad s = 1 \\ \omega_i - b\Big((1-\gamma)\theta_i + \gamma f(\theta_i, \theta_{-i})\Big) - ca_i & \text{if} \quad s = 0, \end{cases} \tag{2.2}$$

where $b > 0$, $c > 0$, and $\gamma \in [0,1]$ are parameters, and $f(\theta_i, \theta_{-i})$ is strictly increasing in both arguments.

According to (2.2), each agent's utility when the principal is convicted is normalized to 0. This is without loss of generality for the purpose of *analyzing agents' incentives*, since each strategic agent's reporting decision depends only on the *difference* between his utility when the principal is convicted and his utility when the principal is acquitted.[14] Such utility difference depends on $b > 0$ that measures an agent's preference for punishing guilty people, $c > 0$ that measures an agent's loss from social stigma or from the principal's retaliation when his accusation fails to convict the principal,[15] $\gamma \in [0,1]$ that measures the intensity of an agent's *social preference*, i.e., the extent to which his payoff internalizes offenses committed against (or witnessed by) other agents, and the realization of his payoff shock ω_i that measures how much agent i dislikes the principal. Intuitively, when ω_i is lower, agent i dislikes the principal more, and has stronger motives to file an accusation.

We compare the equilibrium outcomes when the judge uses the following two decision rules: APP and DPP. These rules differ only in terms of how to measure the probability of guiltiness when the principal is capable of committing *multiple offenses*. Both of them can arise endogenously when the judge maximizes some quadratic payoff functions. Let $\pi^* \in (0,1)$ be an exogenous parameter that measures the judge's standard of proof, and let

$$\overline{\theta} \equiv \max_{i \in \{1,2,\ldots,n\}} \theta_i, \tag{2.3}$$

[14]This statement rules out agents' intrinsic preference to report, for example, each agent strictly benefits from reporting the truth irrespective of the judicial outcome. We study this extension in section 6 and establish the robustness of our results.

[15]In the baseline model, we normalize the social stigma (or loss from retaliation) to zero when the principal is convicted. In general, our results apply as long as the social stigma (or loss from retaliation) is strictly larger when the principal is acquitted rather than convicted.

i.e., the principal *has committed at least one offense* when $\overline{\theta} = 1$, and has committed no offense otherwise.

The first decision rule, APP, *aggregates* the probabilities of committing different offenses and convicts the principal when it is beyond reasonable doubt that the latter has committed *at least one offense*:

$$s \begin{cases} = 1 & \text{if} \quad \Pr\left(\overline{\theta} = 1 \middle| \boldsymbol{a}\right) > \pi^* \\ \in \{0, 1\} & \text{if} \quad \Pr\left(\overline{\theta} = 1 \middle| \boldsymbol{a}\right) = \pi^* \\ = 0 & \text{if} \quad \Pr\left(\overline{\theta} = 1 \middle| \boldsymbol{a}\right) < \pi^*. \end{cases} \tag{2.4}$$

Such a decision rule is frequently used when an individual is accused of wrongdoing outside the criminal process, for example, when managers are charged with discriminating against minority workers, or when supervisors are charged with abusing their subordinates. Schauer and Zeckhauser (1996) argue that the cumulation of multiple low-probability accusations is often appropriate in corporate and other non-judicial settings. Harel and Porat (2009, p. 263) advocate its use in judicial settings by arguing that *acquitting defendants who are almost surely guilty of some unspecified crime is neither just nor efficient*.

The second decision rule is typically used in the criminal justice system, known as the *distinct probabilities principle* or DPP, that when a defendant is charged with a number of offenses, the court examines each charge individually to decide whether a beyond-a-reasonable-doubt standard is satisfied:

$$s \begin{cases} = 1 & \text{if exists } i \in \{1, 2, ..., n\} \text{ s.t. } \Pr\left(\theta_i = 1 \middle| \boldsymbol{a}\right) > \pi^* \\ \in \{0, 1\} & \text{if exists no } i \in \{1, 2, ..., n\} \text{ s.t. } \Pr\left(\theta_i = 1 \middle| \boldsymbol{a}\right) > \pi^*, \text{ but exists } i \in \{1, 2, ..., n\} \text{ s.t. } \Pr\left(\theta_i = 1 \middle| \boldsymbol{a}\right) = \pi^* \\ = 0 & \text{if exists no } i \in \{1, 2, ..., n\} \text{ s.t. } \Pr\left(\theta_i = 1 \middle| \boldsymbol{a}\right) \geq \pi^*. \end{cases}$$

This is equivalent to:

$$s \begin{cases} = 1 & \text{if} \quad \max_{i \in \{1,2,...,n\}} \Pr\left(\theta_i = 1 \middle| \boldsymbol{a}\right) > \pi^* \\ \in \{0, 1\} & \text{if} \quad \max_{i \in \{1,2,...,n\}} \Pr\left(\theta_i = 1 \middle| \boldsymbol{a}\right) = \pi^* \\ = 0 & \text{if} \quad \max_{i \in \{1,2,...,n\}} \Pr\left(\theta_i = 1 \middle| \boldsymbol{a}\right) < \pi^*. \end{cases} \tag{2.5}$$

When there is only one agent, (2.4) and (2.5) coincide. When there are two or more agents, (2.4) and (2.5) are different. In our example of the introduction, when $\pi^* = 0.95$, the defendant who commits two offenses with probability 0.8 each is convicted under decision rule (2.4) but is acquitted under decision rule (2.5).

Solution Concept: We examine proper equilibria (Myerson 1978) that can survive *two refinements*, which we refer to as *equilibrium* for short. Formally, a proper equilibrium consists of a strategy profile $\left\{\sigma^o, \sigma^v, (\sigma_i)_{i=1}^n, q\right\}$:

- $\sigma^o \in \Delta\left(\{0, 1\}^n\right)$ is the opportunistic-type principal's strategy, and $\sigma^v \in \Delta\left(\{0, 1\}^n\right)$ is the virtuous-type principal's strategy, both are distributions of $\boldsymbol{\theta}$;

- $\sigma_i : \mathbb{R} \times \{0,1\} \to \Delta\{0,1\}$ is agent i's strategy, and maps the realization of ω_i and θ_i to the probability with which i accuses the principal,

- $q : \{0,1\}^n \to [0,1]$ is the judge's strategy in which $q(\boldsymbol{a}) \equiv \mathbb{E}[s|\boldsymbol{a}]$ is the probability with which he convicts the principal after observing $\boldsymbol{a} \in \{0,1\}^n$.

Our first refinement requires that the principal is acquitted if no agent reports against him.

Refinement 1 (No Conviction Unless Accused)**.** $q(0,0,...,0) = 0$.

The purpose for Refinement 1 is to rule out equilibria in which the opportunistic-type principal commits offenses against all agents with probability one and is convicted for sure irrespective of the agents' reports. We rule out those equilibria since they violate the principle that a defendant should not be convicted based on the sole basis of a judge's prior belief. Our second refinement endows the agents' messages with meanings.

Refinement 2 (Monotonicity)**.** *For every* $i \in \{1,2,...,n\}$*,* $q(1,a_{-i}) \geq q(0,a_{-i})$ *for every* $a_{-i} \in \{0,1\}^{n-1}$*, and there exists* $a_{-i} \in \{0,1\}^{n-1}$ *such that* $q(1,a_{-i}) > q(0,a_{-i})$*.*

Refinement 2 requires that first, the conviction probability to be nondecreasing when a larger subset of agents accuse the principal, and second, each agent's accusation has a *nontrivial influence* on the conviction probabilities under at least some reporting profile of other agents. This refinement implies that each accusation is a move against the principal. It resonates our interpretation of c as an agent's loss from the principal's retaliation. Intuitively, a principal retaliates against messages that reduce his payoff and does not retaliate against other messages.[16]

Lemma 2.1 establishes the existence of equilibrium that survives both refinements and lists several preliminary observations, all of which apply to *both decision rules* and to *any number of agents*.

Lemma 2.1. *There exists* $\overline{L} \in \mathbb{R}_+$ *such that for every* $L \geq \overline{L}$*, there exists a proper equilibrium that survives Refinements 1 and 2. In every such equilibrium:*

1. *For every* $i \in \{1,2,...,n\}$*, agent* i*'s strategy is characterized by* ω_i^* *and* ω_i^{**} *with* $\omega_i^* > \omega_i^{**}$*, such that when* $\theta_i = 1$*, agent* i *reports if and only if* $\omega_i \leq \omega_i^*$*; when* $\theta_i = 0$*, agent* i *reports if and only if* $\omega_i \leq \omega_i^{**}$*.*

2. *The probability of committing offense* $\Pr\left(\overline{\theta} = 1\right)$ *is strictly between* 0 *and* 1*.*[17]

[16]A microfoundation for this monotonicity refinement is provided by Chassang and Padró i Miquel (2019), in which the principal optimally commits to a retaliation plan $\widetilde{c}_i : \{0,1\} \to [0,c]$ privately against agent i, which maps agent i's report to his loss from the principal's retaliation. Retaliation can only be carried out when the principal is acquitted, and c is the maximal damage the principal can inflict on each agent. The principal's optimal retaliation plan is to retaliate to the maximum against the message that increases his probability of being convicted and not to retaliate against the other message.

[17]The observation that offenses are committed with positive probability is related to albeit different from a well-known result in inspection games (Dresher 1962), in which players misbehave with positive probability due to the inspector's cost of inspection. In our model, positive probability of offense is not driven by the cost of inspection, but is driven by the requirement that the principal is convicted *only when* the posterior probability with which he is guilty exceeds some exogenous $\pi^* \in (0,1)$.

3. *The virtuous-type principal never commits any offense, i.e.,* σ^v *attaches probability* 1 *to* $\boldsymbol{\theta} = (0, ..., 0)$.

The proof is in Appendix A, except for the one on existence in the supplementary appendix The existence of equilibrium requires L to be large enough due to Refinement 1. For example, if $L < 1$ and $\pi^o > \pi^*$, then in every proper equilibrium, the opportunistic principal commits offense against all agents with probability 1 and is convicted even when no agent reports. In this case, no Nash equilibrium survives Refinement 1.

To explore how large L needs to be for our existence result, we perform numerical simulations. For example, when $\pi^* = 0.95$, ω_i follows a normal distribution with mean 0 and variance 1, the fraction of strategic agents is $\delta = 0.95$, $b = 1$ and $c = 10$, a proper equilibrium that survives both refinements exists when $L \geq 5$.

2.2 Discussion of Assumptions

Decision Rules & Conviction Cutoffs: Our analysis takes the conviction cutoff π^* as exogenous and focuses on the comparison between the equilibrium outcomes under APP and DPP. Our results establish the connections between π^* and the effectiveness of deterrence as well as the *quality of judicial decisions*, measured by:

1. the probability of being innocent conditional on being convicted,

2. and the probability of being guilty conditional on being acquitted.

From this perspective, our results also shed light on the tradeoffs a society faces when setting π^* endogenously. For example, we show that under both decision rules, a fraction $1 - \pi^*$ of convicted people are innocent and a fraction π^* of acquitted people are guilty. This suggests that by lowering π^*, the judge increases the probability of convicting guilty people, at the expense of increasing the fraction of innocent people being convicted. Depending on the decision rule, the equilibrium probability of crime is close to either 0 or π^*. As a result, our cutoff decision rules can arise from *maximizing a social welfare function* that trades off the effectiveness of crime deterrence with the fraction of false positives among those who are convicted.

Binary Conviction Decision *vs* General Punishment Function: In our baseline model, the conviction decision is binary. In principle, the punishment of conviction can depend not only on the probability with which the defendant is guilty, but also on the number of offenses he is guilty of. For example, when it is beyond reasonable doubt that a defendant has committed at least two offenses, the punishment should be higher compared to the case in which he has committed only one offense.

Our main result (Theorem 1) is robust with respect to such concerns: When the punishment of conviction is large enough, an opportunistic principal *commits at most one offense in equilibrium even when the punishment for* *committing multiple offenses is the same as the punishment for committing a single offense.* As a result, increasing the punishment for committing multiple offenses only strengthens our predictions.

Simultaneous-Move *vs* Sequential-Move: In our baseline model, agents simultaneously decide whether to file accusations against a potential offender. In practice, such decisions are often made *sequentially* and agents can often choose not only whether to file an accusation but also *when* to file it. We explain in section 6 that the forces underlying our results are also present in these dynamic versions of our model.

3 Main Results

We show that when there are *multiple potential witnesses*, APP leads to uninformative reports and ineffective deterrence (Theorem 1), while DPP leads to informative reports and effective deterrence (Theorem 2).

To highlight the forces at play in their simplest form, this section assumes that the principal is surely opportunistic, i.e., $\pi^o = 1$, agents are not altruistic, i.e., $\gamma = 0$, and compares decision rules (2.4) and (2.5) when there are *one or two potential witnesses*. The results are generalized to multiple principal types in section 4, three or more agents in section 5, and altruistic agents in section 6.

3.1 Benchmark: Single Potential Witness

When there is only one agent, Refinement 1 implies that the principal is convicted with strictly positive probability only when the agent reports, which we denote by $q(1)$. If $\theta = 1$, then the agent prefers to report when:

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

If $\theta = 0$, then the agent prefers to report when:

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

Let $\Pr(\boldsymbol{\theta} = 1)$ be the prior probability that an offense has taken place, and $\Pr(\boldsymbol{\theta} = 1|\boldsymbol{a} = 1)$ be the judge's posterior belief after observing the agent reports. According to Bayes Rule,

$$\underbrace{\frac{\Pr(\text{agent reports} \mid \theta = 1)}{\Pr(\text{agent reports} \mid \theta = 0)}}_{\equiv \mathcal{I}} \cdot \frac{\Pr(\boldsymbol{\theta} = 1)}{1 - \Pr(\boldsymbol{\theta} = 1)} = \frac{\Pr(\boldsymbol{\theta} = 1|\boldsymbol{a} = 1)}{1 - \Pr(\boldsymbol{\theta} = 1|\boldsymbol{a} = 1)}. \tag{3.3}$$

The above formula suggests that $\mathcal{I}$ is a sufficient statistic for the judge's posterior belief after observing the agent's report, i.e., it measures the *informativeness* of his report about crime.

Proposition 1. *Suppose $n = 1$. There exists $\overline{L} > 0$ such that in every equilibrium when $L > \overline{L}$,*

1. *The judge's assigns probability π^* to the principal being guilty when the agent accuses the principal.*
2. *The informativeness $\mathcal{I}$ converges to infinity and the prior probability of crime $\Pr(\boldsymbol{\theta} = 1)$ vanishes to 0.*[18]

The proof is in Appendix B. Proposition 1's second statement suggests that as the principal's punishment in case of conviction becomes large relative to the benefit from committing crime, the agent's report becomes arbitrarily informative, and the equilibrium probability of crime vanishes to 0.

This result may be understood as follows: In equilibrium, the probability of crime is interior (Lemma 2.1), and the principal must be indifferent between committing crime and not committing crime. This indifference condition suggests that $q(1)$ vanishes to 0 as L goes to infinity. According to (3.1) and (3.2), the agent's reporting cutoffs are decreasing in L, and moreover, the distance between the two cutoffs $\omega^* - \omega^{**}$ equals b.

As the fraction of behavioral agent $1-\delta$ vanishes to 0, the informativeness ratio $\mathcal{I}$ is approximately $\Phi(\omega^*)/\Phi(\omega^* - b)$. Given that $\omega^* - \omega^{**} = b$, $\Phi(\omega^*)/\Phi(\omega^* - b)$ goes to infinity as $\omega^* \to -\infty$. Since $q(1)$ is strictly between 0 and 1 when L is large enough, the judge's posterior belief about crime equals π^* after observing the agent's report. Equation (3.3) then suggests that the prior probability of crime vanishes to 0.

3.2 Two Potential Witnesses: Aggregating the Probabilities of Distinct Crimes

We examine equilibrium outcomes under decision rule (2.4) when there are *two agents*. Recall the definition of $\overline{\theta}$ in (2.3). Bayes Rule implies the following equation that connects the prior probability of crime $\Pr(\overline{\theta} = 1)$ with the posterior probability of crime $\Pr(\overline{\theta} = 1|\boldsymbol{a})$:

$$\frac{\Pr(\boldsymbol{a}|\overline{\theta} = 1)}{\Pr(\boldsymbol{a}|\overline{\theta} = 0)} \cdot \frac{\Pr(\overline{\theta} = 1)}{1 - \Pr(\overline{\theta} = 1)} = \frac{\Pr(\overline{\theta} = 1|\boldsymbol{a})}{1 - \Pr(\overline{\theta} = 1|\boldsymbol{a})}. \tag{3.4}$$

Therefore,

$$\mathcal{I}(\boldsymbol{a}) \equiv \frac{\Pr(\boldsymbol{a}|\overline{\theta} = 1)}{\Pr(\boldsymbol{a}|\overline{\theta} = 0)}, \tag{3.5}$$

is a sufficient statistic for the judge's posterior belief about crime after observing $\boldsymbol{a}$, which measures the *informativeness* of reporting profile $\boldsymbol{a}$, i.e., a larger $\mathcal{I}(\boldsymbol{a})$ means that crime is more likely to take place after observing $\boldsymbol{a}$:

Theorem 1. *When $n = 2$ and the judge uses decision rule (2.4), there exists $\overline{L}$ such that when $L > \overline{L}$,*

[18] Throughout the paper, the limit is first taken first with respect to δ and then with respect to L. For example, Proposition 1 states that $\lim_{L \to +\infty}(\lim_{\delta \to 1} \Pr(\boldsymbol{\theta} = 1)) = 0$.

1. ***Endogenous Negative Correlation Between Crimes:*** *In every equilibrium*

$$\Pr(\theta_1 = 1|\theta_2 = 1) < \Pr(\theta_1 = 1|\theta_2 = 0) \text{ and } \Pr(\theta_2 = 1|\theta_1 = 1) < \Pr(\theta_2 = 1|\theta_1 = 0). \tag{3.6}$$

For every $\varepsilon > 0$, there exists $\overline{L}_\varepsilon \in \mathbb{R}_+$, such that in every equilibrium when $L > \overline{L}_\varepsilon$ and $\delta \in (0,1)$,

2. ***Low Informativeness of Reports & Ineffective Deterrence:*** $\max_{\boldsymbol{a}\in\{0,1\}^2} \mathcal{I}(\boldsymbol{a}) < 1 + \varepsilon$.

3. ***Ineffective Deterrence:*** $\Pr(\overline{\theta} = 1) > \pi^* - \varepsilon$.

Theorem 1 suggests that when conviction punishment L is large relative to the benefit from committing crime, aggregating the probabilities of distinct crimes leads to an *endogenous negative correlation* between crimes. Such a negative correlation *undermines the informativeness* of agents' reports, and results in a *high probability of crime*. In the limit where $L \to \infty$,[19] the agents' reports become arbitrarily uninformative and the equilibrium probability of crime converges to π^*. These conclusions contrast to the single-agent benchmark in which the probability of crime vanishes to 0 and the informativeness ratio goes to infinity as the conviction punishment L increases.

Theorem 1 suggests that deterrence may be improved by lowering the conviction threshold π^*, which was taken as exogenous given. However, lowering π^* increases the risk of convicting an innocent defendant. In our model, the probability that a convicted defendant is innocent is equal to $1 - \pi^*$ in equilibrium. For example, if any judge were to use a conviction threshold π^* of 10%, then in every equilibrium, each convicted individual has a 90% change of being innocent. In general, the conviction threshold may be chosen so as to strike a compromise between deterrence and wrongful convictions, and different values of π^* may be justified by a different welfare costs of punishing the innocent.

One may conjecture that low witness credibility and ineffective deterrence are driven by agents' incentives to free-ride on others' reports, which is the intuition behind inefficient public good provision (Chamberlin 1974). The following comparative static result refutes this conjecture, which suggests that by fixing all other parameters of the game and increasing the number of agents from one to two, each agent reports with *strictly higher probability* in every equilibrium. This finding is generalized to three or more agents in Theorem 3.

Proposition 2. *When the judge uses decision rule (2.4), there exists $\overline{L} > 0$ such that for every $L > \overline{L}$,*

- *if $\omega^*_{i,2}(L)$ and $\omega^{**}_{i,2}(L)$ are agent i's equilibrium reporting cutoffs in a two-agent setting, and $\omega^*(L)$ and $\omega^{**}(L)$ are the agent's equilibrium reporting cutoffs in a single-agent setting, then $\omega^*_{i,2}(L) > \omega^*(L)$ and $\omega^{**}_{i,2}(L) > \omega^{**}(L)$ for every $i \in \{1,2\}$.*

[19] According to the statement of Theorem 1, the same conclusion applies under the double limit in Proposition 1 $\lim_{L\to\infty}\lim_{\delta\to 1}$.

In what follows, we explain why agents' private observations of crime are *negatively correlated*, and how this endogenous negative correlation interacts with *agents' coordination motives*.

Step 1: Equilibrium Conviction Probability We start from an observation that when L is large enough, the principal is convicted with positive probability *only when* he is accused by both agents:

Lemma 3.1. *There exists $\overline{L} \in \mathbb{R}_+$ such that when $L > \overline{L}$ and in every equilibrium,*

$$q(0,0) = q(1,0) = q(0,1) = 0 \text{ and } q(1,1) \in (0,1). \tag{3.7}$$

The proof is in Online Appendix A. For an intuitive explanation, suppose by way of contradiction that a single accusation suffices to convict the principal with strictly positive probability. Since each agent is strictly more likely to report when he has witnessed a crime, each additional report makes the principal strictly more likely of having committed at least some crime. This suggests that the principal is surely convicted if two reports are leveled against him. The technical step of our proof is to show under the above presumption, the increase in conviction probability when the principal commits an extra crime is bounded from below. As the magnitude of punishment L increases, the marginal cost from committing crime eventually exceeds the marginal benefit, in which case the opportunistic-type principal has a strict incentive not to commit any crime. This contradicts the conclusion of Lemma 2.1 that the equilibrium probability of crime is strictly positive, which establishes (3.7).

Step 2: Principal's Incentives & Endogenous Negative Correlation We examine the opportunistic principal's incentives to commit crimes, and establish a *necessary and sufficient condition* for his choices of θ_1 and θ_2 to be strategic substitutes or strategic complements.

Lemma 3.2. *The opportunistic principal's choices of θ_1 and θ_2 are strategic substitutes if and only if*

$$q(1,1) + q(0,0) - q(1,0) - q(0,1) > 0, \tag{3.8}$$

and are strategic complements if and only if $q(1,1) + q(0,0) - q(1,0) - q(0,1) < 0$.

Proof of Lemma 3.2: For $i \in \{1,2\}$, let $\Psi_i^* \equiv \delta\Phi(\omega_i^*) + (1-\delta)\alpha$ and let $\Psi_i^{**} \equiv \delta\Phi(\omega_i^{**}) + (1-\delta)\alpha$. The difference in expected conviction probability between $(\theta_1,\theta_2) = (0,0)$ and $(\theta_1,\theta_2) = (1,0)$ is:

$$(\Psi_1^* - \Psi_1^{**})\Big((1-\Psi_2^{**})\big(q(1,0) - q(0,0)\big) + \Psi_2^{**}\big(q(1,1) - q(0,1)\big)\Big). \tag{3.9}$$

The difference in expected conviction probability between $(\theta_1, \theta_2) = (0, 1)$ and $(\theta_1, \theta_2) = (1, 1)$ is:

$$(\Psi_1^* - \Psi_1^{**})\Big((1 - \Psi_2^*)\big(q(1,0) - q(0,0)\big) + \Psi_2^*\big(q(1,1) - q(0,1)\big)\Big). \tag{3.10}$$

The choices of θ_1 and θ_2 are strategic substitutes if and only if (3.9) is less than (3.10), or equivalently,

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

Since $\omega_i^* > \omega_i^{**}$ for every $i \in \{1, 2\}$, we have $(\Psi_1^* - \Psi_1^{**})(\Psi_2^* - \Psi_2^{**}) > 0$, which implies (3.8). □

Lemma 3.1 and Lemma 3.2 imply that in every equilibrium when L is large enough, the conviction probabilities satisfy (3.8). This suggests that the opportunistic principal's decisions are strategic substitutes, and therefore, he *cannot* be indifferent between committing no crime and committing two crimes. Given the conclusion of Lemma 2.1 that the equilibrium probability of crime is interior, the opportunistic principal is indifferent between committing no crime and committing one crime.[20] This leads to a negative correlation between θ_1 and θ_2.

Step 3: Agents' Coordination Motives From (3.7), the principal is convicted only if both agents accuse the principal. To lower the loss from retaliation, each agent has a stronger incentive to accuse the principal when he expects the other agent to accuse the principal with higher probability. Formally, if $\theta_i = 1$, then agent i prefers to report when:

$$\omega_i \leq \omega_i^* \equiv b - c\frac{1 - q(1,1)Q_{1,j}}{q(1,1)Q_{1,j}} = b + c - \frac{c}{q(1,1)Q_{1,j}}, \tag{3.11}$$

where $Q_{1,j}$ is the probability that agent $j \neq i$ accuses the principal *conditional on* $\theta_i = 1$. Similarly, if $\theta_i = 0$, then agent i prefers to report when:

$$\omega_i \leq \omega_i^{**} \equiv -c\frac{1 - q(1,1)Q_{0,j}}{q(1,1)Q_{0,j}} = c - \frac{c}{q(1,1)Q_{0,j}}, \tag{3.12}$$

where $Q_{0,j}$ is the probability that agent $j \neq i$ accuses the principal *conditional on* $\theta_i = 0$. One can verify that ω_i^* is strictly increasing in $Q_{1,j}$, and ω_i^{**} is strictly increasing in $Q_{0,j}$.

Given that each agent is more likely to report when he has witnessed a crime, the negative correlation between θ_1 and θ_2 suggests that $Q_{1,j} < Q_{0,j}$, which implies that $\omega_i^* - \omega_i^{**} \in (0, b)$. We show in Appendix C that $\omega_i^* - \omega_i^{**} \to 0$ as $L \to \infty$. This contrasts to the single-agent benchmark in which the distance between two

[20]When the principal is virtuous with probability greater than $1 - \pi^*$, we argue in section 4 that the opportunistic type principal is indifferent between committing one crime and committing two crimes, and the opportunistic-type principal commits two crimes with strictly positive probability. Nevertheless, we show that the agents' private observations of crime remain *negatively correlated*, i.e., the main insights of Theorem 1 remain robust.

reporting cutoffs equals b. The decrease in the distance between reporting cutoffs undermines the informativeness of agents' reports, measured by $\max_{\boldsymbol{a}\in\{0,1\}^2}\mathcal{I}(\boldsymbol{a})$, which converges to 1 as $L\to\infty$. According to (3.7), the judge's posterior belief about crime equals π^* after observing $\mathbf{a}=(1,1)$. According to (3.4),

$$\underbrace{\frac{\Pr(\boldsymbol{a}=(1,1)|\overline{\theta}=1)}{\Pr(\boldsymbol{a}=(1,1)|\overline{\theta}=0)}}_{\equiv\mathcal{I}(1,1)=\max_{\mathbf{a}\in\{0,1\}^2}\mathcal{I}(\mathbf{a})}\cdot\frac{\Pr(\overline{\theta}=1)}{1-\Pr(\overline{\theta}=1)}=\frac{\pi^*}{1-\pi^*},$$

which suggests that the prior probability of crime $\Pr(\overline{\theta}=1)$ is close to π^* when L is large enough.

3.3 Two Potential Witnesses: Separating the Probabilities of Distinct Crimes

We examine equilibrium outcomes when there are *two agents* and the judge's conviction decision respects the distinct probabilities principle, i.e., based on the probability that the principal is guilty of *each individual crime*.

Theorem 2. *When $n=2$ and the judge uses decision rule (2.5), there exists $\overline{L}\in\mathbb{R}_+$ such that in every equilibrium when $L>\overline{L}$,*

1. ***Uncorrelated Crimes:*** , $\Pr(\theta_i=1|\theta_j=1)=\Pr(\theta_i=1|\theta_j=0)$ *for every $i\neq j$.*
2. ***Linear Conviction Probability:*** $\Pr(\theta_i=1|a_i=1)=\pi^*$ *for every $i\in\{1,2\}$, and the conviction probability is linear in the number of reports.*

As $L\to+\infty$ the following asymptotic results hold.[21]

3. ***Effective Deterrence:*** *The equilibrium probability of crime* $\Pr(\overline{\theta}=1)$ *converges to* 0.
4. ***Highly Informative Reports:*** *For every $i\in\{1,2\}$, the informativeness of agent i's report about θ_i, measured by $\frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)}$, goes to ∞.*[22]

The proof is in Appendix D. According to Theorem 2, potential witnesses' private observations of crimes are *uncorrelated* and the conviction probability must be a linear function of the number of reports. By Lemma 3.2, the principal's incentives to commit different crimes are neither complements nor substitutes, and consequently he has an incentive to commit different crimes independently. This preserves the credibility of the agents' reports, which lowers the probability of crime in equilibrium. We explain the intuition behind Theorem 2 in three steps, focusing primarily on why agents' private observations of crime are uncorrelated, and why uncorrelated private information leads to linear conviction probabilities, informative reports, and effective deterrence in equilibrium.

[21] As noted in Proposition 1, the results are obtained by first taking the limit with respect to δ and then with respect to L.

[22] The same conclusion applies under other measures of informativeness, for example, the one developed in Theorem 1, in which case one can show that $\max_{\boldsymbol{a}\in\{0,1\}^2}\mathcal{I}(\boldsymbol{a})$ also goes to infinity in the $L\to\infty$ limit.

Ruling Out Correlations: Suppose first by way of contradiction that θ_1 and θ_2 are *negatively correlated*. In this case,

$$\Pr(\theta_1 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_1 = 1|\boldsymbol{a} = (1,0)).$$

because an accusation by agent 2 increases the probability that 2 has observed an offense and, given the negative correlation between offenses, reduces the probability that agent 1 has observed one. A similar logic implies that $\Pr(\theta_2 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_2 = 1|\boldsymbol{a} = (0,1))$. Under decision rule (2.5), the above inequalities imply that $q(1,0) \geq q(1,1)$ and $q(0,1) \geq q(1,1)$. For this to survive our Refinements 1 and 2, it has to be the case that

$$q(1,1) = q(1,0) = q(0,1) = 1 \text{ and } q(0,0) = 0. \tag{3.13}$$

When L is large enough, one can show similar to Lemma 3.1 that under such conviction probabilities, the opportunistic principal has a strict incentive not to commit any crime. This contradicts the conclusion of Lemma 2.1 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 decision rule (2.5),

$$q(1,1) \geq \max\{q(1,0), q(0,1)\} \geq q(0,0) = 0, \tag{3.14}$$

with the first inequality being strict unless $\max\{q(1,0), q(0,1)\} = 1$. Similar to the logic behind Lemma 3.1, the opportunistic principal's incentive to commit crime in equilibrium suggests that $q(1,1) \in (0,1)$ when L is large enough. When $q(1,1) \in (0,1)$, the positive correlation between θ_1 and θ_2 implies that $q(0,1) = q(1,0) = 0$. Lemma 3.2 implies that the opportunistic-type principal's decisions to commit crimes are *strategic substitutes*. This contradicts the presumption that θ_1 and θ_2 are positively correlated.

Linear Conviction Probabilities: Given that θ_1 and θ_2 are *uncorrelated*, the principal's incentives to commit distinct crimes are neither substitutes or complements. Lemma 3.2 and Refinement 1 suggest that $q(1,1) = q(1,0) + q(0,1)$. Since the equilibrium probability of crime is interior (Lemma 2.1), the opportunistic principal plays a completely mixed strategy in equilibrium. Taken together, these observations imply that conviction probabilities are symmetric, i.e., $q(1,0) = q(0,1)$. Otherwise, one of the strategies $\boldsymbol{\theta} = (1,0)$ and $\boldsymbol{\theta} = (0,1)$ would be strictly suboptimal for an opportunistic principal.

Informative Report & Effective Deterrence: The fact that θ_1 and θ_2 are uncorrelated also suggests that each agent's belief about whether the other agent has witnessed a crime is independent of his own observation of crime.

According to the formulas for agents' reporting cutoffs (3.11) and (3.12), the independence of θ_1 and θ_2 implies that $Q_{1,j} = Q_{0,j}$ for $j \in \{1, 2\}$, which further suggests that $\omega_j^* - \omega_j^{**} = b$.

Similar to the single-agent benchmark, both reporting cutoffs converge to $-\infty$ as L becomes large, and the informativeness ratio, which is approximately $\Phi(\omega^*)/\Phi(\omega^* - b)$, goes to $+\infty$. Given that the judge's posterior belief about $\theta_i = 1$ equals π^* after observing $a_i = 1$, the informativeness ratio converging to $+\infty$ implies that the prior probability with which $\theta_i = 1$ converges to 0.

Remark: The comparison between Theorem 2 and Theorem 1 provides a justification for the use of DPP in criminal justice systems. It provides a rationale for several aspects of the law, for example, judges are not allowed to aggregate the probabilities of different criminal behaviors when making conviction decisions, and are prohibited from using character evidence to establish the guiltiness of defendants.

Relative to earlier writings of legal scholars that discuss the advantages and disadvantages of aggregating probabilities across cases (e.g., Schauer and Zeckhauser 1996, Harel and Porat 2009), the contribution of our analysis is to take potential offenders' and potential witnesses' strategic motives into account. Our results suggest that aggregating probabilities across criminal cases affects the *correlations* between crimes, which can undermine *the quality of evidence* available to judges when they make conviction decisions.

Nevertheless, implementing DPP in practice requires decision makers (e.g., judges, board of trustees of a firm, etc.) to have commitment power. This is illustrated by the numerical example in the introduction, and is also clearly explained by the following thought experiment: Consider a setting with two defendants and four potential victims. The probabilities that each defendant is guilty of committing crime against the potential victims are:

	Pr(crime against individual 1)	Pr(crime against individual 2)	Pr(crime against at least one person)
Defendant 1	49.5 %	49.5 %	99 %
	Pr(crime against individual 3)	Pr(crime against individual 4)	Pr(crime against at least one person)
Defendant 2	50 %	1 %	51 %

Under decision rule (2.4), defendant 1 is convicted as long as defendant 2 is convicted. Under decision rule (2.5), defendant 1 is acquitted for sure as long as defendant 2 is acquitted. This is the case even when defendant 1 is *almost surely* guilty of at least one crime, and the probability with which defendant 2 is guilty is low.

Such a lack-of-commitment problem is the main difference between judicial and nonjudicial conviction decisions. This has been recognized by Schauer and Zeckhauser (1996) that *...although sound reasons for the criminal law's refusal to cumulate multiple low-probability accusations exist, the reasons for such refusal are often inapt in other settings. Taking adverse decisions based on cumulating multiple low-probability charges is often justifiable both morally and mathematically.* In practice, firms and organizations succumb to social pressure and aggregate the probabilities across different offenses. For example, firms face more social pressure to fire a manager whose probabilities of abusing subordinates are given by the first row, political parties have incentives to ostracize party members with bad reputations (e.g., individuals who are believed to have committed at least some offenses with high probability) in order to restore their popularity.

Motivated by this lack-of-commitment problem in non-judicial decision making, we explore alternative remedies to reduce the probability of crime when conviction decisions are made according to (2.4). Proposition 3 suggests that reducing the magnitude of punishment can help to deter crime since it induces a *positive correlation* in agents' private observations. Under such a positive correlation, agents' coordination motives encourage them to report when they have witnessed crime and vice versa, making their reports more informative.

Proposition 3. *When conviction decisions are made according to (2.4). For every $c > 0$, there exists $[\underline{L}, \overline{L}] \subset \mathbb{R}_+$ such that $q(1,1) + q(0,0) - q(1,0) - q(0,1) < 0$ in every equilibrium when $L \in [\underline{L}, \overline{L}]$.*

The proof is in Online Appendix B. Proposition 3 and Lemma 3.2 together imply that the opportunistic principal's decisions to commit different crimes are *strategic complements*. Given that the equilibrium probability of crime is interior, the opportunistic principal is indifferent between committing no crime and committing two crimes. This generates an *endogenous positive correlation* among agents' private observations of crime. Their coordination motives, induced by the loss from social stigma or retaliation c, encourage them to file accusations when they have witnessed offenses and discourage them from reporting when they have not witnessed any crime. Compared to the case with negatively correlated private information, the reporting cost c improves the informativeness of reports and decreases the equilibrium probability of crime.

Proposition 3 suggests that when the judge aggregates the probabilities of different crimes, the optimal punishment that minimizes $\Pr(\overline{\theta} = 1)$ is *interior*. This finding contrasts to Becker's (1968) seminal analysis of criminal justice and law enforcement, which suggests that increasing the magnitude of punishment helps reducing crime. From this perspective, our finding provides a novel rationale for being lenient to convicted defendants. Our logic applies to settings in which smoking-gun evidence is scarce and the potential victims' claims are hard to verify.

Notably, reducing L comes at the cost of increasing the number of crimes conditional on the principal being guilty. In particular, conditional on the opportunistic principal commits crime, two agents are victimized instead of only one. From this perspective, Proposition 3 unveils a tradeoff between reducing the probability of crime and the number of victims conditional on crime taking place.[23]

Implementing the solution proposed by Proposition 3 is challenging in practice since it requires a careful calibration of L. This is especially the case when the potential offenders' benefits from committing crime are unknown and L is interpreted as the magnitude of punishment *relative to* the benefit from committing crime.

[23] Note that the *expected* number of crimes is smaller when L is intermediate compared to the case in which L is large.

4 Heterogenous Propensity to Commit Crimes

This section extends Theorems 1 and 2 to cases in which with positive probability, the principal is a *virtuous-type* who does not benefit from committing crime. By incorporating the principal's heterogenous propensity to commit crimes, our analysis speaks to the concern that a key advantage of aggregating probabilities is to aggregate information about the potential offender's *type*.

We show that under both decision rules, the opportunistic principal *commits multiple crimes with positive probability*. Nevertheless, agents' private observations of crime are *negatively correlated* under decision rule (2.4), and are *uncorrelated* under decision rule (2.5). Our predictions on the informativeness of witness testimonies and on the effectiveness of crime deterrence remain unchanged. Recall that $\pi^o \in (0,1]$ is the probability of the opportunistic-type principal. Theorem 1' generalizes Theorem 1:

Theorem 1'. *When $n = 2$ and the judge uses decision rule (2.4), there exists $\overline{L}$ such that when $L > \overline{L}$,*

1. ***Endogenous Negative Correlation Between Crimes:*** *In every equilibrium*

$$\Pr(\theta_1 = 1|\theta_2 = 1) < \Pr(\theta_1 = 1|\theta_2 = 0) \text{ and } \Pr(\theta_2 = 1|\theta_1 = 1) < \Pr(\theta_2 = 1|\theta_1 = 0).$$

For every $\varepsilon > 0$, there exists $\overline{L}_\varepsilon \in \mathbb{R}_+$, such that in every equilibrium when $L > \overline{L}_\varepsilon$,

2. ***Low Informativeness of Reports:***

$$\max_{\boldsymbol{a} \in \{0,1\}^2} \mathcal{I}(\boldsymbol{a}) < \left\{ \frac{\pi^*}{1-\pi^*} \Big/ \frac{\min\{\pi^*, \pi^o\}}{1-\min\{\pi^*, \pi^o\}} \right\} + \varepsilon. \tag{4.1}$$

3. ***Ineffective Deterrence:***

$$\Pr(\overline{\theta} = 1) > \min\{\pi^*, \pi^o\} - \varepsilon. \tag{4.2}$$

Theorem 1' suggests that the endogenous negative correlation in agents' private observations of crime and their coordination motives remain robust when there is unobserved heterogeneity in the principal's propensity to commit crimes. Theorem 1 and Theorem 1' differ in terms of the upper bound on reporting informativeness and on the lower bound for the probability of crime:

- When the fraction of opportunistic-type principal π^o exceeds π^*, the equilibrium probability of crime is close to the conviction cutoff π^*. The informativeness of report about crime, measured by $\max_{\boldsymbol{a} \in \{0,1\}^2} \mathcal{I}(\boldsymbol{a})$ is arbitrarily close to 1 as the punishment from conviction increases.

- When the fraction of opportunistic-type principal π^o is less than π^*, Theorem 1' suggests that the prior probability of crime equals π^o, which is the maximal (ex ante) probability of crime given that the virtuous-type principal never commits any crime (Lemma 2.1).

  Given that in every equilibrium, there exists $\boldsymbol{a} \in \{0,1\}^2$ such that the principal is convicted with positive probability under $\boldsymbol{a}$, the judge's posterior belief about crime is no less than π^* after observing $\boldsymbol{a}$. This suggests a lower bound on the informativeness of agents' reports:

$$\max_{\boldsymbol{a}\in\{0,1\}^2} \mathcal{I}(\boldsymbol{a}) \geq \mathcal{I}_{min} \equiv \frac{\pi^*}{1-\pi^*} \Big/ \frac{\pi^o}{1-\pi^o}. \tag{4.3}$$

  The significance of Theorem 1' is to show that the lower bound $\mathcal{I}_{min}$ is attained in *all equilibria*.

The proof of Theorem 1' resembles that of Theorem 1 except for one modification. To begin with, the conclusions of Lemma 3.1 and Lemma 3.2 extend, which imply that the opportunistic-type principal *cannot* be indifferent between committing no crime and committing two crimes. However, whether he is indifferent between committing no crime and one crime, or indifferent between committing one crime and two crimes depends on the prior probability of the virtuous type, leading to the two different cases in Theorem 1':

1. When $\pi^o \geq \pi^*$, the opportunistic-type principal is indifferent between committing no crime and committing one crime, and the equilibrium outcome resembles that in Theorem 1.

2. When $\pi^o < \pi^*$ and L is large enough, the opportunistic-type principal cannot commit no crime. Otherwise, the informativeness of reports is close to 1 when L is large enough, and the posterior probability of crime is strictly below π^* even when the judge observing both agents' filing reports. This implies that the principal is never convicted, which violates the monotonicity refinement (Refinement 2). This suggests that the opportunistic-type principal is indifferent between committing one crime and two crimes.

   Nevertheless, θ_1 and θ_2 remain *negatively correlated*. To understand why, suppose by way of contradiction that θ_1 and θ_2 are independent or positively correlated. Then $Q_{0,j} \leq Q_{1,j}$, and the expressions for the reporting cutoffs (3.11) and (3.12) suggest that $\omega_j^* - \omega_j^{**} \geq b$. According to Lemma 3.1 that $q(1,1)$ vanishes to 0 as $L \to \infty$, both reporting cutoffs go to minus infinity. Similar to the single-agent benchmark, the informativeness of report then converges to infinity. Given that the opportunistic-type principal commits crime for sure, i.e., the prior probability of crime is π^o, the posterior probability of crime after observing two reports will exceed π^*. This contradicts the conclusion of Lemma 3.1 that $q(1,1) \in (0,1)$, which implies that the posterior belief about crime reaches π^* after the judge observing two reports.

The next result generalizes Theorem 2 which studies the equilibrium outcomes under DPP. The logic and proof are similar to those of Theorem 2, which we omit to avoid repetition.

Theorem 2'. *When $n = 2$ and the decision rule is (2.5), there exists $\overline{L} \in \mathbb{R}_+$ such that when $L > \overline{L}$,*

1. ***Uncorrelated Crimes:*** *, $\Pr(\theta_i = 1|\theta_j = 1) = \Pr(\theta_i = 1|\theta_j = 0)$ for every $i \neq j$.*

2. ***Linear Conviction Probability:*** *$\Pr(\theta_i = 1|a_i = 1) = \pi^*$ for every $i \in \{1, 2\}$, and the conviction probability is linear in the number of reports.*

In the limit where $\lim_{L\to\infty} \lim_{\delta\to 1}$,

3. ***Effective Deterrence:*** *The equilibrium probability of crime $\Pr(\overline{\theta} = 1)$ converges to 0.*

4. ***Highly Informative Reports:*** *For every $i \in \{1, 2\}$, the informativeness of agent i's report about θ_i, measured by $\frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)}$, goes to ∞.*

5 Three or More Agents

We extend Theorems 1 and 2 to settings with three or more agents, and establish a comparative static result on the number of agents that generalizes Proposition 2.

To start with, we generalize the measure of informativeness in Theorem 1 to three or more agents. Recall the following implication of Bayes Rule:

$$\frac{\Pr(\boldsymbol{a}|\overline{\theta} = 1)}{\Pr(\boldsymbol{a}|\overline{\theta} = 0)} \cdot \frac{\Pr(\overline{\theta} = 1)}{1 - \Pr(\overline{\theta} = 1)} = \frac{\Pr(\overline{\theta} = 1|\boldsymbol{a})}{1 - \Pr(\overline{\theta} = 1|\boldsymbol{a})} \text{ for every } \boldsymbol{a} \in \{0, 1\}^n. \tag{5.1}$$

That is to say,

$$\mathcal{I}(\boldsymbol{a}) \equiv \frac{\Pr(\boldsymbol{a}|\overline{\theta} = 1)}{\Pr(\boldsymbol{a}|\overline{\theta} = 0)} \tag{5.2}$$

measures the change in a judge's posterior belief after observing report profile $\boldsymbol{a}$, which we use to measure the informativeness of $\boldsymbol{a}$. We say that an equilibrium is *symmetric* if the principal's equilibrium strategy treats different agents symmetrically, and all agents' equilibrium strategies are the same. Proposition 4 generalizes the insights of Theorem 1 to settings with three or more agents:

Proposition 4. *Suppose the judge uses decision rule (2.4). For every $n \geq 2$ and $\varepsilon > 0$, there exists $\overline{L}_{n,\varepsilon} > 0$ such that for every $L > \overline{L}_{n,\varepsilon}$ and $\delta \in (0, 1)$,*

1. *There exists a symmetric equilibrium that survives Refinement 1.*[24]

[24] We show in Online Appendix C.4 that every symmetric Bayes Nash Equilibrium that survives Refinement 1 is a proper equilibrium that survives Refinement 2.

2. *In every symmetric equilibrium that survives Refinement 1,*

(a) $\Pr(\max_{j\neq i}\theta_j = 1|\theta_i = 1) < \Pr(\max_{j\neq i}\theta_j = 1|\theta_i = 0)$ *for every* $i \in \{1, 2, ..., n\}$*;*

(b) $\max_{\boldsymbol{a}\in\{0,1\}^n}\mathcal{I}(\boldsymbol{a}) < 1 + \varepsilon$ *and* $\Pr(\overline{\theta} = 1) > \pi^* - \varepsilon$*.*

According to statement 2(a), when an agent has witnessed an offense, his posterior belief attaches lower probability to other agents having witnessed offenses, compared to his posterior belief when he has not witnessed any offense. Furthermore, we show that in every symmetric equilibrium, the opportunistic-type principal either commits no crime or commits only one crime, and therefore, the above statement suggests that agents' private observations of crime remain *negatively correlated*. According to statement 2(b), agents' reports become arbitrarily uninformative and the total probability of crime converges to π^* as the punishment of conviction becomes large relative to the benefit from committing crime. This is similar to the prediction in Theorem 1.

To distinguish the mechanism behind our result and the one behind inefficient public good provision, we state a comparative statics result on the number of agents, with proof in Appendix E:

Theorem 3. *Suppose the judge uses decision rule (2.4). For every* $k, n \in \mathbb{N}$ *with* $k > n$*, there exists* $\overline{L} > 0$ *such that for every* $L > \overline{L}$*, compare any symmetric equilibrium when there are* k *agents with any symmetric equilibrium when there are* n *agents:*

1. ***Lower Informativeness:*** *The equilibrium value of* $\max_{\boldsymbol{a}\in\{0,1\}^k}\mathcal{I}(\boldsymbol{a})$ *is strictly lower than the equilibrium value of* $\max_{\boldsymbol{a}\in\{0,1\}^n}\mathcal{I}(\boldsymbol{a})$*.*

2. ***Higher Probability of Crime:*** *The equilibrium probability of crime* $\Pr(\overline{\theta} = 1)$ *is strictly higher in an environment with* k *agents compared to that with* n *agents.*

3. ***Higher Probability of Report:*** *Each agent's reporting cutoffs* (ω^*, ω^{**}) *are both strictly higher in an environment with* k *agents compared to that with* n *agents.*

The proof is in Appendix E. Theorem 3 shows that as the number of potential victims increases, the informativeness of reports decreases, and the probability of crime increases. Interestingly, each agent is *more likely* to accuse the principal when there are more potential victims, regardless of whether he has witnessed a crime. The last statement distinguishes our result from those on public good provision (e.g., Chamberlin 1974), where inefficiencies arise due to agents' incentives to free ride on others' contributions. In our model, each agent reports with higher probability when there are more agents, and the high chances of crime are driven by the lack-of-informativeness of these reports, instead of the lack-of reports.

Next, we study the game's equilibrium outcomes when the judge uses decision rule (2.5). Proposition 5 establishes the existence of equilibria in which agents' private observations of crimes are uncorrelated, and the conviction probability is a linear function of the number of reports. Such equilibria lead to arbitrarily informative report and a vanishing probability of crime as the punishment of conviction becomes large.

Proposition 5. *When the judge uses decision rule (2.5), there exists $\overline{L} > 0$, such that for every $L > \overline{L}$, there exist equilibria in which:*

1. θ_i *and* θ_j *are uncorrelated for every* $i \neq j$,
2. *the probability with which the principal is convicted is an affine function of the number of reports.*

In the limit where $\lim_{L\to\infty} \lim_{\delta\to 1}$,

3. *The equilibrium probability of crime* $\Pr(\overline{\theta} = 1)$ *converges to* 0.
4. *For every* $i \in \{1, 2\}$, *the informativeness of agent* i*'s report about* θ_i, $\frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)}$, *goes to* ∞.

6 Extensions & Robustness

We discuss several variants of our baseline model, which include the arrival of ex post evidence that exposes false accusations, the agents and the judge facing uncertainty about the number of potential victims, alternative payoff functions for the principal and agents, and the mechanical types' reports being informative about whether crimes have taken place. The details of the analysis can be found in supplementary appendix C.

Decreasing Marginal Benefits from Crime: Our result remains robust when the principal faces decreasing marginal returns from committing multiple crimes (Becker 1968) or receives a punishment larger than L when he is believed to have committed multiple crimes. These changes motivate the principal to commit fewer crimes and induce, as in the baseline model, negative correlation in the agents' private information. As in the baseline model, the agents' coordination motives undermine the informativeness of their reports and increase the probability of crime. In supplementary appendix C.3, we study an extension of the baseline model that formalizes these arguments. In particular, the principal is convicted of a minor crime and receives punishment L if the probability with which he is guilty of at least one crime exceeds π^*, and is convicted of a felony and receives punishment $L'(> L)$ if the probability with which he is guilty of two crimes exceeds some other cutoff $\pi^{**} \in (0, 1)$. When L is large (e.g., the principal can lose his lucrative position even when he is convicted with a minor offense), the principal commits at most one crime in every symmetric equilibrium that satisfies Refinement 1.

Agent's Reporting Cost: In terms of agents' costs of filing reports, our results remain valid as long as each agent's reporting cost c is strictly higher when the principal is acquitted compared to the case in which the principal is convicted. Therefore, our results are robust when each agent faces strictly positive reporting cost even when the principal is acquitted. Our results also extend when each agent's suffers a strictly lower loss from retaliation when there are more accusations against the principal, as long as the loss from retaliation is strictly positive whenever the principal is acquitted. This variation strengthens the coordination motives among agents without affecting the negative correlation between their private information.

Agent's Interdependent Preferences: As mentioned in section 2, agents may directly care about crimes committed against or witnessed by other agents. Recall that in general, strategic agent i's payoff is normalized to 0 when the principal is convicted, and is

$$\omega_i - b\Big((1-\gamma)\theta_i + \gamma f(\theta_i, \theta_{-i})\Big) - ca_i \tag{6.1}$$

when the principal is acquitted, with $f(\theta_i, \theta_{-i})$ an increasing function of both arguments. The term $\gamma f(\theta_i, \theta_{-i})$ captures the agent's altruism, namely, his payoff directly depends on crimes witnessed by or committed against other agents.

We show in supplementary appendix C.1 that when L is large, agents' reports become arbitrarily uninformative and the equilibrium probability of crime approaches π^*. In fact, introducing social preferences in this way undermines even further the informativeness of agents' reports. The reason is that when agent i's payoff depends directly on θ_j, agent i's report a_i becomes more responsive to his belief about θ_j. This, given the negative correlation between θ_i and θ_j, causes i's report to become even less responsive to θ_i and reduces the report's informativeness even further relative to the baseline model.

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

$$\frac{cl^*}{cl^* - (l^*+2)d}. \tag{6.2}$$

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

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

Uncertainty about the Number of Potential Victims: In applications such as workplace bullying, physical assaults and discrimination, the number of potential victims is usually not observed by the judge and the victims. We consider an extension of our model, in which nature randomly selects a subset $\widetilde{N}$ of $\{1, 2, ..., n\}$, interpreted as the set of agents against whom the principal has opportunities to commit crimes (or, depending on the interpretation, who may have an opportunity of witnessing some crime committed by the principal). We assume that only agents in $\widetilde{N}$ can accuse, which is interpreted as follows: if an agent outside of $\widetilde{N}$ filed an accusation, this accusation would be easily refuted by the defendant (e.g., by using an alibi). Only the principal observes $\widetilde{N}$. Agent i privately observes whether $i \in \widetilde{N}$ or not in addition to his private information in the baseline model.

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

Behavioral-Type Agents' Strategies: Suppose, in contrast to the baseline model, that the behavioral-type agent accuses the principal with probability $\overline{\alpha}$ when he has observed a crime, and accuses the principal with probability $\underline{\alpha}$ otherwise, with $1 > \overline{\alpha} \geq \underline{\alpha} > 0$. In addition to describing alternative behavioral types, this formulation can represent strategic agents who are immune to the principal's retaliation: Without retaliation, a strategic agent maximizes the expected value of $(\omega_i - b\theta_i)(1 - s)$. In equilibrium, the agent's reporting cutoffs are b and 0, depending on whether the agent has witnessed a crime or not. The conditional probabilities with which the agent accuses the principal are $\overline{\alpha} = \Phi(b)$ and $\underline{\alpha} = \Phi(0)$, respectively. Therefore, such a strategic agent behaves *as though* he were a behavioral type playing an informative cutoff-strategy.

7 Concluding Remarks

We put the key assumptions in our analysis in a broader perspective.

Contribution to the Legal Scholarship: The separation of charges against an individual and the use of distinct thresholds for each accusation may seem a priori arbitrary, unfair, and ineffective. This observation has been formalized and explored in legal scholarship and takes on a particular weight for individuals facing multiple accusations, each of which is hard to establish with the sufficient degree of certainty.

The premise, common in the legal literature, that offenses are exogenously and independently distributed is particularly problematic when analyzing the aggregation of offense probabilities. This aggregation creates an incentive for defendants to strategically restrict the number of offenses, which introduces *negative correlation* in the occurrence of offences and violates the premise that offenses are independently distributed. Indeed, our analysis shows that aggregating offense probabilities has severe drawbacks once the incentives of potential offenders and accusers are taken into account.

Underlying this difficulty, the *Aggregate Probabilities Principle* (APP) describes a *rule of punishment* rather than a *social objective function*. While it may be socially desirable to punish a defendant who is sufficiently likely of committing any offense, even unspecified, the principle may be self-defeating and suboptimal from a social welfare perspective once incentives are taken into account.

Equilibrium Analysis *vs.* Nonequilibrium Adjustments: Our results are derived from an equilibrium analysis, which presumes that players have correct expectations about the payoff consequences of their actions and other players' strategies.[25] When social rules change, as in a sudden crackdown on a specific type of crime, 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.[26]

Tradeoffs in Designing Conviction Rules: We identify a tradeoff between deterrence and fairness: while the probability of crime decreases when the standard of proof becomes more inclusive (π^* decreases), the probability that a convicted defendant is innocent increases. This tradeoff is easily quantified when there are multiple potential victims and the punishment in case of conviction is large, because the probability that a convicted principal is guilty

[25] Foundations of Nash equilibrium based on players learning one another's strategies have a long history in economics. See, for example, the textbook of Fudenberg and Levine (1995).

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

is approximately equal to π^* in equilibrium, while the fraction of innocent people among those that are convicted is approximately $1 - \pi^*$.[27]

However, Proposition 3 shows that using a more lenient sentence in case of conviction can be effective in deterring crimes, without increasing the probability of convicting the innocent. Thus, the tradeoff between deterrence and fairness depends on which instrument is considered: the conviction threshold or the sentence.

Suppose that a judge could *commit* to convict the principal if at least one report is filed. This would eliminate agents' coordination motive and, when the punishment L is large, would give the principal a strict incentive not to commit crime. To fulfill his commitment, however, the judge must convict the principal after receiving any accusation from the agents, even when the principal never commits crime in equilibrium. This leads to an undesirable outcome since *all convicted individuals are innocent* and the probability of convicting the innocent is significant.[28]

Shielding Accusers from Stigma through Secret Accusations: To address the potential pressure that is sometimes experienced by lone accusers, institutions have been developed under which reports are submitted to a third party and are only released when enough of them have been filed.[29]

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

In some cases, accusations may be leaked to the principal because of corruption, imperfect institutions, and other reasons. This risk is especially high if the principal is powerful and well-connected. Our results apply when such leakages occur with strictly positive probability: the acquitted principal can retaliate against the reporting agents once the information is leaked. Indeed, the results do not require any condition on the magnitude of c. They require only that (i) an agent's expected loss from retaliation be strictly positive, and (ii) the loss from retaliation be strictly larger when the principal is acquitted compared to the case in which the principal is convicted.

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

[27]This tradeoff has been discussed in reduced-form by Harris (1970) and Miceli (1991), but it does not arise in Becker (1968) and Landes (1970), who ignore wrongful convictions. In Kaplow (2011), punishments are expressed in terms of fines, i.e., zero sum transfers that do not affect the social surplus.

[28]This paradox induced by commitment arises in plea bargaining models, in which agents who reject pleas but are convicted at trial are known to be innocent (Grossman and Katz 1983, Reinganum 1988, and Siegel and Strulovici 2020).

[29]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.

reports made against him. Second, an individual agent has an incentive to coordinate with other agents whenever he is unsure about whether his report is pivotal or not. In a dynamic setting, this incentive can materialize after a *cold start* (i.e., where very few people have reported before and no agent wants to be the first accuser). It can also occur when an agent has observed many reports and is unsure of the number of reports needed to convict the principal (for example, if he faces uncertainty about the conviction standard π^* used by the judge). The inefficiencies and lack of credibility caused by the agents' coordination motives thus still arise in a dynamic environment.[30]

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

A Proof of Lemma 2.1

Statement 1: Given (ω_i, θ_i), strategic-type agent i prefers $a_i = 1$ if and only if:

$$\omega_i \sum_{\mathbf{a}_{-i} \in \{0,1\}^{n-1}} \sigma_{-i}(\mathbf{a}_{-i})\Big(q(1, \mathbf{a}_{-i}) - q(0, \mathbf{a}_{-i})\Big) \leq$$

$$b\theta_i \sum_{\mathbf{a}_{-i} \in \{0,1\}^{n-1}} \sigma_{-i}(\mathbf{a}_{-i})\Big(q(1, \mathbf{a}_{-i}) - q(0, \mathbf{a}_{-i})\Big) - c \sum_{\mathbf{a}_{-i} \in \{0,1\}^{n-1}} \sigma_{-i}(\mathbf{a}_{-i})\Big(1 - q(1, \mathbf{a}_{-i})\Big). \tag{A.1}$$

Refinement 2 implies that $q(1, \mathbf{a}_{-i}) - q(0, \mathbf{a}_{-i}) \geq 0$ for every $\mathbf{a}_{-i} \in \{0,1\}^{n-1}$, and the inequality is strict for some $\mathbf{a}_{-i}$. The existence of behavioral-type agent suggests that every $\mathbf{a}_{-i} \in \{0,1\}^{n-1}$ occurs with strictly positive probability. Therefore,

$$\sum_{a_{-i} \in \{0,1\}^{n-1}} \sigma_{-i}(\mathbf{a}_{-i})\Big(q(1, \mathbf{a}_{-i}) - q(0, \mathbf{a}_{-i})\Big) > 0.$$

According to inequality (A.1), agent i reports if and only if ω_i is below a cutoff, and the cutoff is strictly higher when $\theta_i = 1$ compared to when $\theta_i = 0$.

Statement 2: Suppose toward a contradiction that $\Pr(\overline{\theta} = 1) = 0$, then $\theta_1 = ... = \theta_n = 0$ with probability 1. Since every $\boldsymbol{a}$ occurs with strictly positive probability, we know that $\Pr(\overline{\theta} = 1 | \boldsymbol{a}) = 0$ for every $\boldsymbol{a} \in \{0,1\}^n$. Since $\pi^* \in (0,1)$, then the principal is convicted with probability 0 regardless of $\boldsymbol{a}$, and the opportunistic-type principal has a strict incentive to commit crime. This contradicts the presumption that $\Pr(\overline{\theta} = 1) = 0$.

Statement 3: Since $\omega_i^* > \omega_i^{**}$ for every $i \in \{1, 2, ..., n\}$ and every $\boldsymbol{a} \in \{0,1\}^n$ occurs with positive probability, Refinement 2 implies that for every $\boldsymbol{\theta} \succ \boldsymbol{\theta'}$, the probability with which the principal is convicted is strictly higher under $\boldsymbol{\theta}$ compared to $\boldsymbol{\theta'}$. Since the virtuous-type principal's benefit from committing crime is 0, he strictly prefers to commit no crime in any proper equilibrium that survives Refinements 1 and 2.

B Proof of Proposition 1

For notation simplicity, we replace $q(1)$ with q. The principal's expected cost of committing a crime is

$$\delta L q \Big(\Phi(b - c\frac{1-q}{q}) - \Phi(-c\frac{1-q}{q})\Big), \tag{B.1}$$

which is a continuous function of q. For every $\underline{q} \in (0,1)$, the above expression is bounded away from 0 if $q \geq \underline{q}$, and converges to 0 as $q \to 0$. Since the equilibrium probability of crime is strictly positive (statement 2 of Lemma

2.1), the value of (B.1) is less than or equal to 1.

Suppose toward a contradiction that for every $\overline{L} > 0$, there exists $L > \overline{L}$ under which in some proper equilibrium that survives Refinements 1 and 2, the value of (B.1) is strictly less than 1. Then the opportunistic principal has a strict incentive to commit crime i.e., $\Pr(\theta = 1) = \pi^o$, and moreover, q vanishes to 0 and $\Phi(\omega^*)/\Phi(\omega^{**})$ converge to $+\infty$ as $\overline{L} \to \infty$. This implies the existence of $\overline{L} > 0$ such that for every $L > \overline{L}$, and for every q such that (B.1) is strictly less than 1, $\Pr(\theta = 1|a = 1)$ is strictly greater than π^*. According to the judge's conviction rule, $q = 1$, which contradicts the presumption that q vanishes to 0. This implies that (B.1) equals 1 when L is large enough. It also implies that q is interior, and therefore, $\Pr(\theta = 1|a = 1) = \pi^*$.

When $L \to \infty$ while holding c constant, $1/\delta L$ converges to 0. Suppose toward a contradiction that q_s converges to some strictly positive number $\underline{q}$ along some sequence $\{L_n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} L_n = \infty$. Then ω^* and ω^{**} converge to $b - c(1-\underline{q})/\underline{q}$ and $-c(1-\underline{q})/\underline{q}$, respectively. The LHS of (3.5) converges to

$$\delta\underline{q}\Big(\Phi(b - c\frac{1-\underline{q}}{\underline{q}}) - \Phi(-c\frac{1-\underline{q}}{\underline{q}})\Big) \tag{B.2}$$

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

$$\lim_{\omega^*\to-\infty}\lim_{\delta\to 1}\frac{\delta\Phi(\omega^*) + (1-\delta)\alpha}{\delta\Phi(\omega^* - b) + (1-\delta)\alpha} = \infty. \tag{B.3}$$

The above equation makes use of the observation that $\lim_{\omega\to-\infty}\Phi(\omega)/\Phi(\omega - b) \to \infty$ for every $b > 0$. According to (3.3) and given that $\Pr(\boldsymbol{\theta} = 1|\boldsymbol{a} = 1) = \pi^*$, we have $\Pr(\boldsymbol{\theta} = 1)$ converging to 0.

C Proof of Theorem 1 and 1'

The following lemma establishes the symmetry of all equilibria:

Lemma C.1. *When $n = 2$ and the judge uses conviction rule (2.4), there exists $\overline{L} > 0$ such that when $L > \overline{L}$, the probability with which $(\theta_1, \theta_2) = (1, 0)$ equals the probability with which $(\theta_1, \theta_2) = (0, 1)$, and moveover, $(\omega_1^*, \omega_1^{**}) = (\omega_2^*, \omega_2^{**})$.*

The proof of Lemma C.1 is in Supplementary Appendix A and the proof of Lemma 3.1 is in Online Appendix A. In what follows, we show Theorems 1 and 1' taking the conclusions of Lemmas 3.1 and C.1 as given. We consider two cases separately, depending on the comparison between π^o and π^*.

C.1 Case 1: $\pi^o \geq \pi^*$

Lemma 3.1 implies that $q(1,1) + q(0,0) - q(1,0) - q(0,1) > 0$ and $\Pr(\overline{\theta} = 1|\boldsymbol{a}) \leq \pi^*$ for every $\boldsymbol{a} \in \{0,1\}^2$. Therefore, $\Pr(\overline{\theta} = 1) < \pi^* \leq \pi^o$, which suggests that the opportunistic-type principal chooses $\boldsymbol{\theta} = (0,0)$ with positive probability. According to Lemma 3.2, the principal's actions are strategic substitutes, and therefore, he chooses $\boldsymbol{\theta} = (1,1)$ with zero probability. This implies statement 1.

Let π be the ex ante probability of crime. Lemma C.1 suggests that (θ_1, θ_2) equals $(1,0)$ and $(0,1)$ with equal probabilities. Conditional on $\theta_i = 0$, the probability of $\boldsymbol{\theta} = (0,0)$ is:

$$\beta \equiv \frac{1-\pi}{1-\pi/2}. \tag{C.1}$$

Based on the conclusion of Lemma C.1, let $Q_1 \equiv Q_{1,1} = Q_{1,2}$, and let $Q_0 \equiv Q_{0,1} = Q_{0,2}$. Since $\boldsymbol{\theta} = (1,1)$ occurs with zero probability,

$$Q_1 = \delta\Phi(\omega^{**}) + (1-\delta)\alpha \tag{C.2}$$

and

$$Q_0 = \delta\Big(\beta\Phi(\omega^{**}) + (1-\beta)\Phi(\omega^*)\Big) + (1-\delta)\alpha. \tag{C.3}$$

Subtracting (3.12) from (3.11), we have:

$$\omega^* - \omega^{**} = b - \frac{c}{q(1,1)} \cdot \frac{-1 + Q_0/Q_1}{Q_0}. \tag{C.4}$$

Lemma C.2. $\omega^* - \omega^{**} \in (0, b)$.

Proof of Lemma C.2: According to (C.2) and (C.3), $\omega^* - \omega^{**} > 0$ is equivalent to $Q_0 > Q_1$. To see this, suppose by way of contradiction that $Q_0 \leq Q_1$. Equation (C.4) implies that $\omega^* \geq \omega^{**} + b > \omega^{**}$. The comparison between (3.11) and (3.12) then yields $Q_0 > Q_1$, the desired contradiction. Since $Q_0 > Q_1$, the term $\frac{-1+Q_0/Q_1}{Q_0}$ is strictly positive. Therefore, $\omega^* - \omega^{**} < b$. □

Let

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

Since the opportunistic-type principal mixes between $(0,1)$, $(1,0)$, and $(0,0)$, we have:

$$\mathcal{I} \equiv \frac{\big(\delta\Phi(\omega^*) + (1-\delta)\alpha\big)\big(\delta\Phi(\omega^{**}) + (1-\delta)\alpha\big)}{\big(\delta\Phi(\omega^{**}) + (1-\delta)\alpha\big)^2} = \frac{\delta\Phi(\omega^*) + (1-\delta)\alpha}{\delta\Phi(\omega^{**}) + (1-\delta)\alpha}. \tag{C.5}$$

Since $q(1,1) \in (0,1)$, the judge's posterior assigns probability π^* to $\overline{\theta} = 1$ after observing $(a_1, a_2) = (1,1)$, which implies:

$$\frac{\pi}{1-\pi} = \frac{l^*}{\mathcal{I}}, \quad \text{in which} \quad l^* \equiv \frac{\pi^*}{1-\pi^*}. \tag{C.6}$$

Plugging (C.6) into (C.1), we have the following expressions for β and $1-\beta$:

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

Plugging (C.7) into (C.2) and (C.3), we obtain the following expression for the ratio between Q_0 and Q_1:

$$\frac{Q_0}{Q_1} = \beta + (1-\beta)\mathcal{I} = \frac{(l^*+2)\mathcal{I}}{l^* + 2\mathcal{I}}. \tag{C.8}$$

Plugging (3.11) and (3.12) into (C.8), we obtain

$$\frac{|\omega^* - c - b|}{|\omega^{**} - c|} = \frac{(l^*+2)\mathcal{I}}{l^* + 2\mathcal{I}}. \tag{C.9}$$

This leads to the following lemma:

Lemma C.3. *If* $\omega^* \to -\infty$*, then* $\mathcal{I} \to 1$ *and* $\pi \to \pi^*$.

Proof of Lemma C.3: Since $\omega^* - \omega^{**} \in (0, b)$, the difference between $|\omega^* - c - b|$ and $|\omega^{**} - c|$ is at most b. The LHS of (C.9) converges to 1 as $\omega^* \to -\infty$. Since the RHS of (C.9) is strictly increasing in $\mathcal{I}$, we know that the limiting value of $\mathcal{I}$ is 1. According to (C.6), the limiting value of π is π^*. □

Next, we show that $\omega^* \to -\infty$ as $L \to \infty$. Recall that the opportunistic type principal's indifference condition is:

$$(\delta L)^{-1} = q(1,1)\Big(\delta\Phi(\omega^{**}) + (1-\delta)\alpha\Big)\Big(\Phi(\omega^*) - \Phi(\omega^{**})\Big) \tag{C.10}$$

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

1. $L(n) \geq \overline{L}$ for every $n \in \mathbb{N}$, and $\lim_{n\to\infty} L(n) = \infty$;

2. $(\omega^*(n), \omega^{**}(n), q(n), \pi(n))$ is an equilibrium when $L = L(n)$, for every $n \in \mathbb{N}$;

3. $\lim_{n\to\infty} \omega^{**}(n) = \omega^{**}$ for some finite $\omega^{**} \in \mathbb{R}$.

Since $\delta\Phi(\omega^{**}(n)) + (1-\delta)\alpha$ is bounded away from 0, (C.10) implies:

1. *either* there exists a subsequence $\{k_n\}_{n=1}^{\infty} \subset \mathbb{N}$ such that: $\lim_{n\to\infty} q(k_n) = 0$.

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

First, suppose that $\lim_{n\to\infty} q(k_n) = 0$ for some subsequence $\{k_n\}_{n=1}^{\infty}$. Then (3.11) and (3.12) imply that both $\omega^*(k_n)$ and $\omega^{**}(k_n)$ converge to $-\infty$. This contradicts the third requirement that the sequence $\omega^{**}(n)$ converges to some finite number ω^{**}. Next, suppose that $\lim_{n\to\infty}(\omega^*(k_n) - \omega^{**}(k_n)) = 0$ for some subsequence $\{k_n\}_{n=1}^{\infty}$. Since $\omega^{**}(k_n)$ converges to an interior number, both $Q_0(k_n)$ and $Q_1(k_n)$ are bounded away from 0, which suggests that $Q_0(k_n)/Q_1(k_n)$ converges to 1 as $n \to \infty$. From the previous step, we know that there exists no subsequence of $\{k_n\}_{n=1}^{\infty}$ such that $q(k_n)$ converges to 0. That is to say, there exists $\eta > 0$ such that $q(k_n) \geq \eta$ for every $n \in \mathbb{N}$. Expression (C.4) then suggests that $\omega^*(k_n) - \omega^{**}(k_n)$ converges to b. This contradicts the hypothesis that $\lim_{n\to\infty} \omega^*(k_n) - \omega^{**}(k_n) = 0$.

C.2 Case 2: $\pi^o < \pi^*$

First, we show that the opportunistic-type principal *commits two crimes* with positive probability in every equilibrium. Suppose toward a contradiction that he commits two crimes with probability zero. According to Lemma 2.1, the virtuous-type never commits crime, and therefore, the equilibrium probability of crime is no more than π^o, which is strictly less than π^*. As a result,

$$\mathcal{I} \equiv \frac{\Pr(a_1 = a_2 = 1|\overline{\theta} = 1)}{\Pr(a_1 = a_2 = 1|\overline{\theta} = 0)} \geq \frac{\pi^o}{1-\pi^o} \Big/ \frac{\pi^*}{1-\pi^*} > 1. \tag{C.11}$$

The expressions for Q_0 and Q_1 in (C.2) and (C.3) also apply in this setting. The derivations in Appendix C.1 imply that for every $\varepsilon > 0$, there exists $\overline{L}_\varepsilon > 0$ such that when $L \geq L_\varepsilon$, the informativeness ratio is less than $1 + \varepsilon$. This contradicts (C.11), which requires the informativeness ratio to be strictly bounded away from 1.

Since $q(1,1) \in (0,1)$ and $q(1,1)+q(0,0)-q(1,0)-q(0,1) > 0$, Lemma 3.2 implies that the opportunistic-type principal cannot be indifferent between committing zero crime and two crimes. Lemma C.1 implies that he chooses $\boldsymbol{\theta} = (0,1), (1,0)$ and $(1,1)$ with positive probability, and chooses $\boldsymbol{\theta} = (0,0)$ with zero probability. Therefore, $\Pr(\overline{\theta} = 1) = \pi^o$. The informativeness ratio is pinned down by Bayes Rule, namely,

$$\mathcal{I}\frac{\Pr(\overline{\theta} = 1)}{1 - \Pr(\overline{\theta} = 1)} = \frac{\Pr(\overline{\theta} = 1|a_1 = a_2 = 1)}{1 - \Pr(\overline{\theta} = 1|a_1 = a_2 = 1)}, \tag{C.12}$$

In the last step, we establish the negative correlation between θ_1 and θ_2 in the first statement. Suppose by way of contradiction that θ_1 and θ_2 are independent or positively correlated, Lemma C.1 suggests that $Q_1 \geq Q_0$. The expressions of reporting cutoffs (3.11) and (3.12) suggest that $\omega^* - \omega^{**} \geq b$. When L is large enough, we have $q(1,1) \to 0$ and $\omega^* \to \infty$, and the conclusion in the single-agent benchmark suggests that $\mathcal{I} \to \infty$. This contradicts (C.12) as well as the conclusion that $\Pr(\overline{\theta} = 1|a_1 = a_2 = 1) = \pi^*$ implied by Lemma 3.1.

D Proofs of Theorem 2 and 2'

Our proof uses a result shown in Online Appendix A.

Lemma D.1. *There exists $\overline{L} > 0$ such that when $L > \overline{L}$, $q(1,1) < 1$ in every equilibrium.*

Uncorrelated Crimes: First, suppose by way of contradiction that $\Pr(\theta_1 = 1|\theta_2 = 1) < \Pr(\theta_1 = 1|\theta_2 = 0)$. Since θ_1 and θ_2 are both binary, $\Pr(\theta_2 = 1|\theta_1 = 1) < \Pr(\theta_2 = 1|\theta_1 = 0)$. Since $\omega_i^* > \omega_i^{**}$ for $i \in \{1,2\}$, $\Pr(\theta_1 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_1 = 1|\boldsymbol{a} = (1,0))$ and $\Pr(\theta_2 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_2 = 1|\boldsymbol{a} = (0,1))$. Therefore,

$$\max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a} = (1,1)) < \max\Big\{\max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a} = (1,0)), \max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a} = (0,1))\Big\}. \tag{D.1}$$

Decision rule (2.5) implies that $\max\{q(0,1), q(1,0)\} \geq q(1,1)$. Refinement 2 requires that $q(1,1) \geq \max\{q(0,1), q(1,0)\}$. The two together imply that $q(1,1) = \max\{q(0,1), q(1,0)\}$. Refinement 1 requires that $q(0,0) = 0$ and Lemma 2.1 implies that $\max\{q(1,1), q(1,0), q(0,1), q(0,0)\} > 0$. Therefore, $q(1,1) = \max\{q(0,1), q(1,0)\} > 0$. Inequality (D.1) and decision rule (2.5) together rule out the possibility that $q(1,1) = \max\{q(0,1), q(1,0)\} \in (0,1)$. Therefore, $q(1,1) = \max\{q(0,1), q(1,0)\} = 1$, which contradicts Lemma D.1.

Next, suppose by way of contradiction that $\Pr(\theta_1 = 1|\theta_2 = 1) > \Pr(\theta_1 = 1|\theta_2 = 0)$. Since θ_1 and θ_2 are both binary, $\Pr(\theta_2 = 1|\theta_1 = 1) > \Pr(\theta_2 = 1|\theta_1 = 0)$. Since $\omega_i^* > \omega_i^{**}$ for $i \in \{1,2\}$, $\Pr(\theta_1 = 1|\boldsymbol{a} = (1,1)) > \Pr(\theta_1 = 1|\boldsymbol{a} = (1,0))$ and $\Pr(\theta_2 = 1|\boldsymbol{a} = (1,1)) > \Pr(\theta_2 = 1|\boldsymbol{a} = (0,1))$. Therefore,

$$\max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a} = (1,1)) > \max\Big\{\max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a} = (1,0)), \max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a} = (0,1))\Big\}. \tag{D.2}$$

According to Lemma D.1, we have $q(1,1) < 1$, which further suggests that $q(1,1) \in (0,1)$. According to decision rule (2.5), $\max_{i\in\{1,2\}} \Pr(\theta_i = 1|\boldsymbol{a} = (1,1)) = \pi^*$. As a result, the RHS of (D.2) is strictly less than π^*, which according to decision rule (2.5) leads to $q(0,1) = q(1,0) = 0$. Therefore, $q(1,1) + q(0,0) - q(1,0) - q(0,1) > 0$. Under such conviction probabilities, the opportunistic-type principal's incentives to commit crimes are strategic substitutes, which contradicts the presumption that $\Pr(\theta_1 = 1|\theta_2 = 1) > \Pr(\theta_1 = 1|\theta_2 = 0)$.

Linear Conviction Probabilities: First, we show that choosing $\boldsymbol{\theta} = (0,0)$ is weakly optimal for the opportunistic type principal. Suppose toward a contradiction that it is not, then Lemma 2.1 implies that $\Pr(\overline{\theta} = 1) = \pi^o$. The conclusion in the first part suggests that $\Pr(\theta_1 = 1) = 1 - \sqrt{1 - \pi^o}$, $|\omega_i^* - \omega_i^{**}| = b$ for every $i \in \{1,2\}$, and

$$\Pr(\theta_1 = 1|\boldsymbol{a} = (1,1)) = \Pr(\theta_1 = 1|\boldsymbol{a} = (1,0)) = \frac{\delta\Phi(\omega_1^*) + (1-\delta)\alpha}{\delta\Phi(\omega_1^{**}) + (1-\delta)\alpha} \Pr(\theta_1 = 1).$$

For any $\pi^o > 0$, RHS of the above equation is strictly greater than π^* when L is sufficiently large. Under decision rule (2.5), this implies that $q(1,1) = 1$, which contradicts Lemma D.1.

Next, given that $\Pr(\theta_1 = 1|\theta_2 = 1) = \Pr(\theta_1 = 1|\theta_2 = 0)$ and $\boldsymbol{\theta} = (0,0)$ is weakly optimal for the opportunistic type principal, we know that for every $\pi^o \in (0,1)$, the opportunistic type principal is indifferent between $\boldsymbol{\theta} \in \{(0,0),(0,1),(1,0),(1,1)\}$, i.e., his actions are neither strict complements nor strict substitutes. Lemma 3.2 suggests that $q(1,1) + q(0,0) = q(1,0) + q(0,1)$. Refinement 1 implies that $q(0,0) = 0$. Suppose toward a contradiction that $q(1,0) > q(0,1)$. Then the opportunistic type principal strictly prefers $\boldsymbol{\theta} = (0,1)$ to $\boldsymbol{\theta} = (1,0)$. This contradicts the previous conclusion that he is indifferent between $\boldsymbol{\theta} = (0,1)$ and $\boldsymbol{\theta} = (1,0)$.

Limiting Properties: Part 1 and Part 2 of our proof implies that $|\omega_i^* - \omega_i^{**}| = b$, and the opportunistic type principal is indifferent between $\boldsymbol{\theta} \in \{(0,0),(0,1),(1,0),(1,1)\}$. His indifference condition implies that $\omega_i^* \to -\infty$ in the limit where $\lim_{L\to\infty}\lim_{\delta\to 1}$. Similar to Proposition 1, the informativeness ratio of each agent's report converges to infinity and $\Pr(\theta_i = 1) \to 0$. The latter implies that $\Pr(\overline{\theta} = 1) \to 0$.

E Proofs of Theorem 3 and Proposition 2

Our proof uses the following lemma, which is shown in Online Appendix C.

Lemma E.1. *Fix any $n \geq 2$ and suppose the judge uses (2.4). There exists $\overline{L} > 0$ such that for every $L > \overline{L}$, in every symmetric Bayes Nash Equilibrium that survives Refinement 1,*

1. *the principal is convicted with positive probability only if $\boldsymbol{a} = (1,1,...,1)$;*
2. *the principal commits at most one crime.*

In an environment with n agents, we derive formulas for the agents' reporting cutoffs $(\omega_n^*, \omega_n^{**})$, the informativeness of reports $\mathcal{I}_n$ and the equilibrium probability of crime π_n. Agent i's reporting cutoff is

$$\omega_n^* = b + c - \frac{c}{q_n Q_{1,n}} \text{ when } \theta_i = 1, \tag{E.1}$$

and

$$\omega_n^{**} = c - \frac{c}{q_n Q_{0,n}} \text{ when } \theta_i = 0, \tag{E.2}$$

where

$$Q_{1,n} \equiv \Big(\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha\Big)^{n-1}, \tag{E.3}$$

$$Q_{0,n} \equiv \frac{n\mathcal{I}_n}{(n-1)l^* + n\mathcal{I}_n}\Big(\delta\Phi(\omega_n^{**})+(1-\delta)\alpha\Big)^{n-1}+\frac{(n-1)l^*}{(n-1)l^* + n\mathcal{I}_n}\Big(\delta\Phi(\omega_n^{**})+(1-\delta)\alpha\Big)^{n-2}\Big(\delta\Phi(\omega_n^{*})+(1-\delta)\alpha\Big). \tag{E.4}$$

In any symmetric equilibrium, the aggregate informativeness of reports, defined in (5.2), can be written as:

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

Since the judge is indifferent between convicting and acquitting the principal when there are n reports, we have

$$\mathcal{I}_n = \frac{\pi^*}{1-\pi^*}\Big/\frac{\pi_n}{1-\pi_n}. \tag{E.5}$$

When L is large enough, the principal is indifferent between committing crime against a single agent and committing no crime, which leads to the indifference condition

$$\frac{1}{\delta L} = q_n\Big(\Phi(\omega_n^*) - \Phi(\omega_n^{**})\Big)\Big(\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha\Big)^{n-1}. \tag{E.6}$$

Reporting Cutoffs & Distance Between Cutoffs: In this part, we show that $\omega_k^* > \omega_n^*$. Suppose toward a contradiction that $\omega_k^* \le \omega_n^*$. From (E.1), we have

$$q_k\Big(\delta\Phi(\omega_k^{**}) + (1-\delta)\alpha\Big)^{k-1} \le q_n\Big(\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha\Big)^{n-1}. \tag{E.7}$$

Therefore, $q_kQ_{1,k} \le q_nQ_{1,n}$ which is equivalent to

$$q_k\Big(\delta\Phi(\omega_k^{**}) + (1-\delta)\alpha\Big)^{k-1}\Big(\Phi(\omega_n^*) - \Phi(\omega_n^{**})\Big) \le q_n\Big(\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha\Big)^{n-1}\Big(\Phi(\omega_n^*) - \Phi(\omega_n^{**})\Big)$$

$$= q_k\Big(\delta\Phi(\omega_k^{**}) + (1-\delta)\alpha\Big)^{k-1}\Big(\Phi(\omega_k^*) - \Phi(\omega_k^{**})\Big).$$

This implies that

$$\Phi(\omega_n^*) - \Phi(\omega_n^{**}) \le \Phi(\omega_k^*) - \Phi(\omega_k^{**}). \tag{E.8}$$

Since $\omega_k^* \le \omega_n^*$, (E.8) is true only when

$$\omega_n^* - \omega_n^{**} \le \omega_k^* - \omega_k^{**}, \tag{E.9}$$

which in turn implies that $\omega_k^{**} \leq \omega_n^{**}$ and therefore $q_k Q_{0,k} \leq q_n Q_{0,n}$. Computing the two sides of (E.9) by subtracting (E.2) from (E.1), we have

$$\omega_n^* - \omega_n^{**} = b - \frac{c}{q_n}\frac{Q_{0,n} - Q_{1,n}}{Q_{1,n}Q_{0,n}} \text{ and } \omega_k^* - \omega_k^{**} = b - \frac{c}{q_k}\frac{Q_{0,k} - Q_{1,k}}{Q_{1,k}Q_{0,k}}.$$

Due to the previous conclusion that $q_k Q_{1,k} \leq q_n Q_{1,n}$ and $q_k Q_{0,k} \leq q_n Q_{0,n}$, (E.9) is true only when

$$q_n(Q_{0,n} - Q_{1,n}) \geq q_k(Q_{0,k} - Q_{1,k}). \tag{E.10}$$

Since

$$Q_{0,n} - Q_{1,n} = \frac{(n-1)l^*}{(n-1)l^* + n\mathcal{I}_n}\delta\Big(\Phi(\omega_n^*) - \Phi(\omega_n^{**})\Big)\Big(\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha\Big)^{n-2}$$

and the term

$$\delta\Big(\Phi(\omega_n^*) - \Phi(\omega_n^{**})\Big)\Big(\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha\Big)^{n-2} = L^{-1}q_n^{-1}\frac{1}{\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha}.$$

From (E.6), we know that (E.10) is equivalent to:

$$\frac{(n-1)l^*}{(n-1)l^*\Big(\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha\Big) + n\Big(\delta\Phi(\omega_n^*) + (1-\delta)\alpha\Big)}$$

$$\geq \frac{(k-1)l^*}{(k-1)l^*\Big(\delta\Phi(\omega_k^{**}) + (1-\delta)\alpha\Big) + k\Big(\delta\Phi(\omega_k^*) + (1-\delta)\alpha\Big)}$$

which in turn reduces to:

$$(n-1)(k-1)l^*\Big(\delta\Phi(\omega_k^{**}) + (1-\delta)\alpha\Big) + (n-1)k\Big(\delta\Phi(\omega_k^*) + (1-\delta)\alpha\Big)$$

$$\geq (n-1)(k-1)l^*\Big(\delta\Phi(\omega_n^{**}) + (1-\delta)\alpha\Big) + (k-1)n\Big(\delta\Phi(\omega_n^*) + (1-\delta)\alpha\Big)$$

The above inequality cannot be true since $\delta\Phi(\omega_k^{**}) + (1-\delta)\alpha < \delta\Phi(\omega_n^{**}) + (1-\delta)\alpha$, $\delta\Phi(\omega_k^*) + (1-\delta)\alpha < \delta\Phi(\omega_n^*) + (1-\delta)\alpha$ and $(n-1)k < (k-1)n$. The last inequality holds due to the assumption that $k > n$. This leads to a contradiction which shows that $\omega_k^* > \omega_n^*$ whenever $k > n$.

Notice that up until the last step, we did not use the fact that $k > n$. Given the previous conclusion that $\omega_k^* > \omega_n^*$ and repeat the same reasoning up until (E.9), we have:

$$\omega_n^* - \omega_n^{**} > \omega_k^* - \omega_k^{**}. \tag{E.11}$$

This together with $\omega_k^* > \omega_n^*$ implies that $\omega_k^{**} > \omega_n^{**}$.

Informativeness and Probability of Crime: We show that $\mathcal{I}_n > \mathcal{I}_k$. Since $q(1,1,...,1) = \pi^*$, this together with (5.1) lead to the comparative statics on the ex ante probability of crime.

Applying (E.1) and (E.2) to both n and k, we obtain the following expression for the ratios:

$$\frac{\omega_n^* - b - c}{\omega_k^* - b - c} = \frac{q_k Q_{1,k}}{q_n Q_{1,n}} \quad \text{and} \quad \frac{\omega_n^{**} - c}{\omega_k^{**} - c} = \frac{q_k Q_{1,k}(\beta_k + (1-\beta_k)\mathcal{I}_k)}{q_n Q_{1,n}(\beta_n + (1-\beta_n)\mathcal{I}_n)}. \tag{E.12}$$

First, we show that

$$\frac{\omega_n^* - b - c}{\omega_k^* - b - c} > \frac{\omega_n^{**} - c}{\omega_k^{**} - c}. \tag{E.13}$$

Suppose toward a contradiction that the RHS of (E.13) is at least as large as the LHS of (E.13), then:

$$\frac{\omega_n^{**} - c - (\omega_n^* - b - c)}{\omega_k^{**} - c - (\omega_k^* - b - c)} \geq \frac{\omega_n^* - b - c}{\omega_k^* - b - c}. \tag{E.14}$$

The RHS of (E.14) is strictly greater than 1 since $0 > \omega_k^* > \omega_n^*$ when L is large enough. The LHS of (E.14) being greater than 1 is equivalent to

$$b - (\omega_n^* - \omega_n^{**}) > b - (\omega_k^* - \omega_k^{**})$$

which contradicts the previous conclusion in (E.11). This establishes (E.13). This together with (E.12) imply that

$$\beta_k + (1-\beta_k)\mathcal{I}_k < \beta_n + (1-\beta_n)\mathcal{I}_n.$$

Plugging in the expressions of $\mathcal{I}_n$ and $\mathcal{I}_k$ in (E.5), we have:

$$\mathcal{I}_k\big(k + (k-1)l^*\big)\big(n\mathcal{I}_n + (n-1)l^*\big) < \mathcal{I}_n\big(n + (n-1)l^*\big)\big(k\mathcal{I}_k + (k-1)l^*\big).$$

Let $\Delta \equiv \mathcal{I}_k - \mathcal{I}_n$, the above inequality reduces to:

$$(k-n)\mathcal{I}_n(\mathcal{I}_k - 1) = (k-n)\mathcal{I}_n(\mathcal{I}_n + \Delta - 1) < k\Delta - \Big(l^*(k-1)(n-1) + nk\Big)\Delta.$$

Suppose toward a contradiction that $\Delta \geq 0$. Then, the LHS is strictly positive since $\mathcal{I}_k > 1$ and $k > n$. The RHS is negative since $l^*(k-1)(n-1) + nk > k$. This leads to the desired contradiction, and implies that $\Delta < 0$ and, hence, $\mathcal{I}_n > \mathcal{I}_k$. Since $\mathcal{I}_n > \mathcal{I}_k$, (5.1) implies that $\Pr(\overline{\theta} = 1)$ increases when the number of agents increases from n to k.

References

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

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

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

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

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

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

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

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

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

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

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

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

[13] 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.

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

[15] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[39] Robertson, Bernard and G. A. Vignaux (1993) "Probabilit–The Logic of the Law," *Oxford Journal of Legal Studies*, 13(4), 457-478.

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

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

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

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

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

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

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

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

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