When to Convict Defendants Facing Multiple Accusations? A Strategic Analysis*

Harry Pei† Bruno Strulovici‡

August 31, 2024

Abstract: Over the last few decades, prominent legal scholars have proposed to radically alter the rules of adjudication for defendants accused of multiple offenses. In such cases, they have argued that conviction should be based on the overall probability of guilt of the defendant rather than on the distinct probabilities of guilt for each possible offense. This proposal has gained prominence with the rising number of individuals accused of committing serial assaults. We show however that the proposal may undermine the quality of the information available to adjudicate cases and weaken deterrence once the incentives of potential offenders and witnesses are taken into account. When conviction entails a large punishment, a potential offender's decisions to commit distinct offenses are strategic substitutes, and witnesses' decisions to report offenses are strategic complements. Taken together, these strategic considerations tend to weaken the informativeness of witness reports and increase the probability of offenses in equilibrium. By contrast, convicting defendants based on the probabilities of distinct offenses leads to informative testimonies and a low probability of offenses.

Keywords: Strategic communication, coordination, correlated signals, aggregate probabilities principle.

JEL Codes: C72, D82, D83, K42.

*This paper was previously circulated under the title "Crime Aggregation, Deterrence, and Witness Credibility". We thank S. Nageeb Ali, Bocar Ba, Arjada Bardhi, Laura Doval, Mehmet Ekmekci, Andreas Kleiner, Anton Kolotinin, Frances Xu Lee, Andrew McClellan, Aniko Öry, Bobby Pakzad-Hurson, Alessandro Pavan, Daniel Rappoport, Larry Samuelson, Ron Siegel, Teck Yong Tan, and Boli Xu for helpful comments. We acknowledge financial support from NSF Grants SES-1151410, SES-1947021, and SES-2337566.

†Department of Economics, Northwestern University. Email: harrydp@northwestern.edu.

‡Department of Economics, Northwestern University. Email: b-strulovici@northwestern.edu.

1 Introduction

When a defendant faces multiple charges, the legal norm is to consider these charges separately and to convict the defendant for each individual charge for which the evidence of guilt meets the appropriate standard of proof. However, the desirability of this separation for deterrence and fairness is not 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, then the defendant is acquitted on both counts, even though the probability that he is guilty of *at least one* offense is $1 - 0.2 \times 0.2 = 0.96$. Under the same threshold, a defendant accused of a single offense is convicted when his probability of guilt is 0.91 and thus lower than the first defendant's.

This issue has received significant attention from prominent legal scholars, starting with Cohen (1977), Bar Hillel (1984), and Robertson and Vignaux (1993), who explore the possibility of aggregating accusations against a defendant by considering his overall probability of guilt. Harel and Porat (2009) formalize the *Aggregate Probabilities Principle* ("APP"), according to which a defendant is convicted if the probability that he has committed *at least one* offence exceeds a threshold. Compared to the standard legal norm, which they call the *Distinct Probabilities Principle* ("DPP"), Harel and Porat advocate the use of APP to varying degrees in both civil and criminal cases, arguing that APP can reduce adjudication errors, improve deterrence, and reduce the cost of enforcement.

This issue has gained even more momentum over the last decade in both the legal and the philosophy literatures (see for example, Pundik (2015), Littlejohn (2017), Smith (2018), Di Bello (2019), Backes (2020), Blome-Tilman (2020), Ross (2020)).[1] Schauer (2023) concludes that *"....the legal systems ought to be more willing than they now are to base liability on under-specified wrongs, at least when multiple under-specified acts are relevantly similar and are alleged to be committed by the same defendant"*.

The aforementioned papers rely on two key assumptions. First, the defendant's guilt for each offense is *independently distributed* across charges. Second, the informativeness of the evidence available regarding each offense is *exogenously given*. In particular, this informativeness is independent of the rule used to adjudicate cases with multiple accusations. These two assumptions ignore the impact that APP can have on (i) the *incentives* of potential offenders to commit offenses and (ii) the incentives of potential witnesses to report offenses. This impact may be significant when a defendant faces multiple accusations that are hard to prove beyond a reasonable doubt, such as abuses of power for which the main incriminating evidence consists of witness testimonies. In these cases, the incentives of witnesses to report their observations truthfully are affected by how likely their report is to be

[1]Much of the related work builds on thought experiments such as Cohen's (1977) Gatecrasher paradox, Pundik's (2015) serial burglar, and Littlejohn (2017)'s Prison Yard, to argue that adjudication rules based on DDP lead to counter-intuitive and problematic outcomes. Some of these thought experiments are closer to reality, such as Schauer's (2023) discussion of repeated assaults and driving infractions.

corroborated by other witnesses and their fear of retaliation.

This paper investigates the effects of APP and DPP on incentives, informativeness, and deterrence from a game-theoretic perspective.[2] Our model features one potential offender and multiple witnesses. The potential offender's decisions to commit offenses and the witnesses' decisions of whether to accuse the potential offender are both endogenously determined in equilibrium. We show that the strategic implications of APP are such that a potential offender's decisions to commit distinct offenses are *negatively correlated* in equilibrium when the punishment incurred in case of conviction is large relative to the benefits from committing offenses. This negative correlation, together with witnesses' coordination motives, lowers the informativeness of testimonies and increases the expected number of offenses in equilibrium.

In our model, the potential offender—hereafter the *defendant*—has multiple opportunities to commit offenses,[3] each of which is associated with a distinct witness. Each witness observes whether the offense associated with him takes place. For example, the defendant could be a firm executive with multiple opportunities to commit fraud or to violate workplace regulations, and the witnesses could be employees who may observe distinct offenses and may become whistleblowers.[4]

Once each witness has privately observed whether the offense associated with him has occurred, the witnesses independently decide whether to accuse the defendant. A witness's decision is driven by three considerations: (i) a preference for convicting guilty defendants and acquitting innocent ones, (ii) a risk of retaliation if the witness makes an accusation that fails to result in a conviction (e.g., because no other witness comes forward), and (iii) a private, idiosyncratic bias for or against getting the defendant convicted, whose distribution is independent from the other witness's bias.[5] After observing witnesses' testimonies, a judge decides whether to convict or acquit the defendant.[6]

[2]One could perform a more general mechanism design analysis beyond comparing APP and DPP. We undertook this approach in an earlier version of this paper (Pei and Strulovici 2021) subsumed by the present version. In a somewhat different and more complex model and using a more involved analysis, we showed that when (i) the defendant may commit at most two offenses and (ii) the punishment in case of conviction is large relative to the benefit from committing an offense, the optimal mechanism is approximated by either APP or DPP so that focusing on these two rules is, under these conditions, without loss of optimality. This paper focuses on the simpler and more practical question of comparing the existing rule to the one suggested in the legal and philosophy literature.

[3]Formally, the potential offender becomes a defendant only if at least one accusation is made against him. We nonetheless use the shorter terminology for simplicity.

[4]In cases that concern abuses of power, police brutality, workplace bullying, and harassment, witnesses may be the victims of these offenses. In other cases, they may be mere observers.

[5]In our model, some offenses go unreported and, conversely, some false accusations are made. 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. 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 (2020) and Ba and Rivera (2023) using Chicago data.

[6]In an extension, we consider the case in which the defendant's punishment depends not only on the probability of guilt, but also on the probability that the defendant has committed multiple offenses. We show that this extension does not affect the results of the model with a simple conviction/acquittal decision.

We compare equilibrium outcomes across the adjudication rules APP and DPP. Under APP, a defendant is convicted if the probability that he has committed at least one offense (i.e., that he is guilty of something) exceeds some exogenous threshold. Under DPP, the defendant is convicted if there is one specific offense for which the probability that the defendant committed this offense exceeds some threshold. The APP and DPP criteria are identical when the defendant may commit at most one potential offense, but may lead to different outcomes when the defendant may commit multiple offenses.

Our analysis shows that APP and DPP lead to radically different outcomes when the defendant's loss (e.g., disutility) from being convicted is large relative to his benefit from committing an offense. Under this assumption, we show that (i) under APP, each witness's accusation is arbitrarily *uninformative* and the expected number of offenses is close to the conviction threshold being used, while (ii) under DPP, a witness's accusation is highly informative (in a likelihood-ratio sense) and the expected number of offenses is arbitrarily small.

Turning to judicial errors, we show that APP introduces a tension between deterrence and false convictions: while lowering the guilt-probability threshold used to convict defendants results in a lower expected number of offenses, it also increases the conditional probability that a convicted defendant did not commit any offense. Precisely, we show that when the threshold probability is $p \in (0, 1)$, the expected number of offenses under APP is close to p and the probability that a convicted defendant is innocent is close to $1 - p$. This tension, which may seem intuitive, does not arise under DPP.

We now provide intuition for our results, focusing on settings in which the defendant can commit at most two possible offenses (the extension to three or more offenses is also analyzed in the paper). We start by explaining why, under our maintained assumption that the punishment for conviction is large relative to the benefit from committing offenses, the defendant is convicted in equilibrium under APP only if *both* witnesses accuse him. To see this, suppose by way of contradiction that only one accusation sufficed to get the defendant convicted with positive probability, i.e., sufficed to raise the defendant's probability of guilt to the conviction threshold. Then, two accusations would raise the posterior probability of guilt even further, strictly above the conviction threshold, and result in the defendant being convicted with probability one. Facing a large punishment in case of conviction, the defendant would respond by not committing any offense so as to minimize the probability of facing multiple accusations, contradicting the fact that the defendant is convicted with positive probability in equilibrium.

This observation implies that the defendant's decisions to commit distinct offenses are *strategic substitutes*: the increase in the probability of conviction that results from committing a given offense is higher if the defendant also commits the other offense than if he does not.

The fact that offenses are strategic substitutes implies, in equilibrium, that witnesses' private signals are *negatively correlated*: a witness who observes an offense believes that the other witness is unlikely of also observing an offense, and vice versa.

Separately, because each witness faces retaliation if he is the only one to accuse the defendant, witnesses' decisions to accuse the defendant are *strategic complements*.

Combining the negative correlation of witnesses' signals and the complementarity of their accusation decisions creates a tension between witnesses' incentives to coordinate on the same reports and the fact that the individual signals that drive these reports are negatively correlated. This tension has the effect of reducing the dependence of witnesses' reports on their private signals, thereby diminishing the informativeness of their testimonies. This lower informativeness increases the equilibrium probability that the defendant commits offenses and the expected number of offenses.

DPP is immune to the logic just described. DPP uses each probability that the defendant is guilty of a specific offense as an input. Crucially, the defendant's probability of guilt for a specific offence need not increase if he is accused of another offense. In particular, when the decisions to commit distinct offenses are uncorrelated, a testimony pertaining to a given offense has no bearing on the likelihood of guilt for another offense and thus need not affect the probability of conviction according to DPP. In fact, we show that this lack of correlation across offenses *must* arise in equilibrium when DPP is used. In equilibrium, the judge's belief about the occurrence of any given offense reaches the conviction threshold if and only if the witness whose role it is to testify on this specific offense accuses the defendant, and this belief is independent of other witnesses' testimonies. In equilibrium, witnesses' private observations are uncorrelated, which neutralizes the adverse effect that their coordination motive had, under APP, on the credibility of their testimonies. Finally, the defendant's decisions to commit distinct offenses are no longer strategic substitutes (and neither are they complements), because the probability of conviction is *linear* in the number of accusations made against the defendant.

Our findings are robust with respect to several extensions of our model. First, we consider the case in which the defendant has a private type that affects his benefit from committing offenses. We consider the case of binary types, in which the defendant is either a *virtuous type* who does not benefit from committing offenses or an *opportunistic type* who benefits from committing offenses as before. We show that when the probability of the virtuous type is above some cutoff, the opportunistic type will commit both offenses with positive probability. Nevertheless, our main theorems extend after we account for such heterogeneity.

Next, we extend our theorems to the case of three or more offenses maintaining our assumption that each offense, when it occurs, is observed by a distinct witness. We show that, under APP, the informativeness of witnesses' testimonies, pooled together, regarding the defendant's likelihood of guilt is *decreasing* in the number of potential offenses, even though there are more witness reports available to the judge. Moreover, we find that the unconditional probability that a given witness accuses the defendant *increases* in the number of potential offenses. Since each accusation bears a potential cost for the accuser but increases the value of accusations by other witnesses, accusations may be viewed as a form of contribution to a public good between witnesses. Viewed in this light , the fact that each witness is more likely to accuse the defendant as the number of witnesses increases distinguishes our analysis from standard theories of public good provision, in which contributions become scarcer as the number of players increases.

We also consider the following variations of our model: (i) the defendant faces a larger punishment when the probability that he has committed *multiple* offenses exceeds some exogenous threshold; (ii) the defendant's utility from committing offenses is concave in the number of offenses; (iii) a witness who makes a false accusation against the defendant may be uncovered and punished ex post with positive probability; (iv) witnesses have an intrinsic preference for reporting the truth (above and beyond the effect of their testimony on the verdict); (v) each witness cares not only about the potential offense that he is associated with, but also about all other offenses that the defendant may have committed; (vi) the number of opportunities to commit offenses is not publicly known, and is instead the defendant's private information; and (vii) the witnesses incur a positive cost when they accuse the defendant even when the defendant is convicted.

In summary, our comparison between APP and DPP suggests a rationale for using a conviction criterion that treats each accusation independently of other offenses that the defendant may have committed. Our results echo some critiques of procedures that link accusations across potential witnesses, especially those that are related to the credibility of witnesses' reports.[7]

One caveat for the applicability of our results is that these results are obtained using equilibrium analysis. While equilibrium analysis is the norm in economics, it rests on the assumption that all players know one another's equilibrium strategies and are playing mutual best replies to these strategies. This assumption is more plausible when the environment has remained stable for an extended period of time than after a change in the environment, which might correspond in our context to a change in the conviction rule (e.g., APP vs. DPP) used for adjudication or to a discrete jump in the punishment level imposed when convicted of a given offense.

We review the related literature in the remainder of this section. Section 2 sets up our model. Section 3 analyzes a single-offense benchmark in which the defendant cannot commit one of the offenses. Section 4 states and shows our main results, which compare the equilibrium outcomes under APP and DPP when the defendant can commit multiple offenses. Section 5 studies several extensions which include settings with three or more agents and settings in which the defendant has private information about his propensity to commit offenses. Section 6 concludes.

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

Related Literature: In addition to the debate concerning the potential benefits of aggregating offenses, our paper contributes to several literatures that concern strategic communication, voting, and law and economics.

First, our paper is related to the literature on strategic communication with multiple senders (Battaglini (2002, 2017), Ambrus and Takahashi (2008), Morgan and Stocken (2008), and Ekmekci and Lauermann (2020, 2022)) and when the sender's information is endogenous (Pei (2015), Argenziano, Severinov and Squintani (2016), and Kreutzkamp (2023)). In all these works, senders communicate information about some *exogenous* state of the world. In the present paper, by contrast, witnesses (the senders in our model) communicate information about an *endogenous* object, which is the *defendant's action*. As a result, the correlation between senders' private signals is determined by the defendant's incentives and is therefore endogenous. Our results shed light on how this endogenous correlation structure combined with the senders' coordination motives, affects the informativeness of communication which in turn affects the defendant's equilibrium behavior.

Second, our paper is related to the literature on strategic voting (Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1996)). Witnesses' testimonies under the APP criterion may be viewed as a voting game in which witnesses vote on whether the defendant is guilty of having committed at least one offense. In contrast to the earlier literature, in our model the correlation between voters' private signals is endogenous and the voting rule used for conviction is also endogenously determined. These features also distinguish our analysis from Schmitz and Tröger (2012) and Ali, Mihm, and Siga (2018), who analyze the performances of voting rules when voters' payoffs are negatively but exogenously correlated, and from Persico (2004), Gershkov and Szentes (2009), and Strulovici (2022), who analyze voters' incentives to acquire information about some exogenous state of the world.

Third, our paper contributes to the law and economics literature. Our main contribution is to compare the equilibrium outcomes under different conviction rules when potential offender may commit multiple offenses and when the offender's incentives to commit offenses and witnesses' incentives to testify are both endogenous. In addition to the works from legal scholarship cited above, our paper is most closely related to the contemporaneous works of Lee and Suen (2020) and Cheng and Hsiaw (2022). 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 (2022) adopt a global game perspective to study the reporting incentives of a continuum of agents who observe conditionally independent signals of the state. In contrast to those papers, endogenizing the behaviors of potential offenders is the main innovation of our model.

2 Model

Consider a three-stage game between a defendant, two witnesses (which we will call *agents*),[8] and a judge.

In the first stage, the defendant chooses $\boldsymbol{\theta} \equiv (\theta_1, \theta_2) \in \{0,1\}^2$ where $\theta_i = 1$ stands for the defendant committing an offense witnessed by (or against) agent i. In the second stage, each agent $i \in \{1,2\}$ privately observes θ_i and the realization of a preference shock $\omega_i \in \mathbb{R}$. We assume that ω_1 and ω_2 are independently distributed, both are drawn from a normal distribution with cdf $\Phi(\cdot)$ and pdf $\phi(\cdot)$.[9] Then the agents simultaneously decide whether to accuse the defendant ($a_i = 1$) or not ($a_i = 0$). In the third stage, the judge observes $\boldsymbol{a} \equiv (a_1, a_2) \in \{0,1\}^2$ and chooses $s \in \{0,1\}$, where $s = 1$ stands for convicting the defendant and $s = 0$ stands for acquitting the defendant.

A *strategy profile* consists of a tuple $\left\{\sigma_o, (\sigma_i)_{i=1}^2, q\right\}$ where: $\sigma_o \in \Delta\big(\{0,1\}^n\big)$ is the defendant's strategy and represents a probability distribution over vectors $\boldsymbol{\theta}$ of offenses; $\sigma_i : \mathbb{R} \times \{0,1\} \to [0,1]$ is agent i's strategy and maps each realization of ω_i and θ_i to a probability that i accuses the defendant; and $q : \{0,1\}^n \to [0,1]$ maps agents' reports $\boldsymbol{a} \in \{0,1\}^2$ to the probability of conviction ($s = 1$).

The defendant's payoff is given by $\sum_{i=1}^2 \theta_i - sL$, and interpreted as follows: the defendant receives a benefit normalized to 1 for each committed offense, and loses $L > 0$ if he is convicted. Agent $i \in \{1,2\}$'s payoff is given by

$$u_i(\omega_i, \theta_i, a_i) \equiv \begin{cases} \omega_i + b\theta_i & \text{if} \quad s = 1 \\ -ca_i & \text{if} \quad s = 0. \end{cases} \tag{2.1}$$

This payoff should be understood as follows: if the defendant is convicted, agent i receives an idiosyncratic private benefit ω_i plus an additional benefit $b > 0$ if i had observed offense $\theta_i = 1$ committed by the defendant. This payoff structure implies that, other things being equal, i has a stronger incentive to accuse the defendant when ω_i is higher and when i has observed an offense. In addition, i incurs a cost c if i accuses the defendant but this accusation does not lead to a conviction ($s = 0$). This cost may be interpreted as defendant retaliating against i in case of a failed accusation.

Section 5.3 considers several important extensions, including one in which the defendant has a private type that determines his benefit from from committing offenses and one in which the defendant faces decreasing marginal returns from committing offenses. That section also considers the case in which agents have a private benefit from telling the truth irrespective of the conviction outcome or from convicting defendants who have committed offenses against other agents, in addition to the one that they get to directly to observe.

[8]Section 5.2 studies an extension in which there are three or more witnesses.

[9]Our main result (Theorem 1) and our comparative statics result (Theorem 3) hold regardless of the distribution of agents' preference shocks. Proposition 1 and Theorem 2 also extend to other distributions as long as their left tails are not too fat in the sense that $\lim_{\omega \to +\infty} \frac{\Phi(\omega - b)}{\Phi(\omega)} = 0$ for every $b > 0$. Our condition is satisfied, for example, when the distribution is normal.

As noted in the Introduction, an important consideration to frame our analysis is the debate in legal scholarship regarding whether one should aggregate accusations faced by a defendant to form an overall probability of the defendant's guilt. To engage with this debate, we compare equilibrium outcomes for the two adjudication criteria used in this literature: the *Aggregate Probabilities Principle* (or APP) and the *Distinct Probabilities Principle* (or DPP). Under APP, a defendant is convicted if the probability that he is guilty of committing *at least one offense* (i.e., the probability that $\max_i\{\theta_i\} = 1$) exceeds some threshold, such as 95%, interpreted as the judge's standard of proof. Under DPP, by contrast, a defendant is convicted if there is at least one offense for which the probability that the defendant is guilty of this specific offense exceeds some threshold. Mathematically, DPP leads to a conviction if $\max_i \Pr(\theta_i = 1)$ exceeds some threshold.

In practice, DPP is used in many criminal justice systems. For example, 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. By contrast, APP is more commonly used for punishments meted out outside the *criminal* justice system, for example in the case of managers charged with discriminating against minority workers, or when supervisors are charged with abusing their subordinates, where the accumulation of charges is sometimes used as proxy for the probability of guilt of the accused. This perspective is supported by Schauer and Zeckhauser (1996) who write 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*. Harel and Porat (2009) advocate for a broader use of APP in judicial settings, arguing that acquitting defendants who are almost surely guilty of some unspecified crime is *neither just nor efficient* (page 263).

We now formally define APP and DPP and their corresponding equilibria. A strategy profile $\left\{\sigma_o, (\sigma_i)_{i=1}^2, q\right\}$ is an *equilibrium* under APP and standard of proof $\pi^* \in (0,1)$ if the following holds: σ_o maximizes the defendant's payoff given (σ_1, σ_2, q); σ_i maximizes agent i's payoff given $(\sigma_o, \sigma_{-i}, q)$ for every $i \in \{1,2\}$; and $q(\boldsymbol{a})$ satisfies

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

where $\overline{\theta} \equiv \max_{i\in\{1,2\}} \theta_i$ captures whether the defendant is guilty of at least one offense.

A strategy profile $\left\{\sigma_o, (\sigma_i)_{i=1}^2, q\right\}$ is an *equilibrium* under DPP and standard of proof $\pi^* \in (0,1)$ if the following holds: σ_o maximizes the defendant's payoff given (σ_1, σ_2, q); σ_i maximizes agent i's payoff given $(\sigma_o, \sigma_{-i}, q)$ for every $i \in \{1,2\}$; and $q(\boldsymbol{a})$ satisfies

$$q(\boldsymbol{a}) = \begin{cases} 1 & \text{if} \quad \max_{i\in\{1,2\}} \Pr\left(\theta_i = 1\middle|\boldsymbol{a}\right) > \pi^* \\ 0 & \text{if} \quad \max_{i\in\{1,2\}} \Pr\left(\theta_i = 1\middle|\boldsymbol{a}\right) < \pi^*. \end{cases} \tag{2.3}$$

As is standard in models with discrete action spaces, we allow players to use mixed strategies in order to ensure the existence of equilibrium. In particular, the judge may convict the defendant with any probability when she is indifferent, which corresponds to the condition $\Pr(\overline{\theta} = 1|\boldsymbol{a}) = \pi^*$ under APP and to the condition $\max_{i \in \{1,2\}} \Pr\left(\theta_i = 1\middle|\boldsymbol{a}\right) = \pi^*$ under DPP.

In our model, each agent i signals his private information about θ_i and ω_i by choosing whether to accuse the defendant. As in other signaling games, there generally exist numerous equilibria, including some which are arguably unnatural. Fortunately, it is possible to rule out these equilibria using two refinements which are easy to interpret in our context. First, regardless of the conviction rule and the standard of proof $\pi^* \in (0,1)$, there always exist equilibria in which the defendant commits all offenses and is convicted even when no agent accuses him. These equilibria violate the principle–akin to a form of presumption of innocence–that a defendant should *not* be convicted based on the sole basis of a judge's prior belief. This motivates the following refinement:

Refinement 1 (No Conviction Unless Accused)**.** *(i) The defendant is acquitted for sure if no agent accuses him. (ii) The event that no agent accuses the defendant occurs with strictly positive probability.*

Refinement 1 implies that the defendant cannot, in equilibrium, be guilty with probability 1. Otherwise, the judge's posterior belief would assign probability 1 to the defendant being guilty, regardless of the vector $\boldsymbol{a}$ of accusations made against the defendant on the equilibrium path, including when no agent accuses him, which violates the refinement.

Next, the judge's conviction decision is a function of her *posterior belief* about (θ_1, θ_2), formed based on the set of agents who accuse the defendant. There may exist "contrarian" equilibria in which agents are more likely to accuse the defendant if they did not observe an offense and making an "accusation" (i.e., $a_i = 1$) runs contrary to its intended meaning. Refinement 2 rules out this case and ensures that accusations have their intuitive meaning by requiring that each accusation weakly increases the probability of conviction and, hence, that each accusation is a move against the defendant.

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

In equilibria that satisfy Refinement 2, each accusation is a move against the defendant. This is consistent with the interpretation that the parameter c appearing in agents' payoff functions is a retaliation cost borne by an agent who accused the defendant but failed to secure a conviction. This refinement is consistent with other models of retaliation. In Chassang and Padró i Miquel (2019), for instance, a defendant commits to retaliate against messages that increase the defendant's probability of conviction, but not against messages that do not have this effect.

3 Single-Offense Benchmark

Before considering the case of multiple potential offenses, we begin the analysis with a single-offense benchmark in which the defendant can commit an offense only against agent 1, but *cannot* commit offense against agent 2. Under this assumption, APP and DPP coincide since $\Pr(\overline{\theta} = 1) = \Pr(\theta_1 = 1) = \max_{i \in \{1,2\}} \Pr(\theta_i = 1)$.

We show, in this context, that when the defendant's loss from conviction L is large relative to his benefit from committing the offense, the probability that the defendant commits the offense (equivalently, since there is only possible offense, the expected number of offenses) vanishes in equilibrium.

Proposition 1. *Suppose that the defendant can only commit an offense against agent* 1*. For every* $\varepsilon > 0$ *and* $\pi^* \in (0, 1)$*, there exists* $\overline{L} > 0$ *such that when* $L > \overline{L}$*, there exists a unique equilibrium that satisfies Refinement 1. In this equilibrium, the defendant commits the offense (i.e.,* $\theta_1 = 1$*) with probability less than* ε*.*

Proposition 1 shows that when the punishment in case of a conviction is large relative to the defendant's benefit from committing offenses, the probability with which offenses taking place vanishes to zero. Although the mechanisms are different, this result is reminiscent of the classic finding in Becker (1968) that larger punishments are effective at deterring criminal behaviors.[10]

The proof Proposition 1, provided in Appendix A, has the following intuition. The judge makes her conviction decision based on her *posterior belief* about the defendant's guilt $\overline{\theta}$ after observing agents' accusations $\boldsymbol{a} \equiv (a_1, a_2)$. According to Bayes rule, the judge's posterior belief $\Pr(\overline{\theta} = 1|\boldsymbol{a})$ satisfies

$$\frac{\Pr(\overline{\theta} = 1|\boldsymbol{a})}{\Pr(\overline{\theta} = 0|\boldsymbol{a})} = \frac{\Pr(\boldsymbol{a}|\overline{\theta} = 1)}{\Pr(\boldsymbol{a}|\overline{\theta} = 0)} \cdot \frac{\Pr(\overline{\theta} = 1)}{\Pr(\overline{\theta} = 0)}. \tag{3.1}$$

This equation implies that, fixing the defendant's equilibrium strategy $\Pr(\overline{\theta} = 1)$, or equivalently the judge's *prior belief* about $\overline{\theta}$, the judge's *posterior belief* after observing $\boldsymbol{a}$ is pinned down by the following likelihood ratio:

$$\mathcal{I}(\boldsymbol{a}) \equiv \frac{\Pr(\boldsymbol{a}|\overline{\theta} = 1)}{\Pr(\boldsymbol{a}|\overline{\theta} = 0)}. \tag{3.2}$$

Equation (3.1) suggests that $\mathcal{I}(\boldsymbol{a})$ measures the *informativeness* of the agents' accusations $\boldsymbol{a}$, in the sense that a larger $\mathcal{I}(\boldsymbol{a})$ implies that the defendant is more likely to be guilty following $\boldsymbol{a}$. When the defendant does not have the ability to commit offense against agent 2, $\mathcal{I}(\boldsymbol{a})$ depends only on a_1 on the equilibrium path. Since Refinement 1 requires that $q(0, 0) = 0$, we show that in equilibrium, the defendant is convicted with positive probability only

[10]The settings are quite different. For instance, Becker does not consider the possibility of convicting an innocent defendant and treats the probability of convicting a guilty defendant as an instrument chosen by the regulator, rather than the result of a signal based on witness with endogenous incentives.

if agent 1 accuses him. Let q denote the *expected* probability that the defendant is convicted after agent 1 accuses him. From agent 1's perspective, if he has witnessed an offense, then he prefers to accuse the defendant when

$$q(b+\omega_1)-(1-q)c \geq 0 \quad \text{or equivalently} \quad \omega_1 \geq \frac{(1-q)c}{q}-b. \tag{3.3}$$

If he has not witnessed any offense, then he prefers to accuse the defendant when

$$q\omega_1-(1-q)c \geq 0 \quad \text{or equivalently} \quad \omega_1 \geq \frac{(1-q)c}{q}. \tag{3.4}$$

Hence, the defendant's cost of committing an offense against agent 1 equals $\Big(\Phi(\frac{(1-q)c}{q})-\Phi(\frac{(1-q)c}{q}-b)\Big)L$. In equilibrium, the defendant's expected cost from committing an offense must be the same as his benefit 1. This implies that as L becomes large, q must go to 0, and the informativeness of agent 1's accusation $\mathcal{I}(a_1=1,a_2)$ diverges to $+\infty$.

Since the equilibrium value of q is strictly between 0 and 1, the judge must be indifferent between convicting and acquitting the defendant after observing $a_1=1$, which means that the posterior probability of guilt in this case must be equal to the judge's standard of proof π^*. Equation (3.1) then implies that for any fixed π^*, the prior probability of guilt vanishes to 0 as $\mathcal{I}(1,a_2)$ diverges to infinity.

4 Comparisons between APP and DPP

We now turn our focus on the main case of interest, in which the defendant can commit multiple offenses.

As in the case of Proposition 1, we focus on the case in which the punishment meted out to the defendant in case of a conviction is *large* relative to his benefit from committing offenses and analyze equilibria that satisfy Refinements 1 and 2. We further focus on equilibria in which the defendant uses *symmetric* strategies, i.e., chooses $(\theta_1,\theta_2)=(1,0)$ and $(\theta_1,\theta_2)=(0,1)$ with the same probability. As it turns out, this symmetry requirement is satisfied by every *proper equilibrium* as defined by Myerson (1978). For simplicity and brevity, we directly impose symmetry rather than define proper equilibrium and prove that it implies symmetry.

From here onwards, we will refer to *symmetric equilibria that satisfy Refinements 1 and 2* simply as *equilibria*. Section 4.1 establishes preliminary results and properties of equilibria that hold for both APP and DPP. Sections 4.2 and 4.3 present our main results, which show that APP leads to uninformative accusations and a large expected number of offenses and DPP leads to informative accusations and effective deterrence. Section 4.4 discusses the implications of our results on the use of APP and DPP, and acknowledges their limitations.

4.1 Common Properties of APP and DPP

In this section, we establish properties of equilibria that hold under both APP or DPP. A useful observation is that the defendant's incentive to commit offenses and agents' incentives to accuse the defendant depend on the judge's conviction rule (whether it is APP or DPP) only through the mapping from the agents' accusations to the probabilities of conviction $q : \{0,1\}^2 \to [0,1]$. Hence, changes in the conviction rule affect the defendant's and the agents' incentives only through the mapping q that arises in equilibrium.

We start by showing that regardless of the conviction rule and of the standard of proof $\pi^* \in (0,1)$, the probability that the defendant commits any given offense lies strictly between 0 and 1 in every equilibrium. Furthermore, when L is large enough, each agent's equilibrium strategy must take the form of two cutoffs–one used when the agent has witnessed an offense and one used when he has not–such that the agent accuses the defendant *if and only if* the agent's realized payoff shock ω_i exceeds the cutoff corresponding to agent's observation.

Lemma 4.1. *In every equilibrium,* $\Pr(\theta_i = 1) \in (0,1)$ *for every* $i \in \{1,2\}$. *There exists* $\overline{L} > 0$ *such that in every equilibrium when* $L > \overline{L}$, *agent* i*'s strategy is characterized by two cutoffs* ω_i^* *and* ω_i^{**} *with* $-\infty < \omega_i^* < \omega_i^{**} < +\infty$ *such that*

1. *When* $\theta_i = 1$, *agent* i *accuses the defendant if and only if his preference shock* ω_i *is greater than* ω_i^*.

2. *When* $\theta_i = 0$, *agent* i *accuses the defendant if and only if his preference shock* ω_i *is greater than* ω_i^{**}.

The proof of this result, provided in Appendix B, has the following intuition. Refinement 1 requires that $q(0,0) = 0$ and that the report vector $\boldsymbol{a} = (0,0)$ occurs with strictly positive probability. This, together with Bayes rule, implies that for any equilibrium and under any conviction threshold $\pi^* \in (0,1)$, the defendant must be innocent of each offense with strictly positive probability, i.e., that $\Pr(\theta_i = 1) < 1$ for every i. The intuition for the fact that the defendant's probability of guilt is strictly positive is reminiscent of the equilibrium logic arising in inspection games: if offenses took place with zero probability, the judge's posterior belief about the defendant's guilt would be equal to zero following every report vector $\boldsymbol{a}$ that arises on the equilibrium path, and the defendant would not be punished. This, in turn would give the defendant a strict incentive to commit offenses, contradicting the hypothesis that the defendant commits no crime.

Turning to agents' behavior, we first observe that agent i chooses $a_i = 1$ with strictly higher probability when $\theta_i = 1$. If this were not the case, the defendant would strictly prefer to commit the corresponding offense ($\theta_i = 1$; a situation that we have already shown cannot arise in equilibrium) as this would, in addition to providing the benefit from committing the offense, decrease the probability of being accused by i and lead to a weakly lower probability of conviction.

To understand why agents' equilibrium strategies take the form of two cutoffs, let us compare their payoffs when they accuse the defendant to their payoffs when they do not. For $j \in \{1,2\}$ and $k \in \{0,1\}$, let $Q_{k,j}$ denote the probability that $a_j = 1$ conditional on the event that $\theta_i = k$, where $i \neq j$ denotes the agent other than j. If $\theta_1 = 1$, then agent 1 prefers to choose $a_1 = 1$ when

$$\underbrace{\Big\{Q_{1,2}(q(1,1)-q(0,1)) + (1-Q_{1,2})(q(1,0)-q(0,0))\Big\}}_{\text{the expected increase in conviction probability when agent 1 accuses the defendant}}(\omega_1 + b)$$

$$\geq c \qquad \underbrace{\Big\{1 - Q_{1,2}q(1,1) - (1-Q_{1,2})q(1,0)\Big\}}_{\text{the expected probability that the defendant is acquitted if agent 1 accuses him}} \tag{4.1}$$

where the LHS is his expected payoff when he accuses the defendant and the RHS is his expected payoff when he does not accuse the defendant. Similarly, if $\theta_1 = 0$, then agent 1 prefers to choose $a_1 = 1$ when

$$\Big\{Q_{0,2}(q(1,1)-q(0,1)) + (1-Q_{0,2})(q(1,0)-q(0,0))\Big\}\omega_1 \geq c\Big\{1 - Q_{0,2}q(1,1) - (1-Q_{0,2})q(1,0)\Big\}. \tag{4.2}$$

Refinements 1 and 2 imply that $q(1,1) \geq q(0,1)$ and $q(1,0) \geq q(0,0) = 0$. Hence, the coefficients in front of ω_1 are non-negative for both (4.1) and (4.2). To show that agent 1's equilibrium strategy takes the form of cutoffs, we only need to rule out cases where where agent 1 believes that his accusation has no impact on the probability of conviction, i.e., either $Q_{1,2}(q(1,1)-q(0,1))+(1-Q_{1,2})q(1,0)$ or $Q_{0,2}(q(1,1)-q(0,1))+(1-Q_{0,2})q(1,0)$ is 0. Intuitively, these cases cannot arise since once agent 1's accusation has no impact on the probability of conviction, the defendant will have a strict incentive to commit offenses against against agent 1. This will then contradict our earlier conclusion that the probability of guilt against each agent must be strictly between 0 and 1.

Next, we show that when the punishment for conviction L is above some threshold, the defendant must be convicted with probability less than 1 in every equilibrium regardless of the conviction rule and of the standard of proof used to convict the defendant.

Lemma 4.2. *There exists $\overline{L} \in \mathbb{R}_+$ such that for every $L > \overline{L}$ and $\pi^* \in (0,1)$, $q(1,1) \in (0,1)$ in every equilibrium under APP and standard of proof π^* and in every equilibrium under DPP and standard of proof π^*.*

The proof of this lemma is provided in Online Appendix A. The key step of the proof is to show that, if the defendant were convicted for sure when two accusations are made against him, then the increase in conviction probability when the defendant commits an additional offense would have to be strictly positive and bounded below away from zero for all punishment levels L large enough. As L increases, the marginal cost from committing offenses would then have to eventually exceed the marginal benefit from committing each offense (which, we recall, has been normalized to 1), which would give the defendant a strict incentive not to commit any offense and contradict Lemma 4.1.

Lastly, we provide conditions on q that characterize when the defendant's choices of θ_1 and θ_2 are substitutes, and when they are complements. Conditional on his choice of θ_j, the defendant's expected cost of committing an offense against agent $i \neq j$ is

$$C_i(\theta_j) \equiv \Big(\mathbb{E}[q(a_i, a_j)|\theta_i = 1, \theta_j] - \mathbb{E}[q(a_i, a_j)|\theta_i = 0, \theta_j]\Big)L. \tag{4.3}$$

We say that the defendant's choices of θ_1 and θ_2 are *strict complements* if $C_i(1) < C_i(0)$ for every $i \in \{1, 2\}$, that is, if the defendant's expected cost of committing an offense against agent i is lower if he commits an offense against agent j than when he doesn't. Similarly, say that the defendant's decision variables θ_1 and θ_2 are *strict substitutes* if $C_i(1) > C_i(0)$ for every $i \in \{1, 2\}$.

Lemma 4.3. *The defendant's choices of θ_1 and θ_2 are strict substitutes if and only if*

$$q(1, 1) + q(0, 0) - q(1, 0) - q(0, 1) > 0. \tag{4.4}$$

His choices of θ_1 and θ_2 are strict complements if and only if $q(1, 1) + q(0, 0) - q(1, 0) - q(0, 1) < 0$.

Proof. Recall from Lemma 4.1 that agent i accuses the defendant with probability $\Psi_i^* \equiv 1 - \Phi(\omega_i^*)$ when $\theta_i = 1$, and with probability $\Psi_i^{**} \equiv 1 - \Phi(\omega_i^{**})$ when $\theta_i = 0$, where $\Psi_i^* > \Psi_i^{**}$. If the defendant increases θ_1 from 0 to 1 when $\theta_2 = 0$, the the probability of conviction increases by

$$(\Psi_1^* - \Psi_1^{**})\Big((1 - \Psi_2^{**})\big(q(1, 0) - q(0, 0)\big) + \Psi_2^{**}\big(q(1, 1) - q(0, 1)\big)\Big). \tag{4.5}$$

If the defendant increases θ_1 from 0 to 1 when $\theta_2 = 1$, the the probability of conviction increases by

$$(\Psi_1^* - \Psi_1^{**})\Big((1 - \Psi_2^{*})\big(q(1, 0) - q(0, 0)\big) + \Psi_2^{*}\big(q(1, 1) - q(0, 1)\big)\Big) \tag{4.6}$$

The defendant's choices of θ_1 and θ_2 are strict substitutes *if and only if* (4.5) is strictly less than (4.6), which is the case if and only if $(\Psi_1^* - \Psi_1^{**})(\Psi_2^* - \Psi_2^{**})\big(q(1, 0) + q(0, 1) - q(0, 0) - q(1, 1)\big) < 0$. Since $\Psi_i^* > \Psi_i^{**}$ for every i, the above inequality is equivalent to (4.4). A similar argument applies to the statement on strict complements. $\square$

4.2 Aggregate Probabilities Principle

This section characterizes equilibrium outcomes under APP. Recall the formula for the judge's posterior belief about the defendant's guilt in (3.1) and the measure for the informativeness of agents' accusations in (3.2).

Theorem 1. *For every $\pi^* \in (0,1)$ and $\varepsilon > 0$, there exists $\overline{L} > 0$ such that when $L > \overline{L}$, in every equilibrium under APP and standard of proof π^*, we have:*

1. ***Negative Correlation Across Offenses:*** $\Pr(\theta_i = 1|\theta_j = 1) < \Pr(\theta_i = 1|\theta_j = 0)$ *for every* $i \neq j$.
2. ***Uninformative Accusations:*** $\max_{\boldsymbol{a} \in \{0,1\}^2} \mathcal{I}(\boldsymbol{a}) < 1 + \varepsilon$.[11]
3. ***Ineffective Deterrence:*** $\sum_{i=1}^{2} \Pr(\theta_i = 1) > \pi^* - \varepsilon$.

Theorem 1 suggests that when the punishment in case of conviction L is large relative to the benefit from committing offenses, APP induces *negative correlation* between the occurrence of the distinct offenses. This negative correlation has, in turn, the effect of weakening the credibility of agents' accusations and results in a high expected number of offenses. These conclusions stand in contrast to the single-offense benchmark, in which the expected number of offenses vanishes and the informativeness of the agent's accusation diverges to infinity as the punishment for conviction increases.

Two aspects of Theorem 1 are worth emphasizing before presenting its proof. First, the expected number of offenses in equilibrium is close to the standard of proof π^*. Although a decrease in the standard of proof can lower the expected number of offenses, such a decrease will also *increase* the probability that a convicted defendant is innocent. In fact, we show that when the standard of proof is π^*, the probability that a convicted defendant is innocent equals $1 - \pi^*$. For example, if the standard of proof is only 20%, then a convicted defendant is innocent with probability 80%. From this perspective, Theorem 1 highlights a tension caused by APP between achieving deterrence and reducing the probability that convicted defendants are innocent. As we will see, this tension does not exist under DPP, in the sense a high standard of proof π^* guarantees that both the expected number of offenses and the probability that convicted defendants being innocent are low.

Second, since each agent incurs a cost c of accusing the defendant, one might conjecture that ineffective deterrence is caused by agents' free-riding motives, as in public good provision games. However, we show in Theorem 3 of Section 5.2 that this intuition is incorrect in our setting. In fact, each agent accuses the defendant with *strictly* higher probability when there are more agents.

Next, we present the proof of Theorem 1, which proceeds in three steps.

[11] By definition, $\max_{\boldsymbol{a} \in \{0,1\}^2} \mathcal{I}(\boldsymbol{a})$ is at least 1. Hence, Theorem 1 implies that $\mathcal{I}$ is arbitrarily close to the trivial lower bound.

Step 1: The Defendant's Incentives Since each agent $i \in \{1,2\}$ is strictly more likely to accuse the defendant when $\theta_i = 1$ than when $\theta_i = 0$, Bayes rule implies that

$$\Pr(\overline{\theta} = 1 | a_i = 1, a_j) > \Pr(\overline{\theta} = 1 | a_i = 0, a_j) \text{ for every } a_j \in \{0,1\}. \tag{4.7}$$

That is to say, each additional accusation *strictly increases* the probability that the defendant is guilty of at least one offense. This observation plays a key role in the analysis of APP, and distinguishes it from the analysis of DPP, as explained in Section 4.3.

Suppose by way of contradiction that either $q(0,1) > 0$ or $q(1,0) > 0$. Then, the posterior probability that $\overline{\theta} = 1$ is weakly greater than π^* either when $\boldsymbol{a} = (1,0)$ or when $\boldsymbol{a} = (0,1)$. Inequality (4.7) then implies that the posterior probability that $\overline{\theta} = 1$ is *strictly* greater than π^* when $\boldsymbol{a} = (1,1)$. Hence, in equilibrium under APP with standard of proof π^*, we have $q(1,1) = 1$. This contradicts Lemma 4.2 that $q(1,1) < 1$ when L is large enough.

Therefore, $q(1,0) = q(0,1) = q(0,0) = 0$ and $q(1,1) \in (0,1)$ in every equilibrium. Since the equilibrium conviction probabilities satisfy (4.4), Lemma 4.3 implies that the defendant's decisions to commit different offenses are strict substitutes. Since Refinement 1 requires that $(\theta_1, \theta_2) = (0,0)$ occurs with positive probability, i.e., committing no offense is optimal for the defendant. This implies that the marginal cost of committing the first offense is weakly greater than 1. Since the marginal cost of committing a second offense is strictly greater than the marginal cost of committing the first offense, it is strictly suboptimal for the defendant to commit two offenses.

Since the equilibrium probability of guilt is interior, the defendant must be indifferent between committing no offense and committing one offense. Therefore, his marginal cost of committing the first offense is 1. Since the defendant's strategy is required to be symmetric across agents, he will choose $(0,0)$, $(1,0)$, and $(0,1)$ with positive probability, and choose $(1,0)$ and $(0,1)$ with the same probability. This implies that in equilibrium, θ_1 and θ_2 are *negatively correlated*. Let $\pi \equiv \Pr(\overline{\theta} = 1)$ denote the equilibrium probability that the defendant is guilty of at least one offense. We know that $\Pr(\theta_i = 1) = \pi/2$ for every $i \in \{1,2\}$, and that the conditional on $\theta_i = 0$, the probability with which $\theta_j = 0$ equals $\beta \equiv \frac{1-\pi}{1-\pi/2}$.

Step 2: Symmetry in Agents' Equilibrium Strategies Since the defendant chooses both $(\theta_1, \theta_2) = (1,0)$ and $(\theta_1, \theta_2) = (0,1)$ with positive probability in equilibrium, his indifference condition yields:

$$\frac{1}{qL} = (1 - \Phi(\omega_1^{**}))\Big(\Phi(\omega_2^{**}) - \Phi(\omega_2^*)\Big) = (1 - \Phi(\omega_2^{**}))\Big(\Phi(\omega_1^{**}) - \Phi(\omega_1^*)\Big). \tag{4.8}$$

To simplify notation, let $q \equiv q(1,1)$. From agents' incentive constraints (4.1) and (4.2), $q(0,0) = q(1,0) = q(0,1) = 0$, implies that when $\theta_i = 1$, agent i accuses the defendant if and only if

$$\omega_i \geq \omega_i^* \equiv -b - c + \frac{c}{qQ_{1,j}}, \tag{4.9}$$

and when $\theta_i = 0$, agent i accuses the defendant if and only if:

$$\omega_i \geq \omega_i^{**} \equiv -c + \frac{c}{qQ_{0,j}}. \tag{4.10}$$

Since $\Pr(\theta_i = 1|\theta_j = 1) = 0$ and $\Pr(\theta_i = 1|\theta_j = 0) = 1 - \beta$, we know that $Q_{1,j} = 1 - \Phi(\omega_j^{**})$ and $Q_{0,j} = \beta(1 - \Phi(\omega_j^{**})) + (1-\beta)(1-\Phi(\omega_j^*))$. Suppose by way of contradiction that $\omega_j^{**} > \omega_i^{**}$. Then (4.8) is true only when $\omega_j^* > \omega_i^*$. However, the expression for ω_i^* is strictly decreasing in $Q_{1,j}$ and $Q_{1,j}$ is strictly decreasing in ω_j^{**}. Therefore, $\Phi(\omega_j^{**}) > \Phi(\omega_i^{**})$ together with (4.9) implies that $\omega_i^* > \omega_j^*$. This leads to a contradiction and implies that $\omega_1^{**} = \omega_2^{**}$. Equation (4.8) then implies that $\omega_1^* = \omega_2^*$. In the remainder of this section, we can therefore simplify notation by denoting the cutoffs by ω^* and ω^{**} without subscripts and writing Q_1 and Q_0 instead of $Q_{1,j}$ and $Q_{0,j}$.

Step 3: Coordination under Negative Correlation Since agent i incurs a cost c when he chooses $a_i = 1$ but fails to convict the defendant and since $q(a_1, a_2)$ is weakly increasing in both of its arguments, agent i's incentive to report $a_i = 1$ is increasing in the probability he assigns to the event $a_j = 1$. That is, agents' decisions to accuse the defendant are *strategic complements*. Since $\omega^* < \omega^{**}$, we know that $Q_1 < Q_0$. The expressions for the cutoffs then imply that $\omega^{**} - \omega^* \in (0, b)$. The fact that this difference is strictly less than b stands in contrast to the single-offense benchmark of Section 3, and it plays an important role in the analysis as it implies that agents' reports are less informative relative to the single-offense benchmark. Since the defendant never commits two offenses in equilibrium, we have

$$\mathcal{I}(1,1) \equiv \frac{\Pr(\boldsymbol{a} = (1,1)|\overline{\theta} = 1)}{\Pr(\boldsymbol{a} = (1,1)|\overline{\theta} = 0)} = \frac{1 - \Phi(\omega^*)}{1 - \Phi(\omega^{**})}. \tag{4.11}$$

Since $q(1,1) \in (0,1)$, we also have $\Pr(\overline{\theta} = 1|\boldsymbol{a} = (1,1)) = \pi^*$. From Bayes rule, $\frac{\pi}{1-\pi}\mathcal{I}(1,1) = \frac{\pi^*}{1-\pi^*}$. Letting $l^* \equiv \frac{\pi^*}{1-\pi^*}$, we have

$$\beta = \frac{2\mathcal{I}}{l^* + 2\mathcal{I}} \text{ and } 1 - \beta = \frac{l^*}{l^* + 2\mathcal{I}}. \tag{4.12}$$

Plugging (4.12) into the expressions for Q_1 and Q_0, we obtain

$$\frac{Q_0}{Q_1} = \frac{\beta(1 - \Phi(\omega_j^{**})) + (1-\beta)(1-\Phi(\omega_j^*))}{1 - \Phi(\omega_j^{**})} = \beta + (1-\beta)\mathcal{I}(1,1) = \frac{(l^* + 2)\mathcal{I}(1,1)}{l^* + 2\mathcal{I}(1,1)}. \tag{4.13}$$

The expressions for ω^* and ω^{**} imply that $\omega^* + b + c = \frac{c}{qQ_1}$ and $\omega^{**} + c = \frac{c}{qQ_0}$. Therefore,

$$\frac{\omega^* + b + c}{\omega^{**} + c} = \frac{Q_0}{Q_1} = \frac{(l^* + 2)\mathcal{I}(1,1)}{l^* + 2\mathcal{I}(1,1)}. \tag{4.14}$$

Since $\omega^{**} - \omega^* \in (0, b)$, the difference between $\omega^* + b + c$ and $\omega^{**} + c$ is at most b. Hence, the LHS of (4.14) converges to 1 as $\omega^*, \omega^{**} \to +\infty$. The RHS of (4.14) is strictly increasing in $\mathcal{I}$ and equals 1 only if $\mathcal{I}(1,1) = 1$. Hence $\mathcal{I}(1,1)$ converges to 1 as $\omega^*, \omega^{**} \to +\infty$.

The defendant's indifference condition (4.8) implies that as $L \to +\infty$, either $q \to 0$ or at least one of $1 - \Phi(\omega^{**})$ and $\Phi(\omega^{**}) - \Phi(\omega^*)$ converges to 0. If $q \to 0$, then the expressions for ω^* and ω^{**} in (4.9) and (4.10) imply that $\omega^*, \omega^{**} \to +\infty$. If $1 - \Phi(\omega^{**})$ converges to 0, then $\omega^{**} \to +\infty$ and $\omega^* \to +\infty$ as well since $|\omega^* - \omega^{**}| \in (0, b)$. If $\Phi(\omega^{**}) - \Phi(\omega^*)$ converges to 0, then either $\omega^{**} \to +\infty$ or $-\infty$ or $\omega^{**} - \omega^* \to 0$ or both. Since (4.10) implies that $\omega^{**} \geq 0$, it cannot be the case that $\omega^{**} \to -\infty$. Our earlier conclusion implies that the value of $\mathcal{I}(1,1)$ converges to 1 regardless of whether both ω^* and ω^{**} diverge to $+\infty$ or $\omega^{**} - \omega^* \to 0$ and both thresholds are bounded from above. According to Bayes rule, $\mathcal{I}(1,1) \to 1$ implies that $\pi \to \pi^*$.

4.3 Distinct Probabilities Principle

Next, we characterize equilibrium outcomes under DPP and compare them to those under APP.

Theorem 2. *For every $\pi^* \in (0,1)$ and $\varepsilon > 0$, there exists $\overline{L} \in \mathbb{R}_+$ such that when $L > \overline{L}$, in every equilibrium under DPP with standard of proof π^*,*

1. ***Uncorrelated Offenses:*** $\Pr(\theta_i = 1|\theta_j = 1) = \Pr(\theta_i = 1|\theta_j = 0)$ *for every* $i \neq j$.
2. ***Linear Conviction Probability:*** $q(1,1) = 2q(1,0) = 2q(0,1) > 0$ *and* $q(0,0) = 0$.
3. ***Effective Deterrence & Informative Accusations:*** $\mathbb{E}[\theta_i] < \varepsilon$ *and* $\frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)} > 1/\varepsilon$ *for every* $i \in \{1,2\}$.

Theorem 2 implies that, under DPP, agents' private observations of offenses must be *uncorrelated* and the conviction probability must be a *linear* function of the number of accusations. When the conviction probabilities are linear, the defendant's incentives to commit different offenses are neither complements nor substitutes, and deciding to commit each offense independently of one's decision regarding other offenses is an optimal strategy. In this case, agents' private observations of offenses are *uncorrelated*. Compared to APP, this lack of correlation restores the credibility of the agents' accusations, which has the further effect of lowering the expected number of offenses in equilibrium for any given standard of proof π^*.

The key distinctive feature of DPP is that accusing the defendant of committing an offense under DPP need *not* increase the probability that the defendant is guilty of any other offense, and thus need not affect or increase the maximal probability, over all possible offenses, that the defendant is guilty of a specific offense. If, for instance, θ_1 and θ_2 are negatively correlated, then an accusation by agent 2 can lower the posterior probability that the defendant is guilty of the first offense.

Next, we present the proof of Theorem 2, which leverages the previous observation and sheds light on how APP and DPP lead to different structures of the conviction probabilities as functions of agents' reports and on how these differences affect the defendant's and the agents' incentives.

Step 1: Ruling Out Correlations Suppose by way of contradiction that θ_1 and θ_2 are negatively correlated, i.e., $\Pr(\theta_i = 1|\theta_j = 1) < \Pr(\theta_i = 1|\theta_j = 0)$ for $i \neq j$. Then, $\Pr(\theta_1 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_1 = 1|\boldsymbol{a} = (1,0))$ because an accusation by agent 2 increases the posterior probability that $\theta_2 = 1$ and, when θ_1 and θ_2 are negatively correlated, his accusation lowers the probability that $\theta_1 = 1$. Similarly, one can show that $\Pr(\theta_2 = 1|\boldsymbol{a} = (1,1)) < \Pr(\theta_2 = 1|\boldsymbol{a} = (0,1))$. When conviction is decided according to DPP, the conviction probabilities must satisfy $q(1,0) \geq q(1,1)$ and $q(0,1) \geq q(1,1)$. For these probabilities to satisfy Refinements 1 and 2, it must be the case that

$$q(1,1) = q(1,0) = q(0,1) = 1 \text{ and } q(0,0) = 0. \tag{4.15}$$

According to Lemma 4.3, the defendant's incentives to commit different offenses are *strict complements*. Since Refinement 1 requires that the defendant choosing $(\theta_1, \theta_2) = (0,0)$ with positive probability, he has no incentive to choose $(\theta_1, \theta_2) \in \{(1,0),(0,1)\}$ since it is dominated by either $(\theta_1, \theta_2) = (0,0)$ or $(\theta_1, \theta_2) = (1,1)$. This contradicts our earlier hypothesis that θ_1 and θ_2 are negatively correlated.

Next, suppose by way of contradiction that θ_1 and θ_2 are *positively correlated*, i.e., $\Pr(\theta_i = 1|\theta_j = 1) > \Pr(\theta_i = 1|\theta_j = 0)$ for $i \neq j$. In this case, $\boldsymbol{a} = (1,1)$ is the unique maximizer of both $\Pr\left(\theta_1 = 1\middle|\boldsymbol{a}\right)$ and $\Pr\left(\theta_2 = 1\middle|\boldsymbol{a}\right)$. If the conviction rule is DPP, then it must be the case that

$$q(1,1) \geq \max\{q(1,0), q(0,1)\} \geq q(0,0) = 0, \tag{4.16}$$

and $q(1,1) > \max\{q(1,0), q(0,1)\}$ unless $q(1,1) = q(1,0) = q(0,1) = 1$. If $q(1,1) = 1$, this contradicts Lemma 4.2 when L is large enough. When $q(1,1) \in (0,1)$, we know that $\max_{i\in\{1,2\}}\{\Pr\left(\theta_i = 1\middle|\boldsymbol{a} = (1,1)\right)\} = \pi^*$, which implies that $\max_{i\in\{1,2\},\boldsymbol{a}\neq(1,1)}\{\Pr\left(\theta_i = 1\middle|\boldsymbol{a}\right)\} < \pi^*$. As a result, $q(1,0) = q(0,1) = 0$. Lemma 4.3 then implies that the defendant's decisions to commit different offenses are *strict substitutes*. This contradicts our earlier hypothesis that θ_1 and θ_2 are positively correlated.

Step 2: Linear Conviction Probabilities Since θ_1 and θ_2 must be *uncorrelated* and Refinement 1 implies that $(\theta_1, \theta_2) = (0, 0)$ occurs with positive probability, the defendant's incentives to commit offenses (i) cannot be strict substitutes since he will have no incentive to choose $(\theta_1, \theta_2) = (1, 1)$ and (ii) cannot be strict complements since he will have no incentive to choose $(\theta_1, \theta_2) = (1, 0)$ and $(\theta_1, \theta_2) = (0, 1)$. Lemma 4.3 then implies that $q(1, 1) = q(1, 0) + q(0, 1)$ and $q(0, 0) = 0$. Lemma 4.1 then implies that the defendant must be playing a completely mixed strategy. The defendant's indifference between $(\theta_1, \theta_2) = (1, 0)$ and $(\theta_1, \theta_2) = (0, 1)$ implies that $q(1, 0) = q(0, 1)$. Hence, the conviction probability is a linear function of the number of accusations.

Step 3: Informative Accusations & Effective Deterrence: The fact that θ_1 and θ_2 are uncorrelated implies that each agent's belief about whether the other agent has witnessed an offense is independent of his private observation of offenses. According to the formulas for agents' reporting cutoffs (4.9) and (4.10), the independence of θ_1 and θ_2 implies that $Q_{1,j} = Q_{0,j}$ for $j \in \{1, 2\}$. This together with (4.1) and (4.2) implies that $\omega_j^{**} - \omega_j^* = b$ for every $j \in \{1, 2\}$. Let $\Delta \equiv q(1, 1) - q(1, 0) = q(1, 1) - q(0, 1) = q(0, 1) = q(1, 0)$ denote the increase in the probability of conviction when there is an additional accusation. Since the defendant is indifferent between choosing $(\theta_1, \theta_2) = (1, 0)$ and $(\theta_1, \theta_2) = (0, 1)$, we have:

$$L\Delta = \Phi(\omega_1^* + b) - \Phi(\omega_1^*) = \Phi(\omega_2^* + b) - \Phi(\omega_2^*). \tag{4.17}$$

Since ω_1^*, ω_2^* are both strictly positive when L exceeds some cutoff, the value of $\Phi(\omega^* + b) - \Phi(\omega^*)$ is strictly decreasing in ω^*. The defendant's indifference condition then implies that $\omega_1^* = \omega_2^*$ and $\omega_1^{**} = \omega_2^{**}$. Hence, we will replace ω_i^* and ω_i^{**} with ω^* and ω^{**}, respectively. The defendant's indifference condition also implies that the cutoffs ω^* and ω^{**} diverge to $-\infty$ as L becomes large. Since θ_1 and θ_2 are uncorrelated, the cutoffs satisfy $\omega^{**} - \omega^* = b$. Hence, the informativeness of each agent's accusation, measured by $\frac{1-\Phi(\omega^*)}{1-\Phi(\omega^{**})}$, diverges to $+\infty$ as $L \to +\infty$. Since 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.

4.4 Discussion

The comparison between Theorems 1 and 2 provides a justification for the use of DPP in criminal justice systems. It also suggests a rationale for several aspects of the law such as the fact that 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 guilt 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 cases affects the *correlations* between crimes through the defendant's incentives, 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(offense against individual 1)	Pr(offense against individual 2)	Pr(offense against at least one)
Defendant 1	49.5 %	49.5 %	99 %
	Pr(offense against individual 3)	Pr(offense against individual 4)	Pr(offense against at least one)
Defendant 2	50 %	1 %	51 %

Under DPP, defendant 1 is acquitted whenever defendant 2 is acquitted. This is the case even though defendant 1 is *almost surely* guilty of at least one crime, and the probability that defendant 2 is guilty is significantly lower.

In practice, firms and organizations may lack the commitment required to implement DPP and yield to social pressure and aggregate the probabilities across different offenses. For example, firms may face more social pressure to fire a manager whose probabilities of abusing subordinates are given by the first row and political parties may have a stronger incentive to ostracize party members with bad reputations (e.g., individuals who are believed to have committed at least some offenses with high probability) even without a thorough investigation of specific evidence related to any given accusation.

5 Extensions

Section 5.1 extends our model in Section 2 to the case in which the defendant has a private type that determines his benefit from committing offenses. In Section 5.2, we extend the analysis to three or more offenses and agents. We establish a comparative statics result showing that each agent accuses the defendant with strictly higher probability as the number of agents increases. Section 5.3 contains several other extensions and variations of our model.

5.1 Uncertain Propensity to Commit Offenses

We have assumed until now that the defendant's benefit from committing offenses was common knowledge, and found that when the penalty from conviction is large relative to the benefit from committing offenses, the defendant never commit both offenses under APP. In practice, individuals have different propensities to commit offenses and anecdotal evidence suggests that some defendants are guilty of multiple offenses.

Motivated by these observations, we now extend our model in Section 2 to the case in which the defendant has a private *type* $t \in \{t^v, t^o\}$. Specifically, the defendant is either a virtuous type t^v whose benefit from committing offenses is zero (or, more generally, non-positive) or an opportunistic type t^o whose benefit from committing each offense is normalized to 1. Let π^o denote the prior probability that the defendant has an opportunistic type (the model in Section 2 corresponds to $\pi^o = 1$). We show that under APP, a defendant with an opportunistic type commits multiple offenses with strictly positive probability as long as $\pi^o < \pi^*$, i.e., as long as the virtuous type occurs with high enough probability. This novel feature notwithstanding, we show that the main insight from Theorem 1 that APP induces negative correlation across offenses still applies.

Proposition 2. *For every $\pi^* \in (0,1)$, $\pi^o \in (0,1)$, and $\varepsilon > 0$, there exists $\overline{L} > 0$ such that when $L > \overline{L}$, in every equilibrium under APP and standard of proof π^*,*

1. ***Endogenous Negative Correlation:*** $\Pr(\theta_i = 1 | \theta_j = 1) < \Pr(\theta_i = 1 | \theta_j = 0)$ *for every $i \neq j$.*
2. ***Uninformative Accusations & Ineffective Deterrence:*** $\Pr(\overline{\theta} = 1) > \min\{\pi^*, \pi^o\} - \varepsilon$ *and*

$$\max_{\boldsymbol{a} \in \{0,1\}^2} \mathcal{I}(\boldsymbol{a}) < \left\{ \frac{\pi^*}{1-\pi^*} \Big/ \frac{\min\{\pi^*, \pi^o\}}{1 - \min\{\pi^*, \pi^o\}} \right\} + \varepsilon. \tag{5.1}$$

3. ***Multiple Offenses with Positive Probability:*** $\Pr(\theta_1 = \theta_2 = 1 | t = t^o) > 0$ *if and only if $\pi^o < \pi^*$.*

The proof is in Appendix C. Proposition 2 shows that, even when the defendant has private information regarding his benefit from committing offenses, agents' private observations about offenses are negatively correlated as in our model of Section 2, the informativeness of agents' accusations is low as in that model, and the lack of informativeness still leads to a high probability that offenses take place.

In Proposition 2, the upper bound regarding the informativeness of agents' accusations and the lower bound regarding the probability of offenses are the same as those in Theorem 1 when $\pi^o \geq \pi^*$, i.e., when the probability of the opportunistic type exceeds the conviction threshold. However, these bounds are different when the opportunistic type occurs with probability less than π^*. In this case, we show that the defendant commits offenses with probability π^o, which is the highest possible probability of guilt given that the virtuous type never commits any offense. Given that, in every equilibrium, there exists $\boldsymbol{a} \in \{0,1\}^2$ such that the defendant is convicted with positive probability when the judge observes $\boldsymbol{a}$, the judge's posterior belief that at least one offense has taken place is no less than π^* after observing $\boldsymbol{a}$. This suggests a lower bound on the informativeness of agents' accusations:

$$\max_{\boldsymbol{a} \in \{0,1\}^2} \mathcal{I}(\boldsymbol{a}) \geq \mathcal{I}_{min} \equiv \frac{\pi^*}{1-\pi^*} \Big/ \frac{\pi^o}{1-\pi^o}. \tag{5.2}$$

Proposition 2 shows that this lower bound $\mathcal{I}_{min}$ is indeed attained in *all equilibria* when $\pi^o < \pi^*$.

The proof of Proposition 2 is similar to that of Theorem 1. We start the proof with the observations that the virtuous type never commits any offenses and that both Lemma 4.2 and Lemma 4.3 extend to this more general setting. These observations imply that the opportunistic type's decisions to commit distinct offenses are *strict substitutes* when l is large. As a result, type t^o cannot be indifferent between committing no offense and committing two offenses. That is to say, type t^o is either indifferent between committing no offense and committing only one offense or indifferent between committing only one offense and committing two offenses. This leads to two disjoint cases in Proposition 2, whose separation depends on whether the prior probability of the opportunistic type exceeds π^*:

1. When $\pi^o \geq \pi^*$, the opportunistic type is indifferent between committing no offense and one offense and the qualitative features of the equilibria remain the same as those characterized by Theorem 1.

2. When $\pi^o < \pi^*$ and L is large enough, the opportunistic type commits at least one offense for sure. Otherwise, by the same logic as in Theorem 1, agents' accusations would be arbitrarily uninformative as L becomes large. Since the prior probability of guilt is at most π^o which is strictly lower than π^*, the posterior probability of guilt is also strictly lower than π^* even when both agents accuse the defendant. In this case, the defendant is never convicted under APP, which leads to a contradiction. Hence, the opportunistic type must be indifferent between committing one offense and committing both offenses.

   Nevertheless, we show that 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. In this case, $Q_{0,j} \leq Q_{1,j}$, and the expressions for agents' reporting cutoffs (4.9) and (4.10) would imply that $\omega_j^{**} - \omega_j^* \geq b$. According to Lemma 4.2, $q(1,1)$ converges to 0 as $L \to +\infty$ and agents' reporting cutoffs both diverge to $+\infty$. As in the single-offense benchmark, the informativeness of agents' accusations would then diverge to $+\infty$. Since the opportunistic type commits at least one offense, the prior probability that the defendant commits at least one offense is π^o. Therefore, the posterior probability that the defendant committed an offense when both agents accuse the defendant would be strictly greater than π^*. This would contradict Lemma 4.2, which requires that the judge be indifferent between convicting and acquitting the defendant when both agents accuse him.

Proposition 3 generalizes Theorem 2 to the case in which π^o can take any value in $(0, 1)$. It shows that, under DPP, there is no correlation across different offenses and the conviction probability is a linear function of the number of accusations. These two features result in informative accusations and a low expected number/probability of offenses.

Proposition 3. *For every $\pi^o \in (0,1)$, $\pi^* \in (0,1)$, and $\varepsilon > 0$, there exists $\overline{L} \in \mathbb{R}_+$ such that when $L > \overline{L}$, in every equilibrium under DPP with standard of proof π^*,*

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

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

3. ***Effective Deterrence & Informative Accusations:*** $\mathbb{E}[\theta_i] < \varepsilon$ *and* $\frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)} > 1/\varepsilon$ *for every* $i \in \{1,2\}$.

The proof of Proposition 3 follows from that of Theorem 2 and is omitted in order to avoid repetition.

5.2 Three or More Potential Offenses and Witnesses

This section extends our main results to settings where the defendant can commit n offenses, i.e., choosing $(\theta_1, .., \theta_n) \in \{0,1\}^n$. For each potential offense, there is a distinct agent who observes whether this offense has occurred. The defendant's marginal benefit from committing an offense is still normalized to 1 and the defendant's loss if convicted is still $L > 0$. Agent's payoffs are given by the same formula as in the two-agent case. Given the standard of proof $\pi^* \in (0,1)$, the defendant is convicted under APP if $\Pr(\overline{\theta} = 1|\boldsymbol{a}) \geq \pi^*$ where $\overline{\theta} \equiv \max_{1 \leq i \leq n} \theta_i$, and is convicted under DPP if $\max_{1\leq i \leq n} \Pr(\theta_i = 1|\boldsymbol{a}) \geq \pi^*$.

We show that the takeaways from our main results extend to this multi-agent setting and establish a comparative statics result showing that under APP the informativeness of agents' accusations about the defendant's guilt decreases *even when all agents' reports are aggregated* in the number of agents. Moreover, we show that the unconditional probability that each agent accuses the defendant is *increasing* in the number of agents. We start by generalizing the measure of informativeness used in Theorem 1 to three or more agents. According to Bayes Rule, for every $\boldsymbol{a} \in \{0,1\}^n$,

$$\frac{\Pr(\boldsymbol{a}|\overline{\theta} = 1)}{\Pr(\boldsymbol{a}|\overline{\theta} = 0)} \cdot \frac{\Pr(\overline{\theta} = 1)}{1 - \Pr(\overline{\theta} = 1)} = \frac{\Pr(\overline{\theta} = 1|\boldsymbol{a})}{1 - \Pr(\overline{\theta} = 1|\boldsymbol{a})}. \tag{5.3}$$

Therefore, the ratio

$$\mathcal{I}(\boldsymbol{a}) \equiv \frac{\Pr(\boldsymbol{a}|\overline{\theta} = 1)}{\Pr(\boldsymbol{a}|\overline{\theta} = 0)} \tag{5.4}$$

measures the change in a judge's posterior belief about $\overline{\theta}$ after observing the agents' actions $\boldsymbol{a} \equiv (a_1, ..., a_n)$. For tractability, we again focus on symmetric equilibria that satisfy Refinements 1 and 2 (or, *equilibrium* for short). Proposition 4 generalizes the insights of Theorem 1 to the case of three or more agents.

Proposition 4. *For every $n \geq 2$, $\pi^* \in (0,1)$, and $\varepsilon > 0$, there exists $\overline{L}_{n,\varepsilon} > 0$ such that for every $L > \overline{L}_{n,\varepsilon}$, in every equilibrium under APP and standard of proof π^*,*

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

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

The proof is in Online Appendix B.1. According to Proposition 4, an agent who has witnessed an offense assigns a lower probability to other agents having witnessed offenses. In fact, we show that in every equilibrium, the defendant commits either no offense or only one offense, which implies that agents' private observations of offenses are negatively correlated. We also show that, as the punishment for conviction becomes large, agents' accusations become arbitrarily uninformative and the expected number of offenses is at least π^*.

Next, we establish a comparative statics result with respect to the number of agents, showing that each agent is *more* likely to accuse the defendant as the number of agents increases. Recall that in every equilibrium, each agent i uses a strategy that is characterized by two cutoffs ω^* and ω^{**}, such that when he has witnessed an offense, he accuses the defendant if and only if $\omega_i > \omega^*$, and when has not witnessed an offense, he accuses the defendant if and only if $\omega_i > \omega^{**}$, i.e., lower cutoffs imply higher probabilities of filing accusations.

Theorem 3. *For every* $\pi^* \in (0,1)$ *and* $k, n \in \mathbb{N}$ *with* $k > n$, *there exists* $\overline{L} > 0$ *such that for every* $L > \overline{L}$, *and compare any equilibrium under APP and standard of proof* π^* *with* k *agents to that with* n *agents:*

1. ***Lower Informativeness:*** $\max_{\boldsymbol{a}\in\{0,1\}^k} \mathcal{I}(\boldsymbol{a}) < \max_{\boldsymbol{a}\in\{0,1\}^n} \mathcal{I}(\boldsymbol{a})$.
2. ***Higher Number/Probability of Offenses:*** *The equilibrium probability of offense* $\Pr(\overline{\theta}=1)$ *as well as the expected number of offenses* $\sum_i \Pr(\theta_i = 1)$ *are both strictly higher with* k *agents than with* n *agents.*
3. ***Higher Probability of Filing Accusations:*** *Each agent's reporting cutoffs* (ω^*, ω^{**}) *are both strictly lower with* k *agents than with* n *agents.*

The proof is in Appendix D. Theorem 3 shows that as the number of potential offenses (or, equivalently, the number of agents) increases, the informativeness of accusations decreases, the probability of offense increases, and each agent is *more likely* to accuse the defendant. This last feature distinguishes our result from those on public good provision, in which inefficiencies arise because agents free ride on other agents' contributions.

Next, we study the game's equilibrium outcomes when the judge uses DPP. Proposition 5 establishes the existence of equilibria in which agents' private information is uncorrelated and the conviction probability is a linear function of the number of accusations against the defendant.[12] These equilibria feature arbitrarily informative reports and a vanishing probability of offenses as the punishment L becomes arbitrarily large.

Proposition 5. *For every* $n \geq 3$, $\pi^* \in (0,1)$, *and* $\varepsilon > 0$, *there exists* $\overline{L} \in \mathbb{R}_+$ *such that when* $L > \overline{L}$, *there exists an equilibrium under DPP with standard of proof* π^* *such that* θ_i *is independent of* θ_{-i} *for every* i, *the probability of conviction is a linear function in the number of accusations and*

[12] Whether this equilibrium is the unique equilibrium remains an open question. This is because when there are three or more θ_i, one cannot use the argument in the proof of Theorem 2 that rules out positive and negative correlation between θ_1 and θ_2.

- ***Effective Deterrence:*** $\mathbb{E}[\sum_{i=1}^{n} \theta_i] < \varepsilon$.

- ***Informative Accusations:*** $\frac{\Pr(a_i=1|\theta_i=1)}{\Pr(a_i=1|\theta_i=0)} > 1/\varepsilon$ *for every* $i \in \{1, 2, ..., n\}$.

The proof is in Online Appendix B.2.

5.3 Other Extensions

Decreasing Marginal Benefits from Committing Offenses: Our main result, Theorem 1, continues to hold when the defendant faces decreasing marginal returns from committing multiple offenses or receives a punishment larger than L when the probability that he has committed multiple offenses exceeds some threshold. Indeed, these changes motivate the defendant to commit fewer offenses and induce negative correlation in the agents' private observations. As in the model of Section 2, the agents' coordination motives undermine the informativeness of their accusations and lead to a high probability of offenses taking place. For example, our results extend when the defendant is convicted of a minor crime and receives punishment L if the probability with which he is guilty of at least one offense exceeds π^*, and is convicted of a felony and receives punishment $L'(> L)$ if the probability with which he is guilty of two offenses exceeds some other cutoff π^{**}. When L is large, in every equilibrium under APP and standard of proof π^*, the defendant commits at most one offense and the probability with which he does so approaches π^*.

Agents' Accusation Costs: Our results continue to hold when agents' costs of accusing the defendant are strictly positive even when the defendant is convicted, as long as accusation costs are strictly higher when the defendant is acquitted than when he is convicted. Our results also continue to hold if each agent suffers a lower retaliation cost when there are more accusations filed against the defendant as long as the retaliation cost is strictly positive whenever the defendant is acquitted. Under these variations of our model, the coordination motives among agents become stronger and their private observations remain negatively correlated.

Interdependent Preferences: Agents may have a preference for convicting defendants who have committed offenses against or witnessed by other agents, even if they have not witnesses or experienced these offenses themselves. Our results extend to the general case where agent i's payoff is:

$$u_i(\omega_i, \theta_i, a_i) \equiv \begin{cases} \omega_i + bf(\theta_i, \theta_{-i}) & \text{if} \quad s = 1 \\ -ca_i & \text{if} \quad s = 0, \end{cases} \tag{5.5}$$

where $f(\theta_i, \theta_{-i})$ is strictly increasing in θ_i and is weakly increasing in θ_{-i}. The model of Section 2 considers the special case where $f(\theta_i, \theta_{-i}) = \theta_i$. Equation (5.5) stipulates that agent i has stronger preference for convicting defendants who have committed offenses, no matter whether the offense is witnessed by himself or by other agents. In fact, directly caring about offenses committed against or witnessed by other agents, may lower even further the informativeness of agents' accusations, by prompting agent i's action a_i to become more responsive to his belief about the other agent's private information θ_j, in effect magnifying agents' coordination motive. When θ_i and θ_j are negatively correlated, a_i will become less responsive to his own private information θ_i.

Agents' Preferences for Truth-telling: Suppose that 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. One can show that with two agents and when $d < \frac{l^*}{l^*+2}c$, the informativeness of each agent's accusation is at most

$$\frac{cl^*}{cl^* - (l^* + 2)d}, \tag{5.6}$$

which converges to 1 when $d \to 0$, i.e., our results are robust when agents receive small benefits from truth-telling.

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

Uncertainty about the Number of Potential Victims: In some applications, the number of potential victims/witnesses 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 defendant has opportunities to commit offenses against. We assume that only agents in $\widetilde{N}$ can accuse the defendant: 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 defendant observes $\widetilde{N}$. Agent i privately observes whether $i \in \widetilde{N}$ as well as (θ_i, ω_i).

We informally argue that the logic behind our results is preserved, and even 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 defendant is convicted or not depends only on the number of accusations, but not on the number of potential witnesses. Since the defendant is convicted with weakly higher probability when there are more accusations, he has a stronger incentive to commit offenses when $|\widetilde{N}|$ is smaller, i.e., fewer agents can accuse him. If an agent witnesses an offense, this agent 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 accuse the defendant when he has witnessed an offense which lowers the informativeness of accusations in equilibrium.

6 Concluding Remarks

Motivated by the debate among legal scholars regarding whether judges should aggregate the probabilities of distinct offenses when making conviction decisions, we present a game-theoretic model in which potential offenders' incentives to commit offenses and witnesses' incentives to report offenses, or the absence thereof, are both endogenous. We show that aggregating the probabilities of distinct offenses into a single overall probability of guilt, on which the conviction is based, can induce negative correlation in the private observations of distinct witnesses. This negative correlation, together with witnesses' coordination motive, lowers the informativeness of accusations and leads to a high number of offenses taking place. By contrast, the distinct probabilities principle commonly used in criminal justice systems across the world can restore the credibility of witness testimonies and lower the probability of offenses.

Our conclusions are obtained under the standard assumption that players have correct expectations about other players' strategies, and under the further assumption that the punishment received by the defendant in case of a conviction is large relative to defendants' benefit from committing an offense. Hence, our results are best suited for settings in which (i) a conviction rule has been in place for a long time and players in the game (defendants, witnesses, and judges) know each other's equilibrium strategies and (ii) conviction is effective for deterrence. Our results are less suitable to analyze behavior following a sudden change in the conviction rule or in the model's parameters, after which the assumption that players' behaviors form an equilibrium is less plausible. The results are also less suitable when the benefit from committing an offense is significant relative to the punishment associated with a conviction. In this case, conviction plays a less central role than in our arguments and the comparison between different conviction rules will like require a different analysis.

A Proof of Proposition 1

First, we show that under Refinement 1, $\Pr(\theta_1 = 1) \in (0, 1)$ and $\Pr(a_1 = 1) \in (0, 1)$. Refinement 1 requires that $q(0, 0) = 0$ and $\Pr((a_1, a_2) = (0, 0)) > 0$, which implies that $\Pr(\theta_1 = 1), \Pr(a_1 = 1) < 1$. Suppose by way of contradiction that $\Pr(\theta_1 = 1) = 0$. Then, the defendant is never convicted on the equilibrium path. If

$a_1 = 1$ occurs with positive probability, then $q(1,0) = q(1,1) = 0$ and the defendant has a strict incentive to choose $a_1 = 1$, which leads to a contradiction. If $a_1 = 1$ occurs with zero probability, then agent 1 always chooses $a_1 = 0$ when $\theta_1 = 0$. Hence, the defendant has an incentive to choose $\theta_1 = 0$ only if there exists $a_2 \in \{0,1\}$ that occurs with positive probability such that $q(1,a_2) > 0$. However, this implies that when $\theta_1 = 0$, a type of agent 1 with a sufficiently high ω_1 has a strict incentive to choose $a_1 = 1$, which leads to a contradiction. Given that $\Pr(\theta_1 = 1) \in (0,1)$, suppose by way of contradiction that $\Pr(a_1 = 1) = 0$. Then, agent 1 never chooses $a_1 = 1$ regardless of the realization of θ_1. This provides the defendant a strict incentive to choose $\theta_1 = 1$, which leads to a contradiction.

Next, we show that when $L > (\Phi(0))^{-1}$, either $q(0,1) = 0$ or agent 2 never chooses $a_2 = 1$. Suppose by way of contradiction that $q(0,1) > 0$ and $a_2 = 1$ occurs with positive probability. Let $\pi(a_1,a_2)$ denote the posterior probability that $\overline{\theta} = 1$ after observing (a_1,a_2). We have $\pi(0,0) = \pi(0,1)$ and $\pi(1,0) = \pi(1,1)$. Since $q(0,0) = 0$ and $q(0,1) > 0$, we know that $\pi(0,0) = \pi^*$. If $\pi(1,0) \le \pi^*$, then agent 1 accuses the defendant with weakly lower probability when $\theta_1 = 1$, which gives the defendant a strict incentive to choose $\theta_1 = 1$ and leads to a contradiction. Hence, $\pi(1,0) = \pi(1,1) > \pi^*$ and, therefore, $q(1,0) = q(1,1) = 1$. Agent 2 has an incentive to choose $a_2 = 1$ if

$$\omega_2 \ge \frac{c(1-q(0,1))}{q(0,1)} \ge 0$$

which implies that agent 2 chooses $a_2 = 0$ with probability at least $\Phi(0)$. Since $q(0,0) = 0$ and $q(1,0) = 1$, the defendant's expected cost of choosing $a_1 = 1$ is at least $L\Phi(0)$, which exceeds his benefit from committing offenses as long as $L > (\Phi(0))^{-1}$. This leads to a contradiction for L above this threshold.

Let q denote the expected probability of conviction when $a_1 = 1$. We show that $q \in (0,1)$. Suppose by way of contradiction that $q = 0$. Then, the defendant has a strict incentive to commit offenses, which leads to a contradiction. Suppose now by way of contradiction that $q = 1$. Then, the defendant is accused with probability $1 - \Phi(-b)$ when he commits an offense and with probability $1 - \Phi(0)$ when he does not commit an offense. Hence, he has an incentive to commit an offense only if $L(\Phi(0) - \Phi(-b)) \le 1$, which cannot be true when $L > \frac{1}{\Phi(0)-\Phi(-b)}$.

Hence, when L is large enough, it must be the case that $q \in (0,1)$ in every equilibrium. This implies that the defendant's posterior probability of guilt equals π^* conditional on $a = 1$. Since the probability that he commits an offense is strictly between 0 and 1, we have the following indifference condition:

$$\frac{1}{L} = \Phi\Big(\frac{(1-q)c}{q}\Big) - \Phi\Big(\frac{(1-q)c}{q} - b\Big).$$

When L is large enough, the conviction probability q becomes arbitrarily small, which implies that

$$\mathcal{I} = \frac{1 - \Phi\left(\frac{(1-q)c}{q} - b\right)}{1 - \Phi\left(\frac{(1-q)c}{q}\right)}$$

diverges to infinity. Bayes rule then implies that the probability that the defendant chooses $\theta_1 = 1$ converges to 0.

B Proof of Lemma 4.1

Step 1: We show that for every $i \in \{1, 2\}$, $\Pr(\theta_i = 1) \in (0, 1)$ in every equilibrium. First, suppose by way of contradiction that $\Pr(\theta_i = 1) = 1$ for some $i \in \{1, 2\}$. Then, regardless of the conviction rule used, the defendant is convicted with probability 1 under every $\boldsymbol{a}$ that occurs with positive probability. This cannot happen in any equilibrium that satisfies Refinement 1 since $q(0, 0) = 0$ and $\boldsymbol{a} = (0, 0)$ with positive probability.

Next, suppose by way of contradiction that $\Pr(\theta_i = 1) = 1$. By symmetry, it must be the case that $(\theta_1, \theta_2) = (0, 0)$ with probability 1. In this case, the defendant is acquitted with probability 1 under every $\boldsymbol{a}$ that occurs with positive probability. In order for the defendant to have an incentive to choose $\boldsymbol{\theta} = (0, 0)$, there exists $\boldsymbol{a} \in \{0, 1\}^2$ such that $q(\boldsymbol{a}) > 0$. Consider two cases. First, suppose $q(1, 0) > 0$. By Refinement 1, $a_2 = 0$ with positive probability, which implies that agent 1 has a strict incentive to choose $a_1 = 1$ when ω_1 is sufficiently high. Hence $(a_1, a_2) = (1, 0)$ occurs with positive probability yet $q(1, 0) > 0$, which leads to a contradiction. A symmetric argument rules out $q(0, 1) > 0$. Next, suppose $q(1, 1) > 0$ and $q(1, 0) = q(0, 1) = 0$. If $a_2 = 1$ occurs with zero probability, then the defendant has a strict incentive to choose $\theta_1 = 1$ given that $q(1, 0) = q(0, 0) = 0$, which leads to a contradiction. If $a_2 = 1$ occurs with positive probability, then agent 1 has a strict incentive to choose $a_1 = 1$ when ω_1 is large enough, under which $(a_1, a_2) = (1, 1)$ occurs with positive probability. This contradicts the requirement that $q(\boldsymbol{a}) = 0$ for every $\boldsymbol{a}$ that occurs with positive probability.

Step 2: We show that for every $i \in \{1, 2\}$, agent i accuses the defendant with strictly higher probability when $\theta_i = 1$ compared to when $\theta_i = 0$. To see this, note that if agent i accuses the defendant with weakly higher probability when $\theta_i = 0$, the defendant has a strict incentive to choose $\theta_i = 1$ since the benefit from doing so is 1 and, according to Refinement 2, reducing the probability that $a_i = 1$ weakly decreases the probability of conviction. This contradicts the result obtained in the first step that $\Pr(\theta_i = 1) \in (0, 1)$ for every $i \in \{1, 2\}$.

Step 3: We rule out cases in which either $Q_{1,2}(q(1, 1) - q(0, 1)) + (1 - Q_{1,2})q(1, 0)$ or $Q_{0,2}(q(1, 1) - q(0, 1)) + (1 - Q_{0,2})q(1, 0)$ is 0, which implies that agent 1's equilibrium strategy takes the form of two cutoffs.

Suppose by way of contradiction that $Q_{1,2}(q(1,1) - q(0,1)) + (1 - Q_{1,2})q(1,0) = 0$, in which case the LHS of (4.1) is 0. Refinement 2 implies that the RHS of (4.1) is weakly positive. If it is strictly positive, then agent 1 never chooses $a_1 = 1$ when $\theta_1 = 1$. As a result, the defendant has a strict incentive to choose $\theta_1 = 1$, which leads to a contradiction. If the RHS of (4.1) is 0, then $Q_{1,2} = 1$ and $q(0,1) = 1$, i.e., conditional on $\theta_1 = 1$, agent 2 chooses $a_2 = 1$ with probability 1. We consider two cases separately:

1. Conditional on $\theta_1 = 1$, $\theta_2 = 0$ with positive probability, and agent 2 plays $a_2 = 1$ when $\theta_2 = 0$.

2. Conditional on $\theta_1 = 1$, $\theta_2 = 0$ with zero probability, and agent 2 plays $a_2 = 1$ when $\theta_2 = 1$.

The first case is ruled out since it would give the defendant a strict incentive to choose $\theta_2 = 1$, given that doing so weakly decreases the probability of conviction by decreasing the probability that $a_2 = 1$. In the second case, the symmetry requirement implies that only $(\theta_1, \theta_2) = (0, 0)$ and $(\theta_1, \theta_2) = (1, 1)$ occur with strictly positive probability. This implies that $Q_{1,1} = 1$ and $q(1, 0) = 1$. Plugging $q(1,1) = q(1,0) = q(0,1) = 1$ into agent 1's incentive constraint when $\theta_1 = 0$, we know that he will accuse the defendant if $(1 - Q_{0,2})\omega_1 \geq 0$, and similarly, we know that when $\theta_2 = 0$, he will accuse the defendant if $(1 - Q_{0,1})\omega_2 \geq 0$. If $Q_{0,i} = 1$, the defendant has a strict incentive to choose $\theta_i = 1$, which leads to a contradiction. If $Q_{0,1}, Q_{0,2} < 1$, then when $(\theta_1, \theta_2) = (0, 0)$, agent i accuses the defendant if $\omega_i \geq 0$, which occurs with probability $1 - \Phi(0)$. When L is large enough so that $L\Phi(0)^2 > 2$, the defendant strictly prefers $(\theta_1, \theta_2) = (0, 0)$ to $(\theta_1, \theta_2) = (1, 1)$, which leads to a contradiction.

Suppose by way of contradiction that $Q_{0,2}(q(1,1) - q(0,1)) + (1 - Q_{0,2})q(1,0) = 0$. Refinement 2 implies that $Q_{0,2}(q(1,1) - q(0,1)) = 0$ and $(1 - Q_{0,2})q(1,0) = 0$. That is, either $q(1,1) = q(0,1)$ or $Q_{0,2} = 0$, and either $q(1,0) = 0$ or $Q_{0,2} = 1$. We consider three cases separately. First, if $q(1,1) = q(0,1)$ and $q(1,0) = 0$, then the defendant has a strict incentive to choose $\theta_1 = 1$, which leads to a contradiction. Second, if $q(1,1) = q(0,1)$ and $Q_{0,2} = 1$, i.e., $\theta_1 = 0$ implies that $a_2 = 1$. Refinement 1 implies that $\theta_2 = 0$ occurs with positive probability conditional on $\theta_1 = 0$. However, in this case agent 2 chooses $a_2 = 1$ with probability 1 when $\theta_2 = 0$, which implies that the defendant has a strict incentive to choose $\theta_2 = 1$, leading to a contradiction. Third, if $q(1,0) = 0$ and $Q_{0,2} = 0$, agent 1 believes that agent 2 chooses $a_2 = 0$ for sure after observing $\theta_1 = 0$. If $\theta_2 = 1$ occurs with positive probability when $\theta_1 = 0$, then agent 2 chooses $a_2 = 0$ for sure when $\theta_2 = 1$. In this case, the defendant has a strict incentive to choose $\theta_2 = 1$, which leads to a contradiction. If $\theta_2 = 0$ occurs with probability 1 when $\theta_1 = 0$, then agent 2 chooses $a_2 = 0$ for sure when $\theta_2 = 0$. This together with $q(1,0) = 0$ implies that the defendant strictly prefers $(\theta_1, \theta_2) = (1, 0)$ to $(\theta_1, \theta_2) = (0, 0)$, which leads to a contradiction.

C Proof of Proposition 3

Lemma 4.1, 4.2, and 4.3 apply to this more general setting in which the virtual type occurs with strictly positive probability. We consider two cases separately depending on which of π^o and π^* has the higher value. The case in which $\pi^o \geq \pi^*$ is similar to the model in Section 2, while the case in which $\pi^o < \pi^*$ requires a different argument.

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

Lemma 4.2 implies that $q(0,0) = q(1,0) = q(0,1) = 0$ and $q(1,1) > 0$. Therefore, $q(1,1) + q(0,0) - q(1,0) - q(0,1) > 0$ and the definition of APP implies that $\Pr(\overline{\theta} = 1|\boldsymbol{a}) \leq \pi^*$ for every $\boldsymbol{a} \in \{0,1\}^2$. Therefore, $\Pr(\overline{\theta} = 1) < \pi^* \leq \pi^o$, which implies that the opportunistic type of the defendant chooses $\boldsymbol{\theta} = (0,0)$ with positive probability. From Lemma 4.3, the defendant's choices of committing distinct offenses are strategic substitutes, and as argued in the proof of Theorem 1, the opportunistic type cannot be indifferent between committing no offense and committing two offenses. This implies that the defendant chooses $\boldsymbol{\theta} = (1,1)$ with zero probability.

Let π denote the ex ante probability that the defendant commits at least one offense. Our symmetry requirement implies that (θ_1, θ_2) equals $(1,0)$ and $(0,1)$ with the same probability. Conditional on $\theta_i = 0$, the probability that $\boldsymbol{\theta} = (0,0)$ is

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

Let $Q_1 \equiv Q_{1,1} = Q_{1,2}$ and $Q_0 \equiv Q_{0,1} = Q_{0,2}$. Since $\boldsymbol{\theta} = (1,1)$ occurs with zero probability,

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

and

$$Q_0 = 1 - \beta\Phi(\omega^{**}) - (1-\beta)\Phi(\omega^*) = \beta Q_1 + (1-\beta)(1-\Phi(\omega^*)). \tag{C.3}$$

Subtracting (4.10) from (4.9) yields

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

Lemma C.1. $\omega^{**} - \omega^* \in (0, b)$.

Proof of Lemma C.1: 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 $\omega^* \leq \omega^{**} - b < \omega^{**}$. The comparison between (4.9) and (4.10) then yields $Q_0 > Q_1$, which leads to a contradiction. Since $Q_0 > Q_1$, the term $\frac{-1+Q_0/Q_1}{Q_0}$ is strictly positive, which shows that $\omega^{**} - \omega^* < b$. $\square$

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 of the defendant mixes between $(0,1)$, $(1,0)$, and $(0,0)$, we have

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

Since $q(1,1) \in (0,1)$, the judge assigns probability π^* to $\overline{\theta} = 1$ after observing $(a_1, a_2) = (1,1)$. This implies that

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

Plugging (C.6) into (C.1), we obtain 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) then yields

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

Plugging (4.9) and (4.10) 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.2. *If $\omega^* \to -\infty$, then $\mathcal{I} \to 1$ and $\pi \to \pi^*$.*

Proof of Lemma C.2: 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 $\mathcal{I}$ and is equal to 1 when $\mathcal{I} = 1$, $\mathcal{I}$ must converge to 1 as $\omega^* \to -\infty$. Equation (C.6) then shows that π converges to π^*. □

The opportunistic type's indifference condition (4.8) implies that as $L \to +\infty$, either $q \to 0$ or at least one of $1 - \Phi(\omega^{**})$ and $\Phi(\omega^{**}) - \Phi(\omega^*)$ converges to 0. If $q \to 0$, then the expressions for ω^* and ω^{**} in (4.9) and (4.10) imply that $\omega^*, \omega^{**} \to +\infty$. If $1 - \Phi(\omega^{**})$ converges to 0, then $\omega^{**} \to +\infty$ and $\omega^* \to +\infty$ as well since $|\omega^* - \omega^{**}| \in (0, b)$. If $\Phi(\omega^{**}) - \Phi(\omega^*)$ converges to 0, then either $\omega^{**} \to +\infty$ or $-\infty$ or $\omega^{**} - \omega^* \to 0$ or both. Since (4.10) implies that $\omega^{**} \geq 0$, it cannot be the case that $\omega^{**} \to -\infty$. Our earlier conclusion implies that the value of $\mathcal{I}(1,1)$ converges to 1 regardless of whether both ω^* and ω^{**} diverge to $+\infty$ or $\omega^{**} - \omega^* \to 0$ and both thresholds are bounded from above by some finite number. According to Bayes rule, $\mathcal{I}(1,1) \to 1$ implies that $\pi \to \pi^*$.

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

First, we show that the opportunistic type commits *two* offenses with positive probability in every equilibrium. Suppose by way of contradiction that he never commits two offenses. Since type t^v defendant never commits any offense, the equilibrium probability of that an offense occurs cannot exceed π^o, which is strictly less than π^*. As a result,

$$\mathcal{I} \equiv \frac{\Pr(a_1 = a_2 = 1|\overline{\theta} = 1)}{\Pr(a_1 = a_2 = 1|\overline{\theta} = 0)} \geq \frac{\pi^o}{1-\pi^o} \Big/ \frac{\pi^*}{1-\pi^*} > 1. \tag{C.10}$$

The expressions for Q_0 and Q_1 in (C.2) and (C.3) also apply to this setting. The derivations contained 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.10), which requires that the informativeness ratio be strictly bounded below, 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 4.3 implies that the opportunistic type cannot be indifferent between committing no offense and committing two offenses. Hence, it can only be the case that the opportunistic type chooses $\boldsymbol{\theta} = (0,1), (1,0)$ and $(1,1)$ with strictly positive probability, and chooses $\boldsymbol{\theta} = (0,0)$ with zero probability. This is the case when the expected marginal cost of committing a second offense equals 1, which implies that the expected marginal cost of committing the first offense is strictly less than 1. This implies that $\Pr(\overline{\theta} = 1) = \pi^o$. The informativeness ratio is pinned down by Bayes Rule:

$$\mathcal{I}\frac{\Pr(\overline{\theta} = 1)}{1-\Pr(\overline{\theta} = 1)} = \frac{\Pr(\overline{\theta} = 1|a_1 = a_2 = 1)}{1-\Pr(\overline{\theta} = 1|a_1 = a_2 = 1)}. \tag{C.11}$$

We now show that θ_1 and θ_2 are negative correlated, i.e., that $\Pr(\theta_j = 1|\theta_i = 1) < \Pr(\theta_j = 1|\theta_i = 0)$ for every $i \neq j$. Suppose by way of contradiction that there exist i and j with $i \neq j$ such that $\Pr(\theta_j = 1|\theta_i = 1) \geq \Pr(\theta_j = 1|\theta_i = 0)$ for every $i \neq j$ The definitions of $Q_{1,j}$ and $Q_{0,j}$ imply that $Q_{1,j} \geq Q_{0,j}$. The equations (4.9) and (4.10) for the cutoffs then imply that $\omega_i^{**} - \omega_i^* \geq b$. When L is large enough, we have $q(1,1) \to 0$ and $\omega_i^* \to +\infty$, and the analysis of the single-agent case shows that $\mathcal{I} \to \infty$. This contradicts (C.11) and the conclusion that $\Pr(\overline{\theta} = 1|a_1 = a_2 = 1) = \pi^*$.

D Proof of Theorem 3

Our proofs in this section use the following lemma, which is proved in Online Appendix B.

Lemma D.1. *Fix any $n \geq 2$ and suppose that the judge uses APP. There exists $\overline{L} > 0$ such that for every $L > \overline{L}$ and every symmetric equilibrium that satisfies Refinements 1 and 2,*

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

With $n \geq 3$ agents, we derive formulas for agents' reporting cutoffs $(\omega_n^*, \omega_n^{**})$, the informativeness of reports $\mathcal{I}_n$, and the equilibrium probability that $\overline{\theta} = 1$, i.e., that at least one offense occurs, which we denote by π_n. Agent i's reporting cutoffs are

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

and

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

where

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

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

In any symmetric equilibrium, the aggregate informativeness of reports, defined in (5.5), can be written as

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

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

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

When L is large enough, the defendant is indifferent between committing only one offense and committing no offense, which leads to the following indifference condition

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

Reporting Cutoffs & Distance Between Cutoffs: We show that $\omega_k^* < \omega_n^*$. Suppose by way of contradiction that $\omega_k^* \geq \omega_n^*$. From (D.1), we know that

$$q_k\Big(1 - \Phi(\omega_k^{**})\Big)^{k-1} \leq q_n\Big(1 - \Phi(\omega_n^{**})\Big)^{n-1}. \tag{D.7}$$

Therefore, $q_k Q_{1,k} \leq q_n Q_{1,n}$, which is equivalent to

$$q_k\Big(1-\Phi(\omega_k^{**})\Big)^{k-1}\Big(\Phi(\omega_n^{**})-\Phi(\omega_n^{*})\Big) \leq q_n\Big(1-\Phi(\omega_n^{**})\Big)^{n-1}\Big(\Phi(\omega_n^{**})-\Phi(\omega_n^{*})\Big)$$

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

This implies that

$$\Phi(\omega_n^{**})-\Phi(\omega_n^{*}) \leq \Phi(\omega_k^{**})-\Phi(\omega_k^{*}). \tag{D.8}$$

Under our hypothesis that $\omega_k^* \geq \omega_n^*$, inequality (D.8) is true only if

$$\omega_n^{**}-\omega_n^{*} \leq \omega_k^{**}-\omega_k^{*}, \tag{D.9}$$

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

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

Since we have shown that $q_k Q_{1,k} \leq q_n Q_{1,n}$ and $q_k Q_{0,k} \leq q_n Q_{0,n}$, (D.9) is true only if

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

We have

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

and

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

Combining this with (D.6) and (D.10) yields

$$\frac{(n-1)l^*}{(n-1)l^*\Big(1-\Phi(\omega_n^{**})\Big)+n\Big(1-\Phi(\omega_n^{*})\Big)} \geq \frac{(k-1)l^*}{(k-1)l^*\Big(1-\Phi(\omega_k^{**})\Big)+k\Big(1-\Phi(\omega_k^{*})\Big)},$$

which may be rewritten as

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

This inequality cannot hold because $1-\Phi(\omega_k^{**}) < 1-\Phi(\omega_n^{**})$, $1-\Phi(\omega_k^{*}) < 1-\Phi(\omega_n^{*})$ and $(n-1)k < (k-1)n$. This leads to a contradiction and shows that $\omega_k^* < \omega_n^*$ whenever $k > n$. The relation that $k > n$ was only used in the last step. Using the fact that $\omega_k^* < \omega_n^*$ and repeating our earlier argument up until (D.9), we obtain that

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

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

The Informativeness of Accusations and Expected Number of Offense: We show that $\mathcal{I}_n > \mathcal{I}_k$. The posterior probability that an offense was committed reaches π^* after observing $\boldsymbol{a} = (1, 1, ..., 1)$. This, together with (5.3), implies that the probability of offenses and the expected number of offenses must be ranked as claimed by Theorem 3.

Applying (D.1) and (D.2) to both n and k, we obtain

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

First, we show that

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

Suppose toward a contradiction that the RHS of (D.13) is at least as large as the LHS of (D.13). Then,

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

When L is large enough, the RHS of (D.14) is strictly greater than 1 since $0 < \omega_k^* < \omega_n^*$. This implies that the LHS of (D.14) is greater than 1, which is equivalent to

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

This contradicts (D.11), which was established earlier, and proves (D.13). This, together with (D.12), implies 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 $\mathcal{I}_k$ obtained in (D.5), we get

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

Letting $\Delta \equiv \mathcal{I}_k - \mathcal{I}_n$, the previous 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 $\Delta < 0$ or, equivalently, that $\mathcal{I}_n > \mathcal{I}_k$. Equation (5.3) then implies that the probability with which the defendant chooses $\overline{\theta} = 1$ increases when the number of agents increases from n to k. It also implies that the expected number of offenses increases since the defendant commits at most one offense on the equilibrium path.

References

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

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

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

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

[5] Ba, Bocar (2020) "Going the Extra Mile: The Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago," Working paper

[6] Ba, Bocar and Roman Rivera (2023) "The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago," *Review of Economics and Statistics*, forthcoming.

[7] Backes, Marvin (2020) "Epistemology and the Law: Why There Is No Epistemic Mileage in Legal Cases?" *Philosophical Studies*, 177, 2759-2778.

[8] Baliga, Sandeep, Ethan Bueno de Mesquita and Alexander Wolitzky (2020) "Deterrence with Imperfect Attribution," *American Political Science Review*, 114(4), 1155-1178.

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

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

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

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

[13] Blome-Tilman, Michael (2020) "Statistical Evidence, Normalcy, and the Gatecrasher Paradox," *Mind*, 129, 563-577.

[14] Chassang, Sylvain and Gerard Padró i Miquel (2019) "Crime, Intimidation, and Whistleblowing: A Theory of Inference from Unverifiable Reports," *Review of Economic Studies*, 86(6), 2530-2553.

[15] Cheng, Ing-Haw and Alice Hsiaw (2022) "Reporting Sexual Misconduct in the #MeToo Era," *American Economic Journal: Microeconomics*, 14(4), 761-803.

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

[17] Di Bello, Marcello (2019) "Trial by Statistics: Is a High Probability of Guilt Enough to Convict?" *Mind*, 128(512), 1045-1084.

[18] Ekmekci, Mehmet and Stephan Lauermann (2020) "Manipulated Electorates and Information Aggregation," *Review of Economic Studies*, 87(2), 997-1033.

[19] Ekmekci, Mehmet and Stephan Lauermann (2022) "Information Aggregation in Poisson Elections," *Theoretical Economics*, 17(1), 1-23.

[20] Feddersen, Timothy and Wolfgang Pesendorfer (1996) "The Swing Voter's Curse," *American Economic Review*, 86(3), 408-424.

[21] Gershkov, Alex and Balazs Szentes (2009) "Optimal Voting Schemes with Costly Information Acquisition," *Journal of Economic Theory*, 144, 36-68.

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

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

[24] Littlejohn, Clayton (2017) "Truth, Knowledge, and the Standard of Proof in Criminal Law," *Synthese*, 197, 5253-5286.

[25] Kreutzkamp, Sophie (2023) "Endogenous Information Acquisition in Cheap-Talk Games," Working Paper.

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

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

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

[29] Pei, Harry and Bruno Strulovici (2021) "Crime Aggregation, Deterrence, and Witness Credibility," SSRN 4901931.

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

[31] Pundik, Amit (2015) "Should Criminals be Convicted of Unspecified Offenses? On Efficiency, Condemnation, and Cognitive Psychology," *Criminal Law and Philosophy*, 9, 207-224.

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

[33] Ross, Lewis (2020) "Recent Work on the Proof Paradox," *Philosophy Compass*, 15(6), e12667.

[34] Schauer, Frederick (2023) "Statistical Evidence and the Problem of Specification," *Episteme*, 20, 367-376.

[35] Schmitz, Patrick and Thomas Tröger (2011) "The Sub-Optimality of the Majority Rule," *Games and Economic Behavior*, 74(2), 651-665.

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

[37] Smith, Martin (2018) "When Does Evidence Suffice for Conviction?" *Mind*, 127(508), 1193-1218.

[38] Strulovici, Bruno (2022) "Can Society Function without Ethical Agents? An Informational Perspective," Working Paper, Northwestern University.

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

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