Political Interventions in the Administration of Justice

Carlo M. Horz* Hannah K. Simpson†

Abstract

Politicians may often be tempted to direct the prosecution of their political opponents. We argue that the informational consequences of prosecution are an important determinant of such interference, because only when a prosecutor is free to follow the evidence can his decisions convey information to the public about an opponent's likely guilt—or innocence. We build a game-theoretic model to investigate the implications of this effect. We find that when public opinion is moderately against an incumbent, interference comes at an informational cost, by preventing the public from updating negatively about an opponent. By contrast, when public opinion moderately favors the incumbent, interference confers an informational benefit by preventing the release of potentially exonerating information. Moreover, an accurate court system may sometimes incentivize interference, because it allows citizens to learn even if the initial prosecution was tainted, and, by protecting the truly innocent, may decrease the prosecutor's concerns about wrongful convictions.

*Assistant Professor, Texas A&M University, Department of Political Science, College Station, TX 77843-4348 E-mail: carlo.horz@tamu.edu. Web: www.carlohorz.com.

†Assistant Professor, Texas A&M University, Department of Political Science, College Station, TX 77843-4348 *E-mail*: hannah.simpson@tamu.edu. *Web*: www.hannahksimpson.com.

Criminal prosecutions carry serious consequences: not only can conviction spell imprisonment (or worse), merely being the target of a criminal investigation can impose heavy reputational costs. As a consequence, politicians may be tempted to use prosecutions against their political opponents, and prevent their use against allies. Leaders of authoritarian and semi-democratic regimes often yield to this temptation: well-known examples include the Moscow Trials in Soviet Russia, the political trials in Brazil, Chile, and Argentina in the 1970s and 1980s (Pereira, 2008), and Ukrainian presidents' ongoing penchant for prosecuting their rivals for office.[1] But leaders of consolidated democracies are not immune. For example, Gordon (2009) shows that US district attorneys systematically pursue presidential party opponents more aggressively than copartisans, and the summary dismissal in 2006 of eight US district attorneys for insufficiently partisan prosecution decisions suggests that these prosecution choices may be, in some sense, directed by the executive.[2] Other recent examples include the 2019 SNC-Lavalin scandal in Canada,[3] claims that French prosecutors in the 2017 François Fillon embezzlement case were pressured to bring charges before the French elections,[4] and former Brooklyn District Attorney Charles Hynes' improper prosecution of allies' (and his own) political foes.[5]

When do leaders direct the prosecution of their political opponents—and when do they hold back? The answer is not entirely clear. There is a sizeable literature on politicians' strategic incentives to preserve or demolish high *court* independence (for a summary, see

[1]Kramer, Andrew. "Ukraine Is Threatening to Arrest Its Former President." *The New York Times.* February 28, 2020.

[2]See, e.g., Lichtblau, Eric, and Lipton, Eric. "E-Mail Reveals Rove's Key Role in '06 Dismissals." *New York Times.* August 11, 2009.

[3]"Secret tape increases pressure on Trudeau in SNC-Lavalin affair." *BBC News.* March 30, 2019.

[4]Wheeldon, Tom. "Macron orders inquiry after ex-prosecutor decries political 'pressure' in Fillon probe." *France 24.* June 20, 2020.

[5]See, e.g., Yakowitz, Will. "Hynes Slapped for Wrongful Prosecution of O'Hara. *New York Post.* October 15, 2009.

Helmke and Rosenbluth, 2009; Moustafa, 2014), but the single individual most critical to the course of a criminal prosecution is arguably not a judge, but a prosecutor (Tonry, 2012), and constraints on interference with high court judges often do not apply clearly to prosecutors. For example, strong institutional safeguards, such as lifetime tenure, are important in preserving judicial freedom (e.g., Melton and Ginsburg, 2014; Hayo and Voigt, 2007; Helmke and Rosenbluth, 2009), but prosecutors often enjoy fewer such protections, and may even serve at the will of the executive. Some scholars have argued that leaders may strengthen institutional safeguards for judges as an insurance policy against future loss of power (see, e.g., Ramseyer, 1994; Ginsburg, 2003; Hirschl, 2004; Finkel, 2008), but prosecutors have fewer safeguards to strengthen, and mere promises to refrain from interference if one's opponents do the same are unlikely to be credible, especially if interference improves the odds of remaining in power. The threat of backlash may force leaders to respect popular high court decisions (Vanberg, 2001; Stephenson, 2004; Staton, 2006), but prosecutors target individuals, not popular social policies—and recent work questions whether citizens punish popular leaders who erode even fundamental democratic protections (Graham and Svolik, 2019). There is a small political economy literature on prosecutors, but it focuses primarily on how different principals can affect prosecutorial decisions using the threat of replacement (e.g., Gordon and Huber, 2002; Shotts and Wiseman, 2010), without investigating the strategic incentives for leaders to manipulate prosecutors for political gain.

In this paper, we suggest another, previously overlooked factor that may affect leaders' decisions to intervene in criminal prosecutions: interventions have fundamental informational implications. In particular, an independent prosecutor's decision to act—or not act—against a political opponent can affect that opponent's public support, by providing information about the likelihood that he is guilty of wrongdoing. Political interference, if revealed, destroys this information.[6]

[6]The empirical frequency with which politicians charged with criminal activity argue that their prosecutions are politically motivated suggests that they, at least, believe prosecutions have negative implications for their careers which can be mitigated by the

To explore how these informational implications affect self-interested politicians' incentives to direct prosecutions, we construct a formal model of intervention in prosecution. Our baseline model features three players—an incumbent, a prosecutor, and a citizen—and a passive political opponent who is a potential target of prosecution and whose guilt is unknown. The incumbent cares about both inflicting some negative consequence (indictment, conviction) on the opponent and minimizing the opponent's public support, and chooses how much to intervene in prosecutorial decision-making. The prosecutor cares about both avoiding prosecution errors and obtaining any reward offered by the incumbent, and chooses whether to act against the opponent after receiving a signal of guilt or innocence. If the prosecutor acts, with some probability the justice system imposes a consequence on the target. To identify the effects of these informational incentives as cleanly as possible, in the baseline model, this probability is independent of the target's guilt; we dispense with this assumption later on. The consequence probability is accordingly interpretable in this model as the likelihood of initial arraignment/indictment, the likelihood of conviction in a completely uninformative court, or, in the special case where the probability of the consequence is 1, as conviction by a court that is fully aligned with the incumbent. Finally, the citizen supports the political opponent if her posterior belief in the opponent's guilt is sufficiently low, given her idiosyncratic preference for the opponent relative to the incumbent.

We model political intervention as an unspecified benefit the incumbent gives to a prosecutor who acts against the political opponent. This is formally equivalent to assuming the incumbent can impose a cost on a prosecutor who does not act. The benefit can be interpreted in a variety of ways: engaging in public pressure campaigns—Germany's former attorney general notes in a report on Council of Europe member states that "[w]hilst this method of influencing an ongoing criminal procedure is rather crude and easily detectable, it still occurs surprisingly often" (Leutheusser-Schnarrenburger, 2009, 36)—, conditioning continued employment or promotion on prosecution decisions, or simply offering bribes (as in Besley and Prat 2006). We assume that this benefit is costly

presence of interference.

to confer, due, for example, to institutional constraints or political risks, with the cost increasing in its size. Note that this type of benefit is targeted to a single prosecutorial action. In an extension, we consider a situation in which the incumbent simply installs a partisan who derives some utility from the citizen supporting the incumbent.[7] To focus on the informational consequences of intervention, we abstract away from citizen oversight of prosecutors (Gordon and Huber, 2002), principal's preferences over prosecutorial aggressiveness and accuracy (Shotts and Wiseman, 2010), citizen dislike of interference (Graham and Svolik, 2019) or more broadly, dishonesty (Dziuda and Howell, 2021), and the possibility that incumbents may intervene in prosecutions simply to signal strength (Huang, 2015).

Our main finding is that, when intervention's informational consequences are considered, politicians' incentives to intervene depend on current public opinion in sometimes surprising ways. Specifically, if the citizen's preferences for or against the political opponent are so extreme that no information would change her mind, the informational consequences of intervention are irrelevant, and the incumbent intervenes whenever doing so increases the probability of the consequence by enough to offset intervention's costs. But if the citizen's preferences are mutable, how they may be affected by an independent prosecutor's actions becomes critical to the incumbent's decision.

In particular, the incumbent has the highest incentives to intervene when the citizen currently prefers the incumbent, but would change her support to the opponent if he were revealed to be less corrupt than she currently believes. Here, intervention not only increases the probability of the consequence, but also confers an informational benefit: it keeps the prosecutor from potentially informing the citizen (through non-action) of the opponent's likely innocence. By contrast, the incumbent has the lowest incentives to intervene when the citizen currently prefers the opponent, but would reject him if he were revealed to be *more* corrupt than she currently believes he is. Here, intervention imposes

[7]Given that this portion of the utility function is independent of the state of the world (the guiltiness of the target), this definition of a partisan is similar to the conceptualization of an "ideologue" in Bueno De Mesquita and Friedenberg (2011).

an informational cost by preventing the citizen from ever learning of the opponent's likely guilt.[8]

Further investigation reveals more insights. We first consider the additional implications of partisanship in the prosecutor. We show that if the prosecutor is a partisan who wants the citizen to support the incumbent, political intervention becomes significantly cheaper. However, partisanship can also destroy the prosecutor's informational usefulness: a partisan prosecutor may act, regardless of signal, against political opponents even when the incumbent would prefer he didn't. (A similar dynamic would obtain if the prosecutor were simply hyper-aggressive.) By contrast, the cost of intervening is higher when the prosecutor is politically neutral, but the incumbent can better control the flow of information from the prosecutor to the citizen. This suggests a trade-off between different forms of intervention. Intervening by simply appointing partisan prosecutors may be desirable when the incumbent is moderately popular and merely wishes to obfuscate evidence of her opponents' innocence. However, maintaining neutral prosecutors may be a better choice when the incumbent is moderately unpopular and often needs the prosecutor to provide credible evidence of her opponents' guilt.

Second, we incorporate the presence of a court that decides whether to inflict the consequence based on its own informative signal about the opponent's guilt or innocence. We find that, in general, the court does not alter the dynamics of the relationship between public opinion and the incumbent's incentives to intervene, although it complicates them in a number of ways. For example, if the citizen currently prefers the opponent but would change her mind if the court condemned him, her preferences effectively depend only on the court. Intervention then confers an indirect informational benefit because it increases

[8]These results generate predictions that differ from the theories of judicial independence described above. For example, the insurance policy theory suggests incentives to intervene should be smallest when the citizen highly favors the political opponent, and largest when the citizen highly favors the incumbent. In our model, incentives to intervene are moderate under these conditions, with the highest and lowest incentives both emerging when the political environment is competitive.

the probability that the opponent comes before the court in the first place.

More generally, we show that the court's competence in distinguishing the guilty from the innocent has an ambiguous effect on the incumbent's incentives to interfere. Depending on the citizen's preferences and the likelihood that the political opponent is guilty, an increase in court accuracy may decrease *or increase* both the informational and the instrumental benefits of interference. At the same time, court accuracy *always* decreases the cost of intervention by reassuring the prosecutor that if he acts against an innocent opponent, the court can still acquit. Our finding that court accuracy can sometimes incentivize interference in prosecution contrasts with Gordon and Huber (2002), in which a fully informative court endogenously prevents the prosecutor from knowingly prosecuting innocents, by allowing citizens who care about prosecutor competence to condition on convictions. However, it is consistent with Gordon (2009)'s argument (and evidence) that under political interference, political opponents obtain lighter sentences because weaker cases are brought against them.

We further extend the informative court model to permit prosecutorial action either early or late in an incumbent's term. In this two-period model, if prosecution is begun late (in the second period), the citizen must make her support decision *before* the court reaches a decision. Assuming that if the second period is reached, the incumbent always intervenes, we find that moderately unpopular incumbents prefer to intervene early and gamble that the court will incriminate their opponent, while moderately popular incumbents prefer second period intervention since this ensures that exonerating information about the opponent cannot come to light before the election. This result suggests that there may be substantial heterogeneity in the circumstances surrounding early as opposed to late political prosecutions.

Finally, while in the main text interference is for simplicity observable, we show in the Appendix that our results continue to hold when interference is only revealed with positive probability—for example, if a prosecutor blows the whistle, or the media or some independent monitoring organization reports interference—so long as the probability of revelation exceeds some threshold. We also find that under most conditions, incorporating citizen dislike for interference does not alter our findings. And we consider a situation where the prosecutor must exert costly effort to obtain a signal of the opponent's guilt or innocence (as in Dewatripont and Tirole, 1999), and find that here, nonintervention is never optimal, because the incumbent can now fine-tune interference. In particular, she may now choose interference to either decrease or increase prosecutorial effort, in an attempt to jointly maximize the probability of conviction and of citizen support.

As well as adding to the small scholarship on prosecutors, our paper contributes to the existing scholarship on political interference in the justice system by suggesting that a new feature of political interference—its informational consequences—may be important in determining its occurrence. These informational consequences are relevant in shaping intervention decisions so long as public support is sufficiently important, suggesting that they may help explain variation in interference in both democracies and autocracies. They also provide a possible explanation for within-country variation in politicians' or parties' attitudes towards the prosecution of their political opponents. If a politician, or party, is somewhat unpopular, she may be unwilling to intervene in, or push for, prosecution of even apparently guilty political opponents, for fear intervention destroys the information conveyed by prosecution. By contrast, a moderately popular politician or party may openly intervene in prosecution decisions against political opponents, precisely in order to prevent prosecutors from signaling opponent innocence through non-action. Moreover, our findings suggest that intervention in prosecutions may be especially tempting in precisely those polities where intervention in judicial decision-making is difficult and the court system is highly competent.

Our work also complements recent theoretical (Dziuda and Howell, 2021; Gratton, Holden and Kolotilin, 2018) and empirical (Nyhan, 2015, 2017) work on the strategy behind political scandals. Like these papers, we argue that political actors may sometimes have incentives to damage the reputations of public figures. Unlike them, we focus on criminal prosecutions, which carry the additional benefit of potential criminal consequences. Finally, in addition to the substantive scholarship above, our paper relates methodologically the literature on information manipulation (e.g., Gehlbach and Sonin, 2014; Gehlbach and Simpser, 2015; Luo and Rozenas, 2018; Besley and Prat, 2006), and substantively to the constraints imposed by, and on, systems of justice (Dragu and Polborn, 2013; Fox and Stephenson, 2014; Turner, 2017; Hübert, 2019).

The Baseline Model

The actors are an incumbent Inc ("she"), a prosecutor P ("he"), and a citizen V ("she"). There is also a passive target ("he") whose type $\theta \in \{G, I\}$ determines his guilt or innocence. We assume the target is a political opponent of the incumbent's.[9]

The sequence of moves is as follows. The game begins when Nature chooses θ, with $\Pr(\theta = G) = p$. This choice is unobserved. The incumbent then offers the prosecutor $\lambda \in \mathbb{R}_+$ to take some action $a \in \{0, 1\}$ against the target, e.g., open an investigation, press charges. Next, the prosecutor receives a private signal of the target's guilt, $s \in \{g, i\}$ and decides whether or not to act. Denote the probability of a guilty signal given a guilty (innocent) target as $\Pr(g|G) = \gamma_G$ ($\Pr(g|I) = \gamma_I$). This probability represents the technology of information collection, including the competence of the prosecutor or the investigators that provide him with information, and any bias they have towards receiving a particular signal. We assume the signal is informative, such that $\gamma_G > \gamma_I$.

The prosecutor's action decision $a \in \{0, 1\}$ leads to a consequence $C \in \{0, 1\}$ with probability $\Pr(C = 1|a) = \psi a$. To more clearly explicate the main mechanisms at work, we assume in the baseline model that ψ is exogenous, representing general features of a justice system, such as its efficiency, or the pivotality of the prosecutor's action. This means that—for now—the probability of the consequence is independent of the target's actual guilt. The consequence is thus interpretable as arraignment/indictment in an independent court system, conviction in an ineffective court system, or, in the special case where $\psi = 1$, conviction in a fully corrupted court system.

Finally, the citizen exogenously supports the opponent ($r = 0$) or the incumbent ($r = 1$). Her choice depends on whether the opponent's likely guilt is high enough to

[9]In the Appendix, we solve the mirror game where the target is an ally.

outweigh any idiosyncratic preference for him (μ): $\Pr(G|\cdot) \geq \mu$.

The incumbent's payoff function is given by:

$$U_{\text{Inc}} = \alpha C + rB - K(\lambda).$$

Here, α and B represent the value she places, respectively, on obtaining the consequence C, and on gaining the citizen's support (or more generally, simply on preventing the citizen from supporting the opponent). B thus represents the extent to which citizen support for the opponent threatens the incumbent's hold on power. Lastly, K increasing in λ represents the costs of interference. We may imagine that the cost function is steeper where prosecutors enjoy more institutional protections, or where incumbents who interfere pay some sort of political price. (We do not include changes in citizen support in response to interference in this cost function, but consider this in the Appendix.)

The utility the prosecutor derives from his action choice is affected by his state-dependent preferences, or accuracy concerns, which are given in Table 1. Choosing the

$u_{C\theta}(q)$	Guilty	Innocent
$C = 1$	0	$-q$
$C = 0$	$-(1-q)$	0

Table 1: The Prosecutor's state-dependent preferences

action $a = 1$ also yields an additional benefit of λ to the prosecutor.[10] The prosecutor's payoff function is then:

$$U_P = u_{C\theta}(q) + a\lambda.$$

This payoff function (and the information obtained by the prosecutor) is very similar to the investigator's payoff function in Shotts and Wiseman (2010). There, the investigator also receives a noisy signal of guilt and is concerned by accuracy. However, in that model, the principal's goal is to get investigators to match *her* accuracy concerns (via the threat of replacement), and there is no citizen and thus no incentive for the principal

[10]This of course implies that the incumbent can credibly commit to providing λ if the prosecutor acts.

to manipulate information release.

Finally, the citizen supports the opponent if her posterior belief that the opponent is guilty is sufficiently low given her idiosyncratic preference for the opponent (μ), such that $\Pr(\theta = G|C, a) < \mu$ with $\mu \in (0, 1)$. The larger μ, the more overwhelming evidence of the opponent's guilt must be for the citizen to abandon him. When μ is intermediate, the citizen can be persuaded in one or another direction by an independent prosecutor's choice of action. We refer to the space in which μ is intermediate as *competitive*, because within this range, information affects the citizen's support choice.[11]

Discussion of Key Assumptions

Before moving to the analysis, a few comments are in order. First, in the main text, we assume perfect observability of interference. This is a standard assumption in Bayesian Persuasion games (Kamenica and Gentzkow, 2011), made in similar models of electoral manipulation (e.g., Gehlbach and Simpser, 2015) and media bias (Gehlbach and Sonin, 2014). Moreover, many forms of political interference (reappointments, firings, promotions, pressure campaigns) are in fact highly observable. However, we show in the Appendix that the model's main insights also hold when the citizen only learns about interference if it is revealed—for example, by a whistleblower prosecutor, a watchdog agency, or the media—so long as the probability of discovery is above some threshold.

Second, in the baseline model, we assume that, conditional on prosecutorial action, the probability that the consequence is inflicted, ψ, is independent of the opponent's true guilt or innocence, θ. We make this assumption primarily to establish the standalone informational value (or cost) of prosecutorial action. However, it is also substantively consistent with the legal and non-legal consequences that can accompany a prosecutor's investigation, arraignment, or indictment, long before a final determination of guilt is

[11] Notice that incorporating a citizen disutility from political interference or prosecutorial mistakes would not affect our results because when the citizen decides whom to support, these outcomes have already been realized, and so by sequential rationality, the citizen's calculus remains unchanged.

reached (e.g., loss of reputation or employment, monetary expenses), as well as with, in the special case $\psi = 1$, environments where the rest of the court system is fully aligned with the incumbent. We later examine how the prosecutor's informational value is affected by an informative court system, by extending the model to incorporate a consequence that is informative of guilt.

Third, in the main text, we assume that the citizen's preferences, μ, are unaffected by incumbent interference. We do this because we are interested in whether the informational costs to interference can motivate nonintervention even if citizens are indifferent to intervention. However, we show in the Appendix that the fundamental informational intuitions we develop in the paper do not change when μ is a function of interference. Instead, depending on the value of μ and the strength of intervention's effect on it, it sometimes introduces an additional cost to interference that may or may not overpower the informational dynamic.

Equilibrium

We solve the game by generalized backwards induction. Our solution concept is pure strategy Perfect Bayesian Equilibrium. In the Appendix, we consider mixed strategy equilibria. We begin with the prosecutor. Upon receiving signal $s \in \{g, i\}$, he updates his beliefs about the guilt θ of the political opponent. His posterior beliefs are:

$$\Pr(G|g) = \frac{p\gamma_G}{p\gamma_G + (1-p)\gamma_I} \quad \text{and} \quad \Pr(G|i) = \frac{p(1-\gamma_G)}{p(1-\gamma_G) + (1-p)(1-\gamma_I)} \tag{1}$$

Because we have assumed that the probability of observing the guilty signal g is higher when the opponent is guilty, the prosecutor's belief in the opponent's guilt increases (decreases) upon receiving the guilty (innocent) signal: $\Pr(G|i) < p < \Pr(G|g)$.

Given these posterior beliefs and the prosecutor's relative aversions to convicting the innocent (type I errors) and acquitting the guilty (type II errors), his expected utility from choosing action $a = 1$ is:

$$\Pr(G|s)\left[\psi 0 + (1-\psi)(-(1-q))\right] + (1-\Pr(G|s))\left[\psi(-q) + (1-\psi)0\right] + \lambda,$$

and his expected utility from choosing action $a = 0$ is:

$$\Pr(G|s)\left[-(1-q)\right].$$

Then the prosecutor chooses action $a = 1$ if and only if:

$$\lambda \geq \psi\left[q - \Pr(G|s)\right] \tag{2}$$

Absent interference (i.e., when $\lambda = 0$), we assume the prosecutor only acts against the target after receiving the guilty signal: $\Pr(G|i) < q < \Pr(G|g)$. Now consider how a positive level of interference ($\lambda > 0$) changes the prosecutor's behavior.

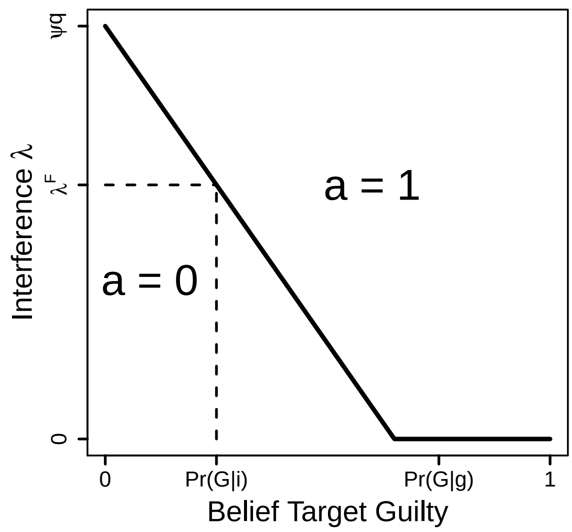

Figure 1: The prosecutor's strategy. The solid line indicates the point at which the prosecutor is indifferent between acting and not acting, for different levels of interference and posterior beliefs in the target's guilt. The dotted lines indicate the minimum level of interference needed to motivate action, given the prosecutor's posterior belief after observing the innocent signal. Parameter values: $p = 0.5$, $\gamma_G = 0.75$, $\gamma_I = 0.25$, $q = 0.65$, and $\psi = 0.2$.

Figure 1 displays how the prosecutor's optimal action changes, depending on his posterior belief that the opponent is guilty, and the amount of interference by the incumbent. As the figure shows, the prosecutor acts after observing a guilty signal whether the incumbent has interfered or not, but acts after observing an innocent signal only if interference is sufficiently high. Thus, incumbent interference induces one of two strategies. If interference is low, such that $\lambda < \lambda^F$, where

$$\lambda^F \equiv \psi \left[q - \Pr(G|i) \right], \tag{3}$$

the prosecutor follows his signal. If interference is high, such that $\lambda \geq \lambda^F$, the prosecutor acts regardless of his signal. Because interference is costly, in equilibrium the incumbent selects either $\lambda = 0$ (nonintervention) or $\lambda = \lambda^F$ (full intervention). If the incumbent chooses nonintervention, the citizen learns from the prosecutor's action: her posterior beliefs are the same as the prosecutor's (represented in Expression 1). By contrast, under full intervention, the citizen learns nothing, and her posterior belief is her prior, p.

Given the prosecutor's strategies and the citizen's beliefs, we can calculate the incumbent's utility for each intervention choice. The attractiveness of intervention to the incumbent depends on several factors. First, while intervention is directly costly, it also directly increases the likelihood of inflicting the consequence. Second, because intervention affects the citizen's posterior beliefs in the opponent's guilt, and the citizen supports the opponent only if her posterior on his guilt is sufficiently low relative to μ ($\Pr(G|a) < \mu$), intervention's attractiveness also depends on both its effect on the citizen's posterior and the size of her preference μ. To better illustrate the connection between incumbent intervention and the citizen's posterior beliefs, we redefine the citizen's posteriors as follows: $\Pr(G|g) \equiv p^N(1)$, $p \equiv p^F$, and $\Pr(G|i) \equiv p^N(0)$, where p^F stands for Full interference, and $p^N(a)$ represents No interference and prosecutorial action choice $a \in \{0, 1\}$. Then the incumbent's expected utility from nonintervention is

$$\Pr(s = g) \left[\psi\alpha + \mathbb{1}(p^N(1) \geq \mu) B \right] + \Pr(s = i) \mathbb{1} \left(p^N(0) \geq \mu \right) B,$$

and her expected utility from full intervention is:

$$\psi\alpha + \mathbb{1}(p^F \geq \mu)B - K\left(\lambda^F\right).$$

Consequently, the incumbent intervenes if and only if:

$$\overbrace{B \cdot \left[\mathbb{1}\left(p^F \geq \mu\right) - \Pr(s=g)\mathbb{1}(p^N(1) \geq \mu) - \Pr(s=i)\mathbb{1}(p^N(0) \geq \mu))\right]}^{\text{Informational cost or benefit of intervention (effect on citizen posterior)}}$$
$$\geq \tag{4}$$
$$\underbrace{K\left(\lambda^F\right) - \alpha\psi\left[1 - \Pr(s=g)\right]}_{\text{Direct cost net of increased probability of the consequence}}$$

The right-hand side of this inequality represents the costs of interference, net of the (constant) benefit the incumbent derives from increasing the probability of the consequence by forcing the prosecutor to act where otherwise he would not have acted. The left-hand side represents the effect of interference on the probability that the citizen supports the incumbent, and therefore naturally depends on μ. We take each case in turn.

Consider first the case of a highly partisan citizen. This is a citizen for whom μ is either very low ($\mu < p^N(0)$), such that the citizen prefers the incumbent even if the opponent is likely innocent, or very high ($\mu > p^N(1)$), such that the citizen prefers the opponent even if he is likely guilty. Here, information cannot change the citizen's mind. Consequently, interference does not affect the probability of gaining the citizen's support, and the incumbent intervenes if

$$0 \geq K\left(\lambda^F\right) - \alpha\psi\Pr(s=i)$$

Now, suppose that the citizen moderately dislikes the incumbent, $\mu \in \left(p^F, p^N(1)\right]$. Here, the citizen supports the opponent unless she learns new information about the opponent's guilt, and the incumbent interferes if:

$$T_{MO} \equiv -B\Pr(s=g) \geq K\left(\lambda^F\right) - \alpha\psi\Pr(s=i)$$

Lastly, suppose the citizen moderately likes the incumbent, $\mu \in \left(p^N(0), p^F\right]$. Now, the citizen supports the incumbent unless she learns new information about the opponent's *innocence*, i.e., unless an independent prosecutor does not act. The incumbent intervenes if:

$$T_{MI} \equiv B\Pr(s = i) \geq K\left(\lambda^F\right) - \alpha\psi\Pr(s = i)$$

Notice that in all three situations, the gains from intervention, T, are increasing in the prosecutor's pivotality ψ, the probability of seeing the innocent signal, $\Pr(s = i)$ (which in turn is decreasing in p, γ_G and γ_I), and the benefit the incumbent derives from the consequence α. Moreover, by inspection, we have the following result:

Proposition 1. *The attractiveness of intervention is strongest when the incumbent is moderately popular, and weakest when the incumbent is moderately unpopular,*

$$T_{MO} < 0 < T_{MI},$$

due to intervention's effect on the citizen's support choice. Furthermore, T_{MO} decreases in the benefit the incumbent derives from the citizen's support B while T_{MI} increases in B.

This result is presented graphically in Figure 2. The reason for the result is that, as shown by Equation 4, the attractiveness of intervention depends on three things: its costs, its effect on the probability of the consequence (and the benefit the incumbent derives therefrom), and its effect on citizen support (and the benefit the incumbent derives therefrom). Intervention is most attractive when the citizen has a moderate preference for the incumbent, because here it not only maximizes the likelihood of the consequence but also provides an informational benefit by precluding the possibility that the citizen learns of the opponent's likely innocence and changes her support choice. Intervention is least attractive when the citizen has a moderate preference for the opponent, because while it still maximizes the likelihood of the consequence, here it has an informational cost: it destroys the chance that the citizen will learn of the opponent's guilt and reject him. Moreover, the size of the benefit B derived from the citizen's support intensifies the

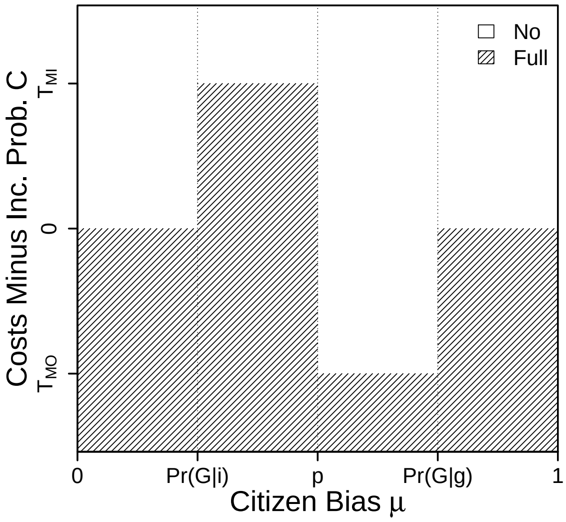

Figure 2: Prosecutorial independence/incumbent intervention under different levels of citizen bias on the x-axis and interference costs net of increased probability of the consequence on the y-axis. Parameter values: $p = 0.5$, $\gamma_G = 0.75$, $\gamma_I = 0.25$ $\alpha = 0.6$, and $\psi = 0.2$.

incumbent's incentive to interfere (to keep that support) in the first of these cases, and intensifies her incentives to refrain from interference (to gain that support) in the second. Finally, when the citizen is too partisan to be swayed by new information, intervention has neither an informational cost nor an informational benefit: its attractiveness lies between the two extremes described above, and is independent of B.

To better understand the result, consider a citizen who currently moderately prefers the incumbent to the opponent, $\mu \in (p^N(0), p^F]$. This ranking is because, from the citizen's perspective, the opponent is too likely to be guilty of wrongdoing to deserve support. If the incumbent pressures the prosecutor to act against the opponent, no new information about the opponent's guilt or innocence is revealed, and the citizen, learning nothing, continues to support the incumbent. If, however, the incumbent refrains from interfering, and the prosecutor does *not* act against the opponent, the citizen learns that the opponent is less blameworthy than she thought, and switches her allegiance to him. Thus, nonintervention not only renders the consequence less likely, it also carries a possible informational cost.

The reverse is true if the citizen moderately prefers the opponent, $\mu \in \left(p^F, p^N(1)\right]$.

Here, the citizen currently believes the opponent deserves support. As before, if the incumbent pressures the prosecutor to act, no new information can be revealed, and the citizen continues to support the opponent. But now, refraining from intervention may carry an informational benefit: if the prosecutor acts, the citizen learns that the opponent is less deserving than she thought and rejects him. Notice that information is effectively wasted for some levels of citizen bias: after learning of prosecutorial action the citzen's posterior may be strictly higher than her bias μ. In the Appendix (section C.2), we show that allowing mixed prosecutor strategies eliminates this feature: from the incumbent's perspective, the best mixed strategy keeps the citizen exactly indifferent between supporting and not supporting, given citizen bias.

Recall that we have abstracted away from the possibility that the citizen dislikes incumbent interference in justice, i.e., that μ is a function of intervention. The above provides some insight into when incorporating this possibility would change our results. Examining Figure 2, it is clear that citizen dislike for interference would only increase its cost to the incumbent if the incumbent were currently preferred by the citizen ($\mu < p$) *and* interference affected μ strongly enough to shift the citizen's support to the opponent (i.e., to $\mu > p$). The likelihood of this occurring would presumably depend, among other things, on the starting level of μ, and even when intervention would indeed shift citizen support, the costliness of the shift to the incumbent would again depend upon the starting level of μ. In particular, the cost would be highest for an extremely popular incumbent, since intervention would now increase the likelihood of the consequence but also destroy her popularity. If the incumbent were moderately popular, however, her popularity would be lost if she intervened, but also if she held back and the independent prosecutor saw the innocent signal—meaning that when the likelihood of the innocent signal is high, the cost of citizen disapproval is low. We show this formally in the Appendix.

Having examined the informational consequences of interference, we now investigate what determines its direct cost, $K(\lambda^F)$. We can show the following:

Proposition 2. *Equilibrium interference, λ^F, is increasing in the likelihood of the consequence ψ, the prosecutor's concern about type I errors q, and the probability that the* *prosecutor sees the innocent signal $Pr(s = i)$.*

Because the cost of intervention K is an increasing function of λ^F, it increases in both the prosecutor's concern for inflicting the consequence on the innocent (q), and the likelihood that such a type I error would actually occur if the prosecutor acted regardless of signal. The likelihood of error in turn increases in two things: the probability of seeing the innocent signal ($\Pr(s = i)$), and the prosecutor's pivotality in inflicting the consequence (ψ). All else equal, the higher these parameters, the more unwilling the prosecutor to ignore his signal, and the larger the reward the incumbent must offer to motivate action.

Despite their effect on the cost of intervention, whether the overall incidence of intervention increases or decreases in ψ and $\Pr(s = i)$ (for a given incumbent popularity level) depends on the steepness of the cost function K and the prosecutor's concern for wrongful prosecutions q. This is because the gains from intervention *also* increase in prosecutorial pivotality and the probability of seeing the true innocent signal; the former because intervention is more valuable the likelier it is to lead to the consequence, and the latter because the higher $\Pr(s = i)$, the less likely the prosecutor to act on his own. High q (and/or a steep cost function) magnify the effects ψ and $\Pr(s = i)$ on the costs of interference, making them more likely to exceed the gains. Low q, by contrast, diminishes these effects, increasing the likelihood that the gains outweigh the costs.

Model Implications

These results provide some real-world insights. Most importantly, they suggest that, when the informational consequences of interference are considered, a politician's or party's current popularity may play an interesting role in intervention decisions. When public support is both necessary to remain in power (high B) and attainable but not assured under all circumstances (moderate μ), prosecutors may be manipulated—or left alone—to control the information they provide to the public about political opponents. Whether manipulation occurs under these conditions depends on whether the incumbent is currently favored or disfavored. Intervention is least likely when a leader or party has lost their competitive edge to an opponent, because the need to regain support motivates them to refrain from politicizing prosecutions and hope that an independent prosecutor will convince the public that the opponent is likely guilty. This suggests one potential explanation for the sometime reluctance of politicians to push for the prosecution of popular opponents even when those opponents appear likely to be guilty.

By contrast, intervention is most likely if a politician or party is currently preferred by the public—for example, because the public suspects the political opponent of shady dealings. This is because the politician/party might lose support if an independent prosecutor refused to act against the opponent and the public updated positively about the opponent's innocence. The need to maintain support then motivates the politician/party to engineer a biased prosecution simply to keep the public from learning that the opponent may be innocent. This perhaps suggests an explanation for the incidence of clearly politicized investigations and prosecutions of opposition figures during competitive election campaigns: the point of the investigation is not to show guilt, but merely to create enough noise to keep the public from learning of innocence.

The informational role of prosecution described above only emerges when maintaining public support is possible but not certain. If a party or leader is either extremely popular (low μ) or deeply unpopular (high μ), the informational repercussions of intervention are irrelevant, because public opinion cannot be changed. The prosecutor's informational role is similarly irrelevant if the leader (or party) does not *need* support (low B), for example, because they head a dictatorship capable of extreme repression, they are at the close of their political career, or the next election is years away.[12] In both cases, only the desire

[12]Some of these predictions diverge substantially from those made by theories of political intervention in the judiciary—an unsurprising outcome, given the potentially very different reasons politicians might have for intervening in prosecution decisions, as opposed to preventing independent judicial review. For example, the well-known insurance policy theory suggest that incumbents are least likely to intervene in judicial review when they are sure to lose, while our model suggests that incumbents should intervene *more* with prosecutions when they are sure to lose.

to inflict the consequence motivates interference, and only its direct costs can prevent it.

When do these direct costs and benefits serve as a strong check on incumbents? First, interference is less appealing when its costs steeply increase in λ; for example, due to institutional protections such as the threat of impeachment for bad behavior. Second, interference is less appealing when λ^F itself is large. Since λ^F increases with prosecutorial pivotality ψ, this recovers the rationale for institutions such as the grand jury system, which lowers ψ by requiring that a large number of independent actors agree to prosecution. At the same time, most systems of prosecution may in some ways be structured to incentivize interference: λ^F increases with prosecutors' concern about prosecuting the innocent, and where careers depend on conviction rates, prosecutors are liable to be primarily worried about failing to prosecute the guilty. Finally, interference is less likely when the benefit the incumbent derives from inflicting the consequence (α) is low. If we assume that the benefit increases with the damage the consequence inflicts, an interesting conclusion is that the difficulty in obtaining political corruption convictions in the United States,[13] and norms of lenient sentencing of politicians more generally,[14] should actually help prevent political intervention.

Extensions

When Prosecutors Are Political

In our baseline model, the prosecutor does not care whether the citizen supports the opposition or the incumbent—in this sense, he is apolitical. However, prosecutors may, for personal or professional reasons, in fact care about whether citizens support the incumbent or her opponent. Indeed, incumbents may sometimes commit another sort of interference by purposefully appointing a partisan supporter as prosecutor. Here, we

[13] See, e.g., Feuer, Alan. "Why Are Corruption Cases Crumbling? Some Blame the Supreme Court." *The New York Times.* November 17, 2017.

[14] For an example, see, e.g., Blank, Jonah. "How the (Once) Most Corrupt Country in the World Got Clean(er)." *The Atlantic.* May 1, 2019.

explore how the prosecutor's incentives, and informational value to the incumbent, might change when he directly benefits from the citizen supporting the incumbent.

Suppose the partisan prosecutor has already been appointed, and any costs of his appointment have been realized. His utility function is the same as in the baseline model, except that he now obtains a benefit $b > 0$ if the citizen supports the incumbent:

$$U_P = u_{C\theta}(q) + a\lambda + rb$$

Because the prosecutor now cares about the citizen's support decision and can influence that decision by revealing his private information about the guilt of the target, we have a signaling game. We solve for Perfect Bayesian Equilibria, subject to some plausibility constraints discussed below.

First consider a situation in which the incumbent wishes to induce the prosecutor to act regardless of his private information. Recall that in this situation the citizen learns nothing from the prosecutor's action. Because now the prosecutor may value using his action to signal to the citizen, we must determine when it is incentive-compatible for all types of prosecutor, i.e., for prosecutors who receive each type of signal $s \in \{g, i\}$, to pool on the action $a = 1$. The attractiveness to the prosecutor of deviating to $a = 0$ naturally depends on the citizen's out-of equilibrium beliefs, which we denote by π. To avoid problems of equilibrium multiplicity and implausibility, we make the assumption that the citizen interprets deviation to $a = 0$ as a sign of innocence, i.e., that $\pi < p$.[15] Then, comparing the expected utilities to the prosecutor of choosing $a = 1$ and $a = 0$, we find that a prosecutor of type s does not deviate from $a = 1$ if

$$\lambda \geq \psi \left[q - \Pr(G|s)\right] + b \cdot \left[\mathbb{1}(\pi \geq \mu) - \mathbb{1}(p \geq \mu)\right]. \tag{5}$$

A comparison of Expression 5 to the baseline case (see Expression 2) reveals that the

[15]Note that, for the g-type of the prosecutor, $a = 0$ is strictly dominated by $a = 1$. Consequently, our restriction immediately follows from common refinements in signaling games.

prosecutor's partisanship changes his decision rule only by adding the term $b \cdot [\mathbb{1}(\pi \geq \mu) - \mathbb{1}(p \geq \mu)]$ to his calculation. As we have assumed that $\pi < p$, this term is weakly negative. Thus, if the incumbent wishes to induce pooling on action $a = 1$, she can do so more cheaply with a partisan than with an apolitical prosecutor.

Now consider a situation in which the incumbent wishes to induce each prosecutor to act in accordance with his private signal, thereby transmitting information to the citizen. The citizen's posterior beliefs about the target's guilt upon observing action a are $\Pr(G|a = 1) = \Pr(G|g)$ and $\Pr(G|a = 0) = \Pr(G|i)$, as before. The incentive compatibility condition for a prosecutor who receives a guilty signal is:

$$\lambda \geq \psi[q - \Pr(G|g)] + b \cdot [\mathbb{1}(\Pr(G|i) \geq \mu) - \mathbb{1}(\Pr(G|g) \geq \mu)]$$

while the condition for a prosecutor who receives an innocent signal is

$$\psi[q - \Pr(G|i)] + b \cdot [\mathbb{1}(\Pr(G|i) \geq \mu) - \mathbb{1}(\Pr(G|g) \geq \mu)] \geq \lambda \tag{6}$$

An examination of these conditions reveals an incentive problem for the prosecutor who receives the innocent signal. Because $b \cdot [\mathbb{1}(\Pr(G|i) \geq \mu) - \mathbb{1}(\Pr(G|g) \geq \mu)]$ is weakly negative, if the prosecutor is sufficiently concerned with undermining opposition support, and sufficiently unconcerned with wrongful convictions, the entire left-hand side of Expression 6 is negative: the prosecutor cannot be induced to signal the target's innocence. This means that in some regions of the parameter space the incumbent is unable to induce a separating equilibrium.

This result implies that a partisan prosecutor is not always good for an incumbent. In certain regions of the parameter space, such as when the prior probability of the target's guilt is high and/or the citizen moderately dislikes the incumbent, the incumbent may not want to intervene, gambling that the prosecutor will receive a guilty signal. But since if the prosecutor receives the innocent signal, he will act even without interference, the informational content of prosecutorial action is destroyed. (In the Appendix, we show that allowing the partisan prosecutor to choose mixed strategies re-introduces the possibility of (partial) citizen learning.)

We conclude that there may sometimes be an interesting trade-off between different forms of interference. Appointing a partisan prosecutor may be a cost-effective interference strategy if the incumbent wishes to ensure that her opponents are prosecuted regardless of evidence of their guilt, and does not need the citizen to learn from these prosecutions—for example, when the incumbent is already moderately popular, or does not value the citizen's support. However, if the incumbent needs the citizen to learn to gain her support, she may prefer a neutral prosecutor, whose behavior is more expensive to alter, but who allows the incumbent to choose whether the citizen learns or not.

When Courts are Informative

In most countries, other actors in the criminal justice system (judges, juries) receive, and through their actions disclose, additional information about a target's guilt. We now account for this by allowing the imposition of the consequence to depend probabilistically on both the prosecutor's action and the target's underlying guilt or innocence. This permits us to interpret the consequence as an informative conviction by a court.

Let the game proceed as in the baseline model, except that the probability of the consequence (conviction) is now:

$$\Pr(C = 1|a, \theta) = \psi_\theta \cdot a \quad \text{with} \;\; \psi_G > \psi_I.$$

Here ψ_G represents the probability that the court observes the guilty signal when the opponent is guilty, and ψ_I the probability that the court observes the guilty signal when the opponent is innocent. It is clear from this set-up that the court follows its signal, i.e., it convicts after seeing a guilty signal and acquits after seeing an innocent signal.[16]

[16]We assume that the court has accuracy concerns, but assume also that these concerns do not prevent the court from following its signal. It is plausible that for some levels of court accuracy, the court may always acquit (e.g., because observing the guilty signal does not raise the probability of guilt 'beyond a reasonable doubt'). For the purposes of

Turning to the equilibrium of the game, the prosecutor now acts ($a = 1$) if and only if:

$$\lambda \geq \psi_I q - \Pr(G|s) \left[\psi_I q + \psi_G(1-q)\right].$$

As before, we assume that absent interference, the prosecutor follows his signal. This now requires that:

$$\Pr(G|i) < \frac{\psi_I q}{\psi_I q + \psi_G(1-q)} < \Pr(G|g).$$

The incumbent's choice is then again between nonintervention ($\lambda = 0$) and the lowest amount of intervention that induces the prosecutor to act for all signals:

$$\lambda^F \equiv \psi_I q - \Pr(G|i) \left[\psi_I q + \psi_G(1-q)\right]. \tag{7}$$

From this expression, the amount of interference required to induce to prosecutor to act decreases in the court's accuracy. The better the court at convicting the truly guilty (high ψ_G) and acquitting the truly innocent (low ψ_I) the more willing the prosecutor to act regardless of signal, since he can rely on the court to fix his mistakes.

The citizen's posterior belief in the opponent's guilt now depends upon the action taken by the court as well as the prosecutor. Thus, even if the incumbent chooses full intervention, destroying the informativeness of prosecutorial action, the citizen can still update her belief in the opponent's guilt based on whether the court convicts or acquits. In particular, after observing a biased prosecution followed by conviction, the citizen believes the opponent is guilty with probability

$$\Pr\left(G|a = 1, C = 1, \lambda = \lambda^F\right) = \frac{p\psi_G}{p\psi_G + (1-p)\psi_I} \equiv p^F(1);$$

after an acquittal, the citizen believes the opponent is guilty with probability

$$\Pr\left(G|a = 1, C = 0, \lambda = \lambda^F\right) = \frac{p(1-\psi_G)}{p(1-\psi_G) + (1-p)(1-\psi_I)} \equiv p^F(0).$$

our model, this situation puts us back in the baseline case, so we ignore it.

The notation $p^F(C)$ denotes the citizen's posterior belief when there there is full intervention (F) and the conviction status is C.

If the incumbent decides not to intervene, the citizen's posterior beliefs depend on her prior p, the prosecutor's action, and the court's decision. If the prosecutor acts and the court convicts, the citizen believes the opponent is guilty with probability

$$\Pr(G|a=1, C=1, \lambda=0) = \frac{p\gamma_G\psi_G}{p\gamma_G\psi_G + (1-p)\gamma_I\psi_I} \equiv p^N(1,1);$$

if the prosecutor acts and the court acquits, she believes he is guilty with probability

$$\Pr(G|a=1, C=0, \lambda=0) = \frac{p\gamma_G(1-\psi_G)}{p\gamma_G(1-\psi_G) + (1-p)\gamma_I(1-\psi_I)} \equiv p^N(1,0);$$

and if the prosecutor does not act, she believes he is guilty with probability

$$\Pr(G|a=0, C=0, \lambda=0) = \frac{p(1-\gamma_G)}{p(1-\gamma_G) + (1-p)(1-\gamma_I)} \equiv p^N(0,0).$$

Here the notation $p^N(a, C)$ denotes the citizen's posterior belief, given a and C, under nonintervention (N).

Notice that, as the court becomes less informative, these posterior beliefs converge to the citizen's baseline posterior beliefs (and indeed the whole model converges to the baseline model). To see this, let $\psi_G = \psi + \varepsilon$ and $\psi_I = \psi - \varepsilon$, and observe that as $\varepsilon \to 0$, i.e., as informativeness approaches zero, the court's role in determining the citizen's posterior beliefs decreases, and at $\varepsilon = 0$ it disappears completely. This means that when the court is not very informative, the difference from the baseline model is small.

Additionally, the exact ordering of these posterior beliefs depends on the relative informativeness of the court's decision ψ_θ and the prosecutor's signal γ_θ, which in turn depends on two conditions. The first determines whether an independent prosecutor's decision *not* to act is more or less informative of innocence than acquittal following coopted prosecutorial action. Prosecutorial inaction is more informative than acquittal if:

$$\frac{1-\psi_I}{1-\psi_G} < \frac{1-\gamma_I}{1-\gamma_G}. \tag{8}$$

This condition implies an upper bound on the informativeness of the court's decisions.[17] We assume that it holds for the remainder of the paper, both because we wish to focus on the informational value of the prosecutor's behavior, and because the burden of proof required for conviction in court is (assuming the court is not fully aligned with the incumbent) generally higher than that required for prosecution.

The second condition involves the *degree* of general court informativeness or accuracy. If the overall informativeness of court decisions is low relative to the informativeness of the prosecutor taking action, we have:

$$\frac{\gamma_G}{\gamma_I} > \frac{\psi_G(1-\psi_I)}{\psi_I(1-\psi_G)}. \tag{9}$$

Whether Condition 9 is met has implications for the citizen's posterior beliefs on the opponent's guilt after observing different combinations of choices by the other players. When it holds (the court's overall informativeness is low), the citizen's posterior beliefs are ordered as follows:

$$p^N(0,0) < p^F(0) < p^F(1) < p^N(1,0) < p^N(1,1). \tag{10}$$

Here, the prosecutor's decisions are always more informative to the citizen than the court's: the citizen is least convinced of guilt when the prosecutor refuses to act, and most convinced of guilt when an independent prosecutor acts.

By contrast, when Condition 9 does not hold, the ordering becomes

$$p^N(0,0) < p^F(0) < p^N(1,0) < p^F(1) < p^N(1,1). \tag{11}$$

The citizen is still least convinced of guilt when the prosecutor refuses to act, but she now finds the court's decision to convict more informative than prosecutorial action.

[17]It also implies that $p^N(0,0) < p^N(1,0)$.

As in the baseline model, the incumbent's decision to intervene depends on the citizen's preferences μ in combination with intervention's effect on the citizen's posterior beliefs, as well as on the costs of intervention K, and the degree to which it increases the likelihood of conviction. The incumbent intervenes if:

$$\underbrace{B \cdot \left[\mathbb{E}^{\lambda^F}\left[\mathbb{1}\left(p^F(C) \geq \mu\right)\right] - \mathbb{E}^0\left[\mathbb{1}\left(p^N(a,C) \geq \mu\right)\right]\right]}_{\text{Informational cost or benefit of intervention (effect on citizen posterior)}} \geq \overbrace{K\left(\lambda^F\right) - \alpha\left[\Pr\left(C=1|\lambda^F\right) - \Pr\left(C=1|0\right)\right]}^{\text{Direct cost net of increased probability of the consequence}} \tag{12}$$

where $\mathbb{E}^\lambda$ denotes the expectation of a and C given λ. As in the baseline model, the right-hand side of this inequality represents interference's beneficial or deleterious effect on the citizen's posterior beliefs and resulting support decision, and denotes the threshold above which intervention is optimal. The size of this threshold is determined by a number of factors, including the informativeness of the prosecutor's and court's decisions and the citizen's idiosyncratic preference, μ. The left hand side again shows the direct costs of intervention to the incumbent, K, net of the benefit she derives from increasing the probability of conviction. These costs still depend on q and the prosecutor's signal, but they now also depend on the likelihood and accuracy of conviction. Intervention's effect on the probability of conviction is constant, as in the baseline, and equal to:

$$p\psi_G(1-\gamma_G) + (1-p)\psi_I(1-\gamma_I) > 0.$$

We denote this term Δ^α. Note that Δ^α can be larger or smaller than the equivalent term in the baseline model ($\psi\left[p(1-\gamma_G) + (1-p)(1-\gamma_I)\right]$), depending on the relative sizes of ψ, ψ_G, and ψ_I.

As in the baseline, we examine the incumbent's incentives to intervene as a function of citizen preferences μ, since where μ falls in the ordering of citizen posterior beliefs determines the threshold above which intervention is optimal. Figure 3 shows the in-

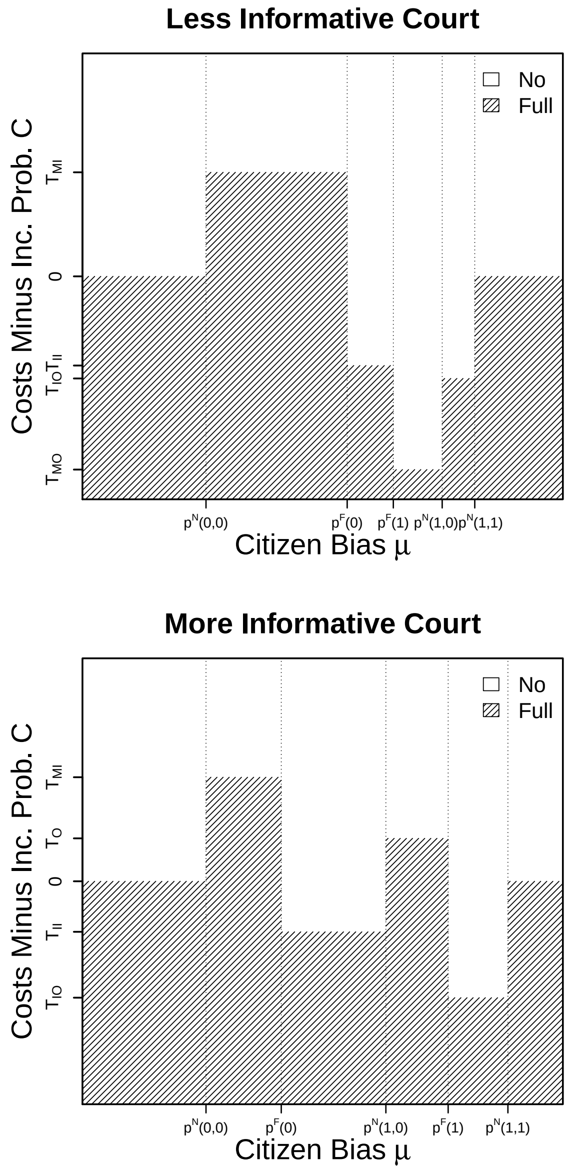

Figure 3: Overview of equilibrium with an informative court. Top panel: court is less accurate. Bottom panel: court is more accurate. Parameter values: both panels: $p = 0.6$, $\gamma_G = 0.85$, $\gamma_I = 0.35$; top panel: $\psi_G = 0.55$ and $\psi_I = 0.45$; bottom panel: $\psi_G = 0.68$ and $\psi_I = 0.32$.

cumbent's optimal intervention threshold as a function of μ when the court has low accuracy/informativeness (top panel) and high accuracy/informativeness (bottom panel). In both cases, there are now six posterior belief regions μ may inhabit, up from four in the baseline model. These regions are demarcated by the posterior belief orderings in Expressions 10 (for the low information case) and 11 (for the high information case), and shown in Figure 3.

We begin with the low information case. Examining the top panel in Figure 3, it is clear that in direct analogue to the baseline model, there are two extreme regions of citizen biases ($\mu < p^N(0,0)$ or $\mu > p^N(1,1)$) in which no information changes the citizen's mind. Because here, intervention does not affect citizen support, the benefit of support B is irrelevant to the incumbent's decision. If μ lies in one of these regions, the incumbent interferes if and only if:

$$0 \geq K\left(\lambda^F\right) - \alpha\Delta^\alpha$$

where the right-hand side intervention threshold is zero, and the left hand side represents the costs of interference net of the increased likelihood of conviction. Note that while the intervention threshold here is identical to the threshold in the analogous regions of the baseline model, the values taken on by the right-hand side are different. This is because the direct costs of intervention and its effect on the probability of conviction are both affected by the court.

Now consider the four middle regions. Two also have direct analogues in, and produce intervention thresholds identical to, the baseline model. First, when the citizen's preferences μ are in the region $\left(p^N(0,0), p^F(0)\right]$, this is analogous to the baseline situation in which the citizen supports the incumbent unless she learns through prosecutorial inaction that the opponent is less blameworthy than she thought. Here, the citizen would support the incumbent even if the court acquitted the opponent after a political prosecution; only prosecutorial inaction would provide sufficient evidence of innocence to change her mind. As in the baseline, the incumbent's incentives to intervene are strongest here, because intervention prevents the prosecutor from revealing the opponent's innocence. The incumbent interferes if:

$$\overbrace{B\left(1 - p\gamma_G - (1-p)\gamma_I\right)}^{=T_{MI}} \geq K\left(\lambda^F\right) - \alpha\Delta^\alpha.$$

Second, the region $\left(p^F(1), p^N(1,0)\right]$ is analogous to the baseline region in which the citizen supports the opponent unless she learns more proof of his guilt. When μ is in this region, the citizen supports the opponent even if a political prosecution resulted in his conviction, because conviction is not very informative of guilt; she only changes her mind if an independent prosecutor acts against the opponent. Because here (as in the baseline model), intervention precludes the incumbent's ever gaining the citizen's support, the incumbent's incentives to intervene are weakest when μ lies in this region. The incumbent intervenes if

$$\overbrace{B\left(-p\gamma_G-(1-p)\gamma_I\right)}^{=T_{MO}} \geq K\left(\lambda^F\right)-\alpha\Delta^\alpha.$$

The two remaining, in-between regions have no analogue in the baseline model. They emerge because the court provides some—but not very precise—information: as court informativeness approaches zero, they disappear. In these regions, incentives to interfere are intermediate, effectively bridging the regions where incentives are very high or very low. First, if $\mu \in \left(p^F(0), p^F(1)\right]$, the citizen supports the incumbent, after observing a biased prosecution, only if the opponent is convicted rather than acquitted. Then the incumbent intervenes if:

$$\overbrace{B\left(p(\psi_G-\gamma_G)+(1-p)(\psi_I-\gamma_I)\right)}^{\equiv T_{II}} \geq K\left(\lambda^F\right)-\alpha\Delta^\alpha$$

Here, intervention can have a wide range of effects. It is beneficial if the prosecutor is unlikely to see the guilty signal but conviction is probable, and quite damaging if the prosecutor is highly likely to see the guilty signal but acquittal is probable. In fact, the threshold T_{II} is constrained only by being smaller than the largest intervention threshold (T_{MI}) and larger than the smallest (T_{MO}).

Finally, if $\mu \in \left(p^N(1,0), p^N(1,1)\right]$, the citizen favors the opposition enough to require significant evidence of guilt before changing her mind: both an independent prosecution and a conviction are necessary. Under these conditions, the incumbent intervenes unless the probability that these events occur is sufficiently high:

$$\overbrace{B\left(-p\psi_G\gamma_G-(1-p)\psi_I\gamma_I\right)}^{\equiv T_{IO}} \geq K\left(\lambda^F\right)-\alpha\Delta^\alpha.$$

By inspection, we have the following result:

Proposition 3. *When conviction is informative of guilt, but its informativeness is low relative to prosecutorial action, intervention thresholds can be ordered as follows:*

$$T_{MI} > 0 > T_{IO} > T_{MO}, \text{ and } T_{II} \in (T_{MO}, T_{MI}).$$

Consider how the situation changes when the court is highly informative (bottom panel of Figure 3). Recall that when the court was relatively uninformative, the citizen's posterior belief in the opponent's guilt was higher after observing an independent prosecution than a coopted prosecution followed by conviction, $p^F(1) < p^N(1,0)$. If the court is highly informative, the reverse is true: the citizen's posterior is higher when the opponent is convicted after a biased prosecution than when he is acquitted after an independent prosecution, $p^N(1,0) < p^F(1)$. Due to this reversal, the region $\left(p^F(1), p^N(1,0)\right]$ that exists in the low-information case disappears, along with its corresponding threshold T_{MO}, and is replaced by a new region: $\left(p^N(1,0), p^F(1)\right]$. In this new region, the citizen moderately prefers the opponent, such that—because the court is highly informative—only his conviction can motivate her to support the incumbent. But since conviction requires prosecution, the incumbent now has a *positive* incentive to interfere. Formally, the incumbent now intervenes if:

$$\overbrace{B\left(p\psi_G(1-\gamma_G) + (1-p)\psi_I(1-\gamma_I)\right)}^{\equiv T_O} \geq K\left(\lambda^F\right) - \alpha\Delta^\alpha$$

This yields the following result:

Proposition 4. *When conviction is highly informative of guilt relative to prosecution, intervention thresholds can be ordered as follows:*

$$T_{MI} > T_O > 0 > T_{IO}, \text{ and } T_{II} < T_{MI}.$$

Notice that even though, by Propositions 3 and 4, all intervention thresholds but one are the same in the low and high court informativeness cases, it is not easy to discern how the actual incidence of intervention changes with court informativeness, either overall, or within different support regions. This is because court informativeness affects both the relative sizes of the support regions *and* the cost-benefit analysis the incumbent conducts for each region. We consider the relationship between court informativeness and the incidence of intervention in more detail below. To do so we again let $\psi_G = \psi + \varepsilon$ and $\psi_I = \psi - \varepsilon$. Then we have the following result:

Proposition 5. *For any citizen bias μ, the incidence of political interference in prosecutions is ambiguous in court informativeness ε. However, if p is sufficiently high and μ is such that the intervention threshold is not T_{IO}, court accuracy increases interference.*

Recall that for interference to be optimal, its informational consequences must positively exceed its direct cost net of any increase in the probability of the consequence. The result described in Proposition 5 is due primarily to the fact that for any level of citizen support μ, the net cost of interference is ambiguous in court informativeness. This in turn is because, while the court's accuracy always decreases the direct costs $K(\lambda^F)$ of intervention, it also decreases the probability of obtaining the consequence unless the prior probability of guilt, p, is very high.

To understand more clearly how court informativeness affects the net costs of intervention, remember that by Expression 7, the more competent a court is in convicting the truly guilty (high ψ_G) and acquitting the truly innocent (low ψ_I), the more willing a prosecutor to act regardless of his signal, and the lower the level of interference λ^F necessary to induce prosecutorial action. This means that the more informative the court, the lower the direct costs of intervention. At the same time, however, if the opponent is not extremely likely to be guilty (in particular, $p > \frac{1-\gamma_I}{1-\gamma_I+1-\gamma_G} > \frac{1}{2}$) the court's greater accuracy makes conviction relatively less likely.

For the regions of citizen preference μ where the court's decisions alone can determine citizen support for the incumbent, the result in Proposition 5 is also due to the ambiguous effect of court accuracy on the informational consequences of intervention. (In regions where the court's decisions are never decisive, court accuracy is irrelevant to citizen support and consequently has no effect on interference's informational consequences.)

First, when the intervention threshold is T_{II}, the citizen would support the opponent if the court acquitted after a biased prosecution, but if any new evidence of guilt were revealed, she would switch to the incumbent. Here, intervention can carry an informational benefit or a cost, depending on the opponent's likelihood of guilt and the court and prosecutor's relative accuracy in determining it, and the size of this cost/benefit is ambiguous in court informativeness. Informativeness decreases the informational cost (increases the benefit) of intervention when the opponent's guilt is more likely than not, $p > 1/2$, and increases (decreases) it otherwise.

Second, if the relevant intervention threshold is T_{IO}, the citizen is very anti-incumbent, supporting her only after observing both both independent prosecution and conviction of the opponent. Here, intervention always carries an informational cost. However, the severity of this cost decreases in court informativeness when the opponent's guilt is unlikely anyway, $p < \frac{\gamma_I}{\gamma_I + \gamma_G} < \frac{1}{2}$. Finally, for the intervention threshold unique to the high-information case, T_O, where the citizen is moderately pro-opponent but only requires conviction to support the incumbent, interference is always beneficial. It makes the court more likely to hear the case, which makes it, ceteris paribus, more likely to convict. However, the size of this benefit only increases with court accuracy when the opponent's guilt is very likely: $p > \frac{1-\gamma_I}{1-\gamma_I - 1-\gamma_G} > \frac{1}{2}$.

We conclude that both the simple presence in a country of an independent, informative court, and that court's relative competence in correctly determining guilt and innocence, have complex implications for the incidence and consequences of intervention. Conditional on prosecutor action, the more informative the court, the lower the likelihood of false convictions. Under some conditions, however, the more informative the court, the higher the likelihood of political prosecutions.

These results have significant implications regarding optimal institutional design. In particular, many reforms to the court system aim to increase its accuracy, whether by increasing judicial expertise, providing judges with more assistance, or disincentivizing political interference in judicial decisions. Because an increase in court accuracy always lowers the cost of intervention, such reforms risk simultaneously increasing the incumbent's incentives to interfere with the prosecutor's work. Moreover, the size of this risk to prosecutorial independence will vary unpredictably over time. For example, the risk would be especially high in circumstances where politicians are generally corrupt, or if a particular opponent politician is likely guilty of malfeasance (high p), or for certain levels of citizen bias μ. Thus, our analysis suggests that institutional designers who wish to discourage intervention in the justice system might want to complement court system reforms with reforms that increase the cost of prosecutor interference, e.g., by shielding prosecutors from dismissal or increasing their wages.

Timing Interventions

Thus far, we have not formally considered that many months, or years, may pass between a prosecutor's initial action and the (informative) conclusion of a case in court. Yet prosecutions begun late in an incumbent's term in office, with no chance of a court deciding the case before the next election, will have very different informational consequences than those begun early. This suggests that informational incentives to intervene may vary substantially over time.

To explore this possibility, we extend the model further. In this extension, formally introduced and analyzed in the appendix, there are two periods: *early* and *late* in the incumbent's term. In each, the prosecutor receives a noisy signal and the incumbent chooses a level of intervention. However, if the prosecutor acts in the early period, the citizen observes the court's finding of guilt or innocence prior to the election, while if the prosecutor acts in the late period, the citizen must make her support choice before the court's decision is reached. For simplicity, we assume that there is no uncertainty about the court reaching a decision prior to the election.[18]

Because we are interested in *when*, rather than *if*, incumbents intervene, we assume that the incumbent certainly intervenes in the second period if this stage of the game

[18] Similar to the literature on pandering (e.g., Canes-Wrone, Herron and Shotts, 2001), one could assume that with some probability, the citizen receives an informative signal, and with complementary probability, the citizen does not.

is reached. This supposes that the benefit of inflicting the consequence is sufficiently large. We show that if the prosecutor anticipates being interfered with in the second period, he never acts in the first period absent a bribe (because he will receive it for sure later on—and can act on a first-period guilty signal then as well). The incumbent then decides whether to intervene early to leave the court time to decide the case, or late, to effectively eliminate all information. We find that moderately unpopular incumbents prefer to intervene early and gamble that the court will release incriminating information, while moderately popular incumbents intervene late to ensure that an acquittal comes too late to inform the citizen's choice. Analogous to the previous model variations, very popular or unpopular incumbents are indifferent between intervening early or late. Thus, while the same general mechanism of information manipulation operates here, the behavioral incentives are different, in part because only the court's informativeness is relevant on the path of play.

Other Extensions

In the Appendix, we additionally consider when the model is robust to loosening our other major assumptions. In particular, we show that all our results hold when intervention is unobservable and only discovered with some probability (e.g., by a whistleblower, the media, or the political opponent), so long as the probability of observation exceeds some threshold. As discussed above, we also show that citizen dislike for intervention is often irrelevant to equilibrium dynamics. Such citizen dislike only affects incumbent behavior if the dislike is extremely strong *and* the incumbent is currently popular; even then, the incentives for intervention do not entirely disappear. Finally, we show that if the prosecutor must exert costly effort to learn about guilt or innocence, there is always *some* interference, because interference now continuously maps onto citizen beliefs. Moreover, depending on the prosecutor's relative concern about type I and type II errors, such interference can increase or decrease prosecutorial effort.

Discussion

Our model relates to a number of existing empirical findings. Within the small political economy literature on political prosecutions, it is particularly relevant to Gordon (2009). First, our main result on interference given an informative court is consistent with Gordon's empirical strategy and main finding. In our model, interference has, overall, a positive effect on the likelihood of obtaining a conviction from the court, because it ensures prosecutor action. But *conditional on prosecution*, interference makes conviction less likely.[19] Empirically, this should translate to political opponents receiving shorter sentences on average, both because more of them should be acquitted and because their plea bargains should be more favorable.[20] This, of course, is precisely what Gordon finds.

Second, the general relationship we identify between an incumbent's popularity and her incentives to intervene suggests that care should be taken when comparing criminal prosecutions decided at different points in an incumbent's term. This is for several reasons. One is that incumbent popularity may vary over time. Another is that even fixing incumbent popularity, our investigation into the timing of interventions reveals that the pools of cases decided at different points in an incumbent's term may be substantially different due to their different informational relevance. For example, cases decided late in a term are informative for reelection, while those decided early are not. This implies that there may be some unobserved time-based heterogeneity in Gordon (2009)'s case data.

Our results additionally relate to the more general scholarship on scandals and electoral bombshells. For example, Nyhan (2015) finds that the likelihood of an (often opposition-generated) scandal about an incumbent president decreases, while the inten-

[19]Under interference and $a = 1$, the probability of a conviction is $p\psi_G + (1-p)\psi_I$. When there is no interference and $a = 1$, it is $\Pr(G|g)\psi_G + (1-\Pr(G|g))\psi_I$ because the prosecutor only acts when receiving a guilty signal. The second term is larger than the first term.

[20]A sentence in a plea bargain can be interpreted as representing the expected sentence at trial, taking into account the probability of acquittal.

sity of media coverage *increases*, with presidential popularity among opposition voters.[21] While Nyhan explains this result as an artifact of scandals' greater likelihood in a president's second term, it also seems broadly consistent with the logic behind our argument that incumbents should refrain from interference when their opponents are moderately popular. In particular, if the opposition in Nyhan refrains from manufacturing a questionable scandal if the president is moderately popular, in the hopes that a real scandal will independently emerge (as our incument does with prosecution), there would be fewer total scandals under popular presidents. And if the media covers real scandals more intensely than manufactured ones, the scandals that *do* occur under popular presidents should have higher-intensity media coverage.

Our timing extension also dovetails interestingly with Gratton, Holden and Kolotilin (2018), who provide some empirical evidence in favor of the argument that October surprises in U.S. presidential campaigns are driven by the strategic behavior of "bad" senders who wish to release false information about the president, because this minimizes the amount of time receivers have to learn that the information is in fact false. The partial analogue in our model is that relatively popular incumbents begin prosecutions late in order to avoid releasing potentially damaging information before reelection.

Our model also provides some additional, testable empirical implications relevant to existing and future work. First, we find that the effect of a politician's popularity on prosecution (and perhaps scandal) is non-monotone: political interference in justice should occur at approximately equal rates (all else equal) when an incumbent has extremely low and extremely high poll numbers, is most likely when the incumbent is polling slightly ahead, and least likely when she is polling slightly behind. Moreover, popularity in our model moderates the effect of other variables, such as the value of holding office, on the likelihood of interference. Failing to consider this nonmonotonicity when assessing the relationship between incentives to intervene and a politician's popularity or office benefits may lead to bias (in either direction) in empirical results. Future research might inves-

[21] This finding may be specific to presidents because it does not seem to generalize to governors (see Nyhan, 2017).

tigate whether apparently monotonic relationships between popularity and incentives to intervene (or, as in Nyhan (2015) and Nyhan (2017), to propagate scandals about political figures) conceal interesting variation across different levels of incumbent popularity.

Second, as discussed above, our timing extension suggests that late-term and early-term prosecutions may be substantially different depending on incumbent popularity. This result can be directly tested, and it also suggests that empirical researchers may wish to consider the electoral calendar carefully when examining patterns in prosecution over time. Third, our model implies that incumbents who are polling far ahead should be more likely to appoint partisan prosecutors than those who are not. Fourth, our results suggest that laws or court decisions that increase the difficulty of obtaining political corruption convictions (as several recent U.S. Supreme Court decisions purportedly did) may decrease incentives for political interference. Fifth, it implies that especially in situations where corruption is endemic and the incumbent is unpopular, accurate, independent courts may increase the incentives for political prosecutions. These latter findings also have implications for institutional design, suggesting (for example) that there may be benefits to the lenient treatment of political corruption (possibly reducing α and hence providing a counter to the incumbent's use of interference).

We leave a number of avenues unexplored. For example, we do not consider that prosecutors may also be vulnerable to interference by other outside actors, such as members of the political opposition. Building on Dziuda and Howell (2021), who let both parties influence scandals, future research could investigate the competing effects of opposition and incumbent interference in prosecutorial decision making. Likewise, we do not consider the role of the media in broadcasting information about criminal prosecutions (Nyhan, 2015, 2017). And we do not allow politicians to corrupt the court system more broadly, by buying judges, juries, etc, nor do we consider how the vulnerability of different actors to interference might vary across justice systems (e.g., common-law versus civil-law systems). Future research might amend the model to address these various possibilities.

References

Besley, Timothy and Andrea Prat. 2006. "Handcuffs for the grabbing hand? Media capture and government accountability." *American economic review* 96(3):720–736.

Bueno De Mesquita, Ethan and Amanda Friedenberg. 2011. "Ideologues or Pragmatists?" *Journal of the European Economic Association* 9(5):931–951.

Canes-Wrone, Brandice, Michael C Herron and Kenneth W Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* pp. 532–550.

Dewatripont, Mathias and Jean Tirole. 1999. "Advocates." *Journal of Political Economy* 107(1):1–39.

Dragu, Tiberiu and Mattias Polborn. 2013. "The administrative foundation of the rule of law." *The Journal of Politics* 75(4):1038–1050.

Dziuda, Wioletta and William G. Howell. 2021. "Political Scandal: A Theory." *American Journal of Political Science* 65(1):197–209.

Finkel, Jodi. 2008. *Judicial reform as political insurance.* University of Notre Dame.

Fox, Justin and Matthew C. Stephenson. 2014. "The constraining, liberating, and informational effects of nonbinding law." *The Journal of Law, Economics, & Organization* 31(2):320–346.

Gehlbach, Scott and Alberto Simpser. 2015. "Electoral Manipulation as Bureaucratic Control." *American Journal of Political Science* 59(1):212–224.

Gehlbach, Scott and Konstantin Sonin. 2014. "Government Control of the Media." *Journal of Public Economics* 118:163–171.

Ginsburg, Tom. 2003. *Judicial Review in New Democracies.* Cambridge University Press.

Gordon, Sanford C. 2009. "Assessing Partisan Bias in Federal Public Corruption Prosecutions." *American Political Science Review* 103(4):534–554.

Gordon, Sanford C and Gregory A Huber. 2002. "Citizen oversight and the electoral incentives of criminal prosecutors." *American Journal of Political Science* pp. 334–351.

Graham, Matthew and Milan W Svolik. 2019. "Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States." *Working Paper* pp. 1–54.

Gratton, Gabriele, Richard Holden and Anton Kolotilin. 2018. "When to Drop a Bombshell." *The Review of Economic Studies* 85(4):2139–2172.

Hayo, Bernd and Stefan Voigt. 2007. "Explaining de facto judicial independence." *International Review of Law and Economics* 27(3):269–290.

Helmke, Gretchen and Frances Rosenbluth. 2009. "Regimes and the rule of law: Judicial independence in comparative perspective." *Annual Review of Political Science* 12:345–366.

Hirschl, R. 2004. *Towards Juristocracy. The Origins and Consequences of the New Constitutionalism.* Harvard University Press.

Huang, Haifeng. 2015. "Propaganda as signaling." *Comparative Politics* 47(4):419–444.

Hübert, Ryan. 2019. "Biased Judgments without Biased Judges: How Legal Institutions Cause Errors." *The Journal of Politics* pp. 1–33.

Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101(6):2590–2615.

Leutheusser-Schnarrenburger, Sabine. 2009. Allegations of politically-motivated abuses of the criminal justice system in Council of Europe member state, Document 11993. Technical report Council of Europe Parliamentary Assembly.
URL: *http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=12276&lang=en*

Luo, Zhaotian and Arturas Rozenas. 2018. "Strategies of Election Rigging: Trade-offs, Determinants, and Consequences." *Quarterly Journal of Political Science* 13(1):1–28.

Melton, James and Tom Ginsburg. 2014. "Does de jure judicial independence really matter? A reevaluation of explanations for judicial independence." *Journal of Law and Courts* 2(2):187–217.

Moustafa, Tamir. 2014. "Law and Courts in Authoritarian Regimes." *Annual Review of Law and Social Science* 10:281–299.

Nyhan, Brendan. 2015. "Scandal potential: How political context and news congestion affect the president's vulnerability to media scandal." *British Journal of Political Science* 45(2):435–466.

Nyhan, Brendan. 2017. "Media Scandals are Political Events: How contextual factors affect public controversies over alleged misconduct by US governors." *Political Research Quarterly* 70(1):223–236.

Pereira, Anthony W. 2008. Of Judges and Generals: Security Courts under Authoritarian Regimes in Argentina, Brazil, and Chile. In *Rule by Law: The Politics of Courts in Authoritarian Regimes*, ed. Tom Ginsburg and Tamir Moustafa. Cambridge University Press.

Ramseyer, J. Mark. 1994. "The Puzzling (In)dependence of Courts: A Comparative Approach." *The Journal of Legal Studies* 23(2):721–747.

Shotts, Kenneth W. and Alan E. Wiseman. 2010. "The Politics of Investigations and Regulatory Enforcement by Independent Agents and Cabinet Appointees." *The Journal of Politics* 72(1):209–226.

Staton, Jeffrey K. 2006. "Constitutional review and the selective promotion of case results." *American Journal of Political Science* 50(1):98–112.

Stephenson, Matthew C. 2004. "Court of public opinion: Government accountability and judicial independence." *Journal of Law, Economics, and Organization* 20(2):379–399.

Tonry, Michael. 2012. "Prosecutors and Politics in Comparative Perspective." *Crime and Justice* 41(1):1–33.

Turner, Ian R. 2017. "Working Smart and Hard? Agency Effort, Judicial Review, and Policy Precision." *Journal of Theoretical Politics* 29(1):69–96.

Vanberg, Georg. 2001. "Legislative-Judicial Relations: A Game-Theoretic Approach to Constitutional Review." *American Journal of Political Science* 45(2):346–361.

Contents

A Omitted Proofs

Proof of Proposition 1:

Proof. The expected utility of choosing no interference is:

$$\mathbb{E}\left[U_{\text{Inc}}(\lambda = 0)\right] = \Pr(s = g)\left[\alpha\psi + \mathbb{1}\left(\Pr(G|g) \geq \mu\right)B\right] + \Pr(s = i)\left[\mathbb{1}\left(\Pr(G|i) \geq \mu\right)B\right]$$

The expected utility of choosing full interference is:

$$\mathbb{E}\left[U_{\text{Inc}}(\lambda = \lambda^F)\right] = \alpha\psi + \mathbb{1}\left(p \geq \mu\right)B$$

We thus have four cases:

Case 1: $\mu \leq \Pr(G|i)$: In this case, the expected utility of full interference is $\alpha\psi + B - K(\lambda^F)$ while the expected utility of no interference is $\Pr(g)\alpha\psi + B$. Re-arranging yields that full interference is optimal if $0 \geq K\left(\lambda^F\right) - \alpha\psi\Pr(i)$.

Case 2: $\mu \in (\Pr(G|i), p]$. In this case, the expected utility of full interference is $\alpha\psi + B - K(\lambda^F)$ while the expected utility of no interference is $\Pr(g)\left[\alpha\psi + B\right]$. Re-arranging yields that full interference is optimal if $T_{MI} = B\Pr(i) \geq K\left(\lambda^F\right) - \alpha\psi\Pr(i)$.

Case 3: $\mu \in (p, \Pr(G|g)]$. In this case, the expected utility of full interference is $\alpha\psi - K(\lambda^F)$ while the expected utility of no interference is $\Pr(g)\left[\alpha\psi + B\right]$. Re-arranging yields that full interference is optimal if $T_{MO} = -B\Pr(g) \geq K\left(\lambda^F\right) - \alpha\psi\Pr(i)$.

Case 4: $\mu > \Pr(G|g)$. In this case, the expected utility of full interference is $\alpha\psi - K(\lambda^F)$ while the expected utility of no interference is $\Pr(g)\left[\alpha\psi\right]$. Re-arranging yields that full interference is optimal if $0 \geq K\left(\lambda^F\right) - \alpha\psi\Pr(i)$.

By inspection: $T_{MI} > 0$ while $T_{MO} < 0$. Moreover, $\frac{\partial T_{MI}}{\partial B} = \Pr(i) > 0$ and $\frac{\partial T_{MO}}{\partial B} = -\Pr(g) < 0$. □

Proof of Proposition 2:

Proof. λ^F is defined as $\psi\left[q - \Pr(G|i)\right] = \psi\left[q - \frac{p(1-\gamma_G)}{\Pr(i)}\right]$, where $\Pr(i) = p(1-\gamma_G) + (1-p)(1-\gamma_I)$. Hence:

$$\begin{aligned}
\frac{\partial \lambda^F}{\partial \psi} &= q - \Pr(G|i) > 0 \\
\frac{\partial \lambda^F}{\partial q} &= \psi > 0 \\
\frac{\partial \lambda^F}{\partial \Pr(i)} &= \frac{\psi p(1-\gamma_G)}{\left[\Pr(i)\right]^2} > 0
\end{aligned}$$

Moreover:

$$\begin{aligned}
\frac{\partial \Pr(i)}{\partial p} &= -(\gamma_G - \gamma_I) < 0 \\
\frac{\partial \Pr(i)}{\partial \gamma_G} &= -p < 0 \\
\frac{\partial \Pr(i)}{\partial \gamma_I} &= -(1-p) < 0
\end{aligned}$$

Finally, since $\frac{\partial \Pr(G|i)}{\partial p} = \frac{(1-\gamma_G)(1-\gamma_I)}{\left[\Pr(i)\right]^2} > 0$ and $\frac{\partial \Pr(G|i)}{\partial \gamma_G} = \frac{-p(1-p)(1-\gamma_I)}{\left[\Pr(i)\right]^2} < 0$, we also have:

$$\begin{aligned}
\frac{\partial \lambda^F}{\partial p} &= \frac{\partial \lambda^F}{\partial \Pr(G|i)} \frac{\partial \Pr(G|i)}{\partial p} = -\psi \frac{\partial \Pr(G|i)}{\partial p} < 0 \\
\frac{\partial \lambda^F}{\partial \gamma_G} &= \frac{\partial \lambda^F}{\partial \Pr(G|i)} \frac{\partial \Pr(G|i)}{\partial \gamma_G} = -\psi \frac{\partial \Pr(G|i)}{\partial \gamma_G} > 0
\end{aligned}$$

□

Proof of Proposition 3:

Proof. Recall the threshold derived in the main text:

$$\begin{aligned}
&T_{MI} = B\left[1 - p\gamma_G - (1-p)\gamma_I\right] > 0 \\
&T_{MO} = B\left[-p\gamma_G - (1-p)\gamma_I\right] < 0 \\
&T_{II} = B\left[p\left(\psi_G - \gamma_G\right) + (1-p)\left(\psi_I - \gamma_I\right)\right] \\
&T_{IO} = B\left[-p\psi_G\gamma_G - (1-p)\psi_I\gamma_I\right] < 0
\end{aligned}$$

The comparisons of 0, T_{MI}, and T_{MO} follow from Proposition 1. To see that $T_{IO} > T_{MO}$, suppose not:

$$T_{MO} \geq T_{IO}$$
$$B\left[-p\gamma_G - (1-p)\gamma_I\right] \geq B\left[-p\psi_G\gamma_G - (1-p)\psi_I\gamma_I\right]$$

Re-arranging yields:

$$-p\gamma_G(1-\psi_G) - (1-p)\gamma_I(1-\psi_I) \geq 0$$

which is a contradiction.

Finally, consider T_{II}. We have $T_{II} > T_{MO}$ because $p\psi_G + (1-p)\psi_I > 0$ but $T_{II} < T_{MI}$ because $p\psi_G + (1-p)\psi_I < 1$. □

Proof of Proposition 4:

Proof. T_O is defined as $B\left[p\psi_G(1-\gamma_G) + (1-p)\psi_I(1-\gamma_I)\right] > 0$. To see that $T_O < T_{MI}$, suppose not:

$$T_O \geq T_{MI}$$
$$B\left[p\psi_G(1-\gamma_G) + (1-p)\psi_I(1-\gamma_I)\right] \geq B\left[1 - p\gamma_G - (1-p)\gamma_I\right]$$

This rearranges to:

$$1 \geq p\left[\gamma_G - \psi_G(1-\gamma_G)\right] + (1-p)\left[\gamma_G - \psi_G(1-\gamma_G)\right]$$

But this is a contradiction because $\gamma_G - \psi_G(1-\gamma_G) \in (0,1)$ and $\gamma_I - \psi_I(1-\gamma_I) \in (0,1)$. □

Proof of Proposition 5:

Proof. We begin by rewriting Conditions (12) and (13) using the accuracy specification $\psi_G = \psi + \varepsilon$ and $\psi_I = \psi - \varepsilon$, for $\varepsilon \geq 0$.

Conditions First, plug in $\psi + \varepsilon$ for ψ_G and $\psi - \varepsilon$ for ψ_I and rearrange Condition 8 to obtain that the accuracy bound is now

$$\varepsilon < (1-\psi) \cdot \frac{\gamma_G - \gamma_I}{1 - \gamma_G + 1 - \gamma_I} \equiv \overline{\varepsilon}$$

Second, do the same for Condition 9 to obtain that

$$\frac{\gamma_G}{\gamma_I} > \frac{(\psi+\varepsilon)(1-(\psi-\varepsilon))}{(\psi-\varepsilon)(1-(\psi+\varepsilon))}$$

Defining $\gamma \equiv \frac{\gamma_G}{\gamma_I}$, this condition is equivalent to

$$F(\varepsilon) \equiv (\gamma-1)\varepsilon^2 + \varepsilon(\gamma+1) + \psi(1-\psi)(\gamma-1) > 0$$

F is a quadratic function with roots:

$$\varepsilon_1 = \frac{\gamma+1-\sqrt{(\gamma+1)^2 - 4\psi(1-\psi)(\gamma-1)^2}}{2(\gamma-1)}$$

and

$$\varepsilon_2 = \frac{\gamma+1+\sqrt{(\gamma+1)^2 - 4\psi(1-\psi)(\gamma-1)^2}}{2(\gamma-1)}$$

However, it is immediate that only ε_1 satisfies the Condition 8 bound on accuracy, $\varepsilon < \overline{\varepsilon}$. Thus, the court features Condition 9's *low accuracy* if $\varepsilon < \varepsilon_1$, and *high accuracy* if $\varepsilon > \varepsilon_1$.

We first investigate how the incumbent's calculus changes for a given citizen support region when court accuracy changes. We then analyze how the citizen support regions change as court accuracy changes.

Change in Conviction Probability. Intervention's effect on the probability of conviction is given by:

$$\begin{aligned}
\Delta^\alpha &= p\psi_G(1-\gamma_G) + (1-p)\psi_I(1-\gamma_I) \\
&= p(\psi+\varepsilon)(1-\gamma_G) + (1-p)(\psi-\varepsilon)(1-\gamma_I)
\end{aligned}$$

Then:

$$\frac{\partial \Delta^{\alpha}}{\partial \varepsilon} = p(1-\gamma_G) - (1-p)(1-\gamma_I)$$

Thus accuracy increases the size of this effect if $p > \frac{1-\gamma_I}{1-\gamma_I+1-\gamma_G}$ and decreases it otherwise.

Costs of Interference. The level of intervention that guarantees the prosecutor acts after receiving the innocent signal is given by:

$$\begin{aligned}\lambda^F &= \psi_I q - \Pr(G|i)\left[\psi_I q + \psi_G(1-q)\right] \\ &= (\psi-\varepsilon)q - \Pr(G|i)\left[(\psi-\varepsilon)q + (\psi+\varepsilon)(1-q)\right]\end{aligned}$$

Then:

$$\frac{\partial \lambda^F}{\partial \varepsilon} = -q - \Pr(G|i)\left[-q+1-q\right] < 0$$

To prove that this is indeed negative, suppose not and rearrange to obtain:

$$\frac{2q-1}{q} \cdot \Pr(G|i) \geq 1.$$

This is a contradiction since $\frac{2q-1}{q} \in (-\infty, 1)$ and $\Pr(G|i) \in (0,1)$. Because K is assumed to be increasing, an increase in court accuracy always decreases the costs of interference.

Thresholds: Low Accuracy. Recall from the main text that the thresholds for intervention are as follows:

$$T_{MI} = B\left[1 - p\gamma_G - (1-p)\gamma_I\right]$$

$$T_{MO} = B\left[-p\gamma_G - (1-p)\gamma_I\right]$$

$$T_{II} = B\left[p\left(\psi_G - \gamma_G\right) + (1-p)\left(\psi_I - \gamma_I\right)\right] = B\left[p\left((\psi+\varepsilon) - \gamma_G\right) + (1-p)\left((\psi-\varepsilon) - \gamma_I\right)\right]$$

$$T_{IO} = B\left[-p\psi_G\gamma_G - (1-p)\psi_I\gamma_I\right] = B\left[-p(\psi+\varepsilon)\gamma_G - (1-p)(\psi-\varepsilon)\gamma_I\right]$$

Evidently:

$$\frac{\partial T_{MI}}{\partial \varepsilon} = 0$$
$$\frac{\partial T_{MO}}{\partial \varepsilon} = 0$$
$$\frac{\partial T_{II}}{\partial \varepsilon} = B(2p-1) > 0 \text{ iff } p > \frac{1}{2}$$
$$\frac{\partial T_{IO}}{\partial \varepsilon} = B\left[-p\gamma_G + (1-p)\gamma_I\right] > 0 \text{ iff } p < \frac{\gamma_I}{\gamma_I + \gamma_G}$$

Thus, accuracy ε either has no effect or the sign of the effect depends on other parameters.

Thresholds: High Accuracy. The only distinct threshold is:

$$\begin{aligned} T_O &= B\left[p\psi_G(1-\gamma_G) + (1-p)\psi_I(1-\gamma_I)\right] \\ &= B\left[p(\psi+\varepsilon)(1-\gamma_G) + (1-p)(\psi-\varepsilon)(1-\gamma_I)\right] \end{aligned}$$

Then:

$$\frac{\partial T_O}{\partial \varepsilon} = B\left[p(1-\gamma_G) - (1-p)(1-\gamma_I)\right] > 0 \text{ iff } p > \frac{1-\gamma_I}{1-\gamma_I+1-\gamma_G}$$

To summarize, supposing for the time being that a change in ε does not affect the citizen support region (see below), for the low accuracy case:

- If $\mu \leq p^N(0,0)$ or $\mu > p^N(1,1)$, an increase in ε does not affect the threshold for interference (0), decreases the costs of interference, and can increase or decrease the size of intervention's effect on conviction probability. The effect is thus ambiguous.
- If $\mu \in \left(p^N(0,0), p^F(0)\right]$, an increase in ε does not affect T_{MI}, decreases the costs of interference, and can increase or decrease the size of intervention's effect on conviction probability. The effect is thus ambiguous.
- If $\mu \in \left(p^F(0), p^F(1)\right]$, an increase in ε can increase or decrease T_{II}, decreases the costs of interference, and can increase or decrease the size of intervention's effect on conviction probability. The effect is thus ambiguous.
- If $\mu \in \left(p^F(1), p^N(1,0)\right]$, an increase in ε does not affect T_{MO}, decreases the costs of interference, and can increase or decrease the size of intervention's effect on

conviction probability. The effect is thus ambiguous.

- If $\mu \in \big(p^N(1,0), p^N(1,1)\big]$, an increase in ε can increase or decrease T_{IO}, decreases the costs of interference, and can increase or decrease the size of intervention's effect on conviction probability. The effect is thus ambiguous.

For the high accuracy case:

- If $\mu \leq p^N(0,0)$ or $\mu > p^N(1,1)$, an increase in ε does not affect the threshold for interference (0), decreases the costs of interference, and can increase or decrease the size of intervention's effect on conviction probability. The effect is thus ambiguous.
- If $\mu \in \big(p^N(0,0), p^F(0)\big]$,an increase in ε does not affect T_{MI}, decreases the costs of interference, and can increase or decrease the size of intervention's effect on conviction probability.The effect is thus ambiguous.
- If $\mu \in \big(p^F(0), p^N(1,0)\big]$, an increase in ε can increase or decrease T_{II}, decreases the costs of interference, and can increase or decrease the size of intervention's effect on conviction probability. The effect is thus ambiguous.
- If $\mu \in \big(p^N(1,0), p^F(1)\big]$, an increase in ε can increase or decrease T_O, decreases the costs of interference, band can increase or decrease the size of intervention's effect on conviction probability.
- If $\mu \in \big(p^F(1), p^N(1,1)\big]$, an increase in ε can increase or decrease T_{IO}, decreases the costs of interference, and can increase or decrease the size of intervention's effect on conviction probability. The effect is thus ambiguous.

Figure 4 summarizes the results graphically. It displays the regions in which an increase in court accuracy ε unequivocally increases the attractiveness of full intervention (white regions), or has competing effects on intervention (shaded regions), as a function of the relative popularity of the opposition (μ) and the prior probability that the target is guilty (p). As the figure shows, if the prior probability is relatively large ($p > \frac{1-\gamma_I}{1-\gamma_I+1-\gamma_G}$) and μ is *not* between $p^N(0,1)$ and $p^N(1,1)$ ($p^F(1)$ and $p^N(1,1)$) for the low (high) accuracy court, an increase in court accuracy increases the attractiveness of intervention. Otherwise, the there are competing effects. First, if $p < \frac{1-\gamma_I}{1-\gamma_I+1-\gamma_G}$, court accuracy makes

Figure 4: Effect of an Increase in Court Accuracy on the Attractiveness of an Intervention. Shaded Regions: Competeting Effects; White Regions: Interventions Increase. $p_1 \equiv \frac{\gamma_I}{\gamma_I+\gamma_G} < \frac{1}{2}$ and $p_2 \equiv \frac{1-\gamma_I}{1-\gamma_I+1-\gamma_G} > \frac{1}{2}$. Parameter values are the same as in Figure 3.

it less likely that a target is convicted. Second, depending on the value of μ, the threshold T may or may not be affected. Of particular importance is T_{IO} which increases in ε if the prior is *low*, which explains the competing effects for the cases in which μ and p are relatively high.

Beliefs. The above establishes that for all μ, the effect of accuracy on intervention is ambiguous. However, for completeness, we now consider how the citizen's posterior beliefs change when court accuracy, ε, improves. First, note that $p^N(0,0) = \frac{p(1-\gamma_G)}{p(1-\gamma_G)+(1-p)(1-\gamma_I)}$ is independent of court accuracy. We show that $p^F(1)$ and $p^N(1,1)$ are increasing court accuracy while $p^F(0)$ and $p^N(1,0)$ are decreasing in court accuracy:

$$
\begin{aligned}
p^F(1) &= \frac{p\psi_G}{p\psi_G + (1-p)\psi_I} = \frac{p(\psi+\varepsilon)}{p(\psi+\varepsilon)+(1-p)(\psi-\varepsilon)} \\
p^N(1,1) &= \frac{p\psi_G\gamma_G}{p\psi_G\gamma_G + (1-p)\psi_I\gamma_I} = \frac{p(\psi+\varepsilon)\gamma_G}{p(\psi+\varepsilon)\gamma_G+(1-p)(\psi-\varepsilon)\gamma_I} \\
p^F(0) &= \frac{p(1-\psi_G)}{p(1-\psi_G) + (1-p)(1-\psi_I)} = \frac{p(1-\psi-\varepsilon)}{p(1-\psi-\varepsilon)+(1-p)(1-\psi+\varepsilon)} \\
p^N(1,0) &= \frac{p(1-\psi_G)\gamma_G}{p(1-\psi_G)\gamma_G + (1-p)(1-\psi_I)\gamma_I} = \frac{p(1-\psi-\varepsilon)\gamma_G}{p(1-\psi-\varepsilon)\gamma_G+(1-p)(1-\psi+\varepsilon)\gamma_I}
\end{aligned}
$$

Then:

$$\begin{aligned}
\frac{\partial p^F(1)}{\partial \varepsilon} &= \frac{2\psi p(1-p)}{\left[p(\psi+\varepsilon)+(1-p)(\psi-\varepsilon)\right]^2} > 0 \\
\frac{\partial p^N(1,1)}{\partial \varepsilon} &= \frac{2\psi p(1-p)\gamma_I\gamma_G}{\left[p(\psi+\varepsilon)\gamma_G+(1-p)(\psi-\varepsilon)\gamma_I\right]^2} > 0 \\
\frac{\partial p^F(0)}{\partial \varepsilon} &= \frac{-2(1-\psi)p(1-p)}{\left[p(1-\psi-\varepsilon)+(1-p)(1-\psi+\varepsilon)\right]^2} < 0 \\
\frac{\partial p^N(1,0)}{\partial \varepsilon} &= \frac{-2(1-\psi)p(1-p)\gamma_I\gamma_G}{\left[p(1-\psi-\varepsilon)\gamma_G+(1-p)(1-\psi+\varepsilon)\gamma_I\right]^2} < 0
\end{aligned}$$

Summarizing, an increase in court accuracy can alter the size of the various citizen support regions in both the high and low accuracy cases. (It is also clear from the above that an increase can alter which accuracy case applies.) □

B Robustness

B.1 Partially Observed Interference

Technology. We consider the following technology of observability: If $\lambda = 0$, then it is unobserved with probability 1. If $\lambda > 0$, it is observed with probability φ and unobserved with probability $1-\varphi$. In other words:

$$Pr\left(\lambda \text{ observed}|\lambda\right) = \begin{cases} 0 & \text{if } \lambda = 0 \\ \varphi & \text{if } \lambda > 0 \end{cases} \tag{13}$$

The baseline case is equivalent to a situation in which $\varphi = 1$, and if $\varphi = 0$, then all levels of interference are unobserved.

Analysis Because the prosecutor can observe the incumbent's choice, his strategy is the same as before. He acts, $a = 1$, if

$$\lambda \geq \psi\left[q - \Pr(G|s)\right].$$

However, since λ may be unobserved, the meaning of this action may not be immediately clear to the citizen.

We assume that the incumbent must choose $\lambda \in \{0, \lambda^F\}$. This is to avoid counterintuitive situations (byproducts of the stark observability technology we employ) in which the incumbent makes an arbitrarily small deviation from zero in an attempt to reveal prosecutorial independence (because such a small deviation would not change the prosecutor's action but would be observable with probability φ).

To assess the robustness of the analysis in the main text, we search for an equilibrium in which the incumbent chooses no interference, i.e., $\lambda = 0$. In such a profile, the incumbent's expected utility is:

$$\Pr(s = g)\left[\alpha\psi + \mathbb{1}(\Pr(G|g) \geq \mu)B\right] + \Pr(s = i)\mathbb{1}(\Pr(G|i) \geq \mu)B$$

This is because, when the citizen expects no intervention, her posteriors after observing a are:

$$\begin{aligned}
\Pr(G|a = 1) &= \frac{p\gamma_G}{p\gamma_G + (1-p)\gamma_I} = \Pr(G|g) \\
\Pr(G|a = 0) &= \frac{p(1-\gamma_G)}{p(1-\gamma_G) + (1-p)(1-\gamma_I)} = \Pr(G|i)
\end{aligned}$$

The expected utility from deviating to full interference, i.e., $\lambda = \lambda^F$, is:

$$\varphi[\alpha\psi + \mathbb{1}(p \geq \mu)B] + (1-\varphi)[\alpha\psi + \mathbb{1}(\Pr(G|g) \geq \mu)B] - K\left(\lambda^F\right)$$

With probability φ, the deviation is observed, and the citizen correctly infers that prosecutorial action no longer conveys information; with probability $1-\varphi$, the deviation is not observed, and the citizen believes that prosecutorial action still conveys information.

We have four cases, depending on the size of μ:

Case 1. Suppose that $\mu \leq \Pr(G|i)$. The incumbent does not deviate if:

$$\begin{aligned}
\Pr(s = g)\alpha\psi + B &\geq \alpha\psi + B - K(\lambda^F) \\
K(\lambda^F) - \alpha\psi\Pr(s = i) &\geq 0
\end{aligned}$$

This is a condition analogous to the one derived in the main text.

Case 2. Suppose that $\mu \in (\Pr(G|i), p]$. The incumbent does not deviate if:

$$\Pr(s = g)\,(\alpha\psi + B) \geq \alpha\psi + B - K(\lambda^F)$$
$$K(\lambda^F) - \alpha\psi \Pr(s = i) \geq B \Pr(s = i)$$

This is a condition analogous to the one derived in the main text.

Case 3. Suppose that $\mu \in (p, \Pr(G|g)]$. The incumbent does not deviate if:

$$\Pr(s = g)\,(\alpha\psi + B) \geq \alpha\psi + (1 - \varphi)B - K(\lambda^F)$$
$$K(\lambda^F) - \alpha\psi \Pr(s = i) \geq B\,[(1 - \varphi) - \Pr(s = g)]$$

This is a generalization of the condition derived in the main text. Rather than the informational costs emphasized in the main text, deviating to full interference could have a benefit if the probability of discovery is sufficiently low. However, deviating to full interference still carries a cost if the probability of interference being observable is sufficiently high, i.e., $\varphi > \Pr(i)$.

Case 4. Suppose that $\mu > \Pr(G|g)$. The incumbent does not deviate if:

$$\Pr(s = g)\alpha\psi \geq \alpha\psi - K(\lambda^F)$$
$$K(\lambda^F) - \alpha\psi \Pr(s = i) \geq 0$$

This is a condition analogous to the one derived in the main text.

Summarizing, our findings are broadly robust to making interference imperfectly observed. The only substantively interesting effect occurs if the citizen is moderately biased towards the opposition, so that there is an informational cost to interference when interference is observable. If it is partially observable, the informational costs becomes smaller and can even turn into a benefit if the probability of observing interference is sufficiently low. (Notice that in a larger game, an incumbent might want to commit to engaging only in fully observable interference so that when she refrains from interference, citizens are certain that the prosecutor's decisions are informative.)

B.2 Citizen Punishment

Suppose that the citizen intrinsically dislikes interference *and* can commit to punishing it. Specifically, suppose that μ is an increasing function of interference; the important quantity is $\mu\left(\lambda^F\right) > \mu(0)$.

The incumbent decides between full and no intervention—no other intervention level can be optimal. Her expected utility from full intervention is:

$$\alpha\psi + \mathbb{1}\left(p^F \geq \mu\left(\lambda^F\right)\right)B - K\left(\lambda^F\right)$$

Her expected utility from nonintervention is:

$$\Pr(g)\left[\alpha\psi + \mathbb{1}\left(p^N(1) \geq \mu(0)\right)B\right] + \Pr(i)\mathbb{1}\left(p^N(0) \geq \mu(0)\right)B$$

Obviously, the attractiveness of intervention now also depends on its effect on the citizen's bias, i.e., the extent to which $\mu\left(\lambda^F\right)$ differs from $\mu(0)$. Basically, if either the citizen's bias for the opposition is already fairly large ($\mu > p$) or if the change in the citizen's preferences after observing interference is fairly small, the equilibrium is unchanged, and the relevant thresholds remain T_{MO} and 0, respectively. However, if the citizen is predisposed toward the incumbent and punishment is strong, the equilibrium can change. Specifically, there is now an additional cost to interference because the citizen is more biased towards the opposition.

Specifically, suppose that $\mu(0) \leq p^N(0)$ but $\mu(\lambda^F) > p^F$. In this case, the incumbent loses citizen support by intervening. The incumbent nevertheless intervenes if:

$$T_E \equiv -B \geq K(\lambda^F) - \alpha\psi\Pr(i)$$

The other interesting case occurs when $\mu(0) \in \left(p^N(0), p^F\right]$ and $\mu(\lambda^F) > p^F$. Fully intervening yields $\alpha\psi - K(\lambda^F)$ while not intervening yields $\Pr(g)(\alpha\psi + B)$. Hence, intervention

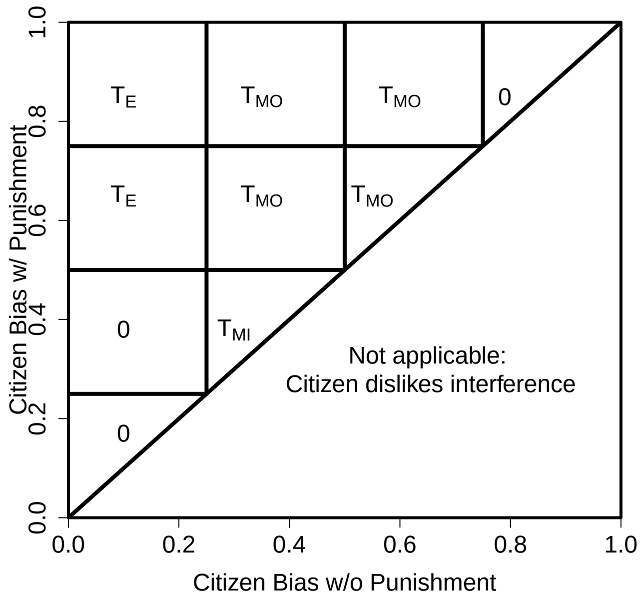

Figure 5: Overview of equilibrium thresholds with citizen punishment

is optimal if

$$T_{MO} = -B\Pr(g) \geq K(\lambda^F) - \alpha\psi\Pr(i)$$

Thus, in contrast to the case when there is no punishment, there is now an informational cost associated with intervention.

Figure 5 gives an overview of the equilibrium thresholds for intervention for any combination of $\mu(0)$ and $\mu\left(\lambda^F\right)$.

B.3 Incumbent Protects Ally

In this section, we analyze a situation in which the incumbent wishes to protect an ally from prosecution, but does not know whether the ally is guilty or not. To incorporate this preference, we change the utility functions as follows. For the incumbent:

$$U_{\text{Inc}} = \alpha(1 - C) + rB - K(\lambda)$$

For the prosecutor:

$$U_P = u_{C\theta}(q) + (1 - a)\lambda$$

Finally, the citizen supports the incumbent if and only if $\Pr(G|\cdot) \leq \mu$.

Comparing the prosecutor's expected utility of acting and not acting, we find that the prosecutor chooses not to act if:

$$\lambda \geq \psi \left[\Pr(G|s) - q\right]$$

As before, when there is no intervention, $\lambda = 0$, the prosecutor acts only if he receives the guilty signal. Therefore, the incumbent can choose to interfere fully, with full intervention now defined by $\lambda^F \equiv \psi \left[\Pr(G|g) - q\right]$, or not at all: $\lambda = 0$.

If the incumbent fully interferes, her expected utility is:

$$\alpha + \mathbb{1}(\mu \geq p)B - K\left(\lambda^F\right)$$

If she does not interfere at all, she receives:

$$\Pr(s = g)\left[\alpha(1 - \psi) + \mathbb{1}(\mu \geq \Pr(G|g))B\right] + \Pr(s = i)\left[\alpha + \mathbb{1}(\mu \geq \Pr(G|i)B\right]$$

We have four cases:

Case 1. Suppose that $\mu < \Pr(G|i)$ so that the citizen never supports the incumbent. The incumbent chooses λ^F if

$$\alpha - K(\lambda^F) \geq \alpha\left[\Pr(g)(1 - \psi) + \Pr(i)\right]$$
$$0 \geq K(\lambda^F) - \alpha\psi\Pr(g)$$

This is analogous to the condition derived in the main text.

Case 2. Suppose that $\mu \in [\Pr(G|i), p)$ so that the citizen supports the incumbent if she becomes aware of the innocent signal. The incumbent chooses λ^F if

$$\alpha - K(\lambda^F) \geq \Pr(g)(1 - \psi)\alpha + \Pr(i)\left(\alpha + B\right)$$
$$-B\Pr(i) \geq K(\lambda^F) - \alpha\psi\Pr(g)$$

This is analogous to the condition derived in the main text: there is now an informational cost to interference (the left-hand side is negative).

Case 3. Suppose that $\mu \in [p, \Pr(G|g),)$ so that the citizen supports the incumbent unless she becomes aware of the innocent signal. The incumbent chooses λ^F if

$$\alpha + B - K(\lambda^F) \geq \Pr(g)(1-\psi)\alpha + \Pr(i)\,(\alpha + B)$$
$$B\Pr(g) \geq K(\lambda^F) - \alpha\psi\Pr(g)$$

This is analogous to the condition derived in the main text: there is now an informational benefit to interference (the left-hand side is positive).

Case 4. Suppose that $\mu \geq \Pr(G|g)$ so that the citizen always supports the incumbent. The incumbent chooses λ^F if

$$\alpha + B - K(\lambda^F) \geq (\alpha + B)\,[\Pr(g)(1-\psi) + \Pr(i)]$$
$$0 \geq K(\lambda^F) - \alpha\psi\Pr(g)$$

This is analogous to the condition derived in the main text.

In sum, the equilibrium outcomes are analogous to the analysis in the baseline case: when public opinion is solidly anti- or pro-incumbent, information does not matter and the incumbent decides on interference based on the costs and the decrease in the probability of getting a bad outcome (here: conviction). The highest incentives to interfere occur if the incumbent is moderately popular (there is a benefit to suppressing unfavorable information) while the lowest incentives occur if the incumbent is moderately unpopular (there is a cost to suppressing potentially helpful information).

C Additional Results

C.1 Citizen Welfare

In the main text, we simply assume that the citizen reelects the incumbent whenever her posterior is greater than the parameter μ. We now microfound this behavior and conduct a welfare analysis. Specifically, we assume that the citizen's utility function is:

$$U_V = u_{\theta C}(z) - M(\lambda) + r v_{\text{Inc}} + (1-r)[\mathbb{1}(\theta = G) v_O(G) + \mathbb{1}(\theta = I) v_O(I)] \tag{14}$$

where M increasing in λ represents the citizen's inherent dislike for interference, and $u_{\theta C}(z)$ is of the same form as the prosecutor's utility function, with z (rather than q) representing the citizen's concern for convicting the innocent.

When the citizen chooses whether to support the opposition ($r = 0$) or not ($r = 1$), C and λ are already determined. They are therefore irrelevant for her calculus: the citizen exclusively focuses on the change in utility associated with her support decision. Specifically, the expected utility from supporting the uncumbent is

$$\mathbb{E}\left[U_V(r = 1)\right] = v_{\text{Inc}}$$

whereas that from supporting the opposition is

$$\mathbb{E}\left[U_V(r = 0)\right] = \Pr(G|\cdot) v_O(G) + (1 - \Pr(G|\cdot)) v_O(I)$$

If $v_O(I) > v_{\text{Inc}} > v_O(G)$ (the citizen's preferences are state-dependent), rearranging yields that the citizen supports the incumbent ($r = 1$) if and only if

$$\Pr(G|\cdot) \geq \frac{v_O(I) - v_{\text{Inc}}}{v_O(I) - v_O(G)} \equiv \mu$$

Therefore, our assumed decision rule is consistent with a citizen's expected utility maximization of a utility function like the one described by Expression 14 above.

We now conduct a welfare analysis of the baseline model. Specifically, we compare the citizen's equilibrium utility when the incumbent fully interferences, $\lambda = \lambda^F$, and when the incumbent does not interfere, $\lambda = 0$. Under full interference, the citizen's welfare is:

$$p\left[-(1-\psi)(1-z)\right] + (1-p)(-\psi z) - M\left(\lambda^F\right) + pv_O(G) + (1-p)v_O(I) \equiv W^V_{\text{full}}$$

By contrast, under noninterference, the citizen's welfare is:

$$\begin{aligned}\Pr(s=g)\left[\Pr(G|g)\left(-(1-\psi)(1-z)\right) + (1-\Pr(G|g))(-\psi z) + v_{\text{Inc}}\right] + \\ \Pr(s=i)\left[\Pr(G|i)\left(-(1-z)\right) + \Pr(G|i)v_O(G) + (1-\Pr(G|i))v_O(I)\right] \equiv W^V_{\text{no}}\end{aligned}$$

Inspecting the inequality $W^V_{\text{no}} \geq W^V_{\text{full}}$ yields that the citizen is better off when there is no interference if and only if:

$$\overbrace{M\left(\lambda^F\right)}^{\text{Costs Interference}} + \psi \underbrace{\left(z\Pr(s=i) - p(1-\gamma_G)\right)}_{\text{Different Outcome } C} \geq \underbrace{p\gamma_G\left[v_O(G) - v_{\text{Inc}}\right] + (1-p)\gamma_I\left[v_O(I) - v_{\text{Inc}}\right]}_{\text{Different Support}}$$

The costs of interference, $M\left(\lambda^F\right)$, are always positive. By contrast, the signs of the terms "Different Outcome C" and "Different Support" are ambiguous. The former is positive if and only if $z \geq \frac{p(1-\gamma_G)}{\Pr(s=i)}$. The latter is positive when the citizen's utility from supporting the incumbent is sufficiently low (v_{Inc} is close to $v_O(G)$) and negative when the citizen's utility from supporting the incumbent is sufficiently high (v_{Inc} is close to $v_O(I)$).

Thus, the citizen is better off under nonintervention if, for example, the concern for convicting the innocent is sufficiently high, $z \geq z^*$, where z^* is given by

$$z \geq \frac{p\gamma_G[v_O(G) - v_{\text{Inc}}] + (1-p)\gamma_I[v_O(I) - v_{\text{Inc}}] + \psi p(1-\gamma_G) - M(\lambda^F)}{\psi \Pr(s=i)} \equiv z^*$$

This is illustrated by Figure 6, in which we plot the citizen's welfare under no and full interference as a function of her concern for type I and II errors (i.e., $W^V_{\text{full}}(z)$ and $W^V_{\text{no}}(z)$).

To summarize, the citizen is not always better off under nonintervention. First, while a citizen who cares a great deal about shielding the innocent prefers nonintervention, if

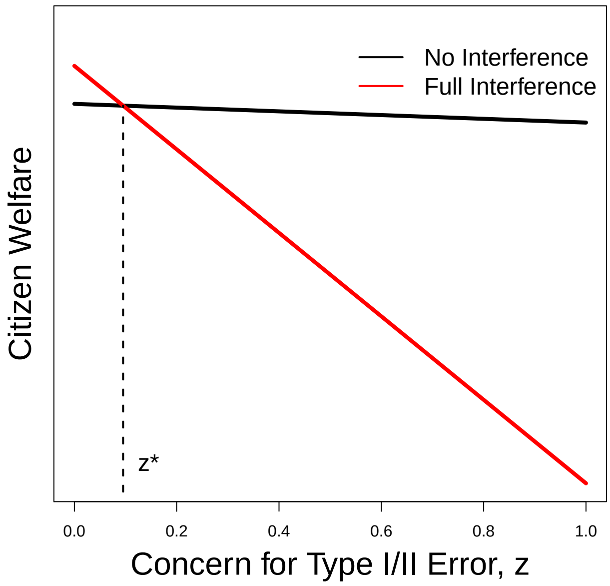

Figure 6: The citizen's welfare. Parameter values: $p = 0.3$, $\gamma_G = 0.76$, $\gamma_I = 0.25$, $\psi = 0.8$, $v_O(G) = -0.1$, $v_O(I) = 1.5$, $v_{\text{Inc}} = 0.6$, and $M(\lambda^F) = 0.01$.

instead she cares a great deal about punishing the guilty, she is happy for the prosecutor to act under all circumstances, even at the cost of some information about the opponent. Second, if the citizen is relatively happy (ex-ante) with the incumbent, the risk that she will be forced to switch to the opposition under more information (i.e., under no interference) is relatively low, so she prefers nonintervention. However, if her ex-ante utility from supporting the incumbent is relatively low, sticking with the default choice of supporting the opposition is better from an ex ante perspective and so the citizen's welfare is higher when there is full interference.

C.2 Mixed Strategy Equilibria

In the main text, we focused on pure strategy equilibria due to their intuitiveness. Here, we consider mixed strategy equilibria for completeness. We cover both the baseline case and the extension in which the prosecutor is political, i.e., obtains $b > 0$ when the citizen supports the incumbent. We assume that $\mu \neq \Pr(G|s)$ for all $s \in \{g, i\}$, so that generically, if the prosecutor plays a pure strategy, the citizen does so as well.

Preliminaries. Let $\tau(a)$ be the probability that the citizen supports the incumbent as a function of prosecutorial action a. Let β_s be the probability of prosecutorial action conditional on signal s. Note that for the prosecutor who sees the guilty signal, action strictly dominates inaction. Hence, $\beta_g = 1$ in any equilibrium. For the prosecutor who sees the innocent signal to be willing to mix, he must be indifferent. This can happen for two reasons. First, the level of intervention may be such that the prosecutor is indifferent (e.g., full intervention). Second, in the political prosecutor case, the citizen may keep the prosecutor indifferent by mixing between supporting and not supporting the incumbent.

Given $\beta_g = 1$ and $\beta_i \in (0,1)$, the citizen's beliefs are as follows:

$$\Pr(G|1, \beta_i) = \frac{p(\gamma_G + (1-\gamma_G)\beta_i)}{Pr(s=g) + \beta_i Pr(s=i)}$$
$$\Pr(G|0, \beta_i) = \frac{p(1-\gamma_G)}{\Pr(s=i)}$$

Observe that $\Pr(G|0, \beta_i)$ is independent of β_i. First consider when $a = 0$. Here, the only way that the citizen can be indifferent is if μ is exactly equal to $\Pr(G|i)$—this is a knife-edge case and as mentioned above, we ignore it. Now consider when $a = 1$. $\Pr(G|1, \beta_i)$ ranges from p to $\Pr(G|g)$, we therefore assume that the citizen's bias μ is in this interval.

Baseline If $b = 0$, then the type i-prosecutor is indifferent only if:

$$\lambda = \psi\left[q - \Pr(G|i)\right] \Rightarrow \lambda = \lambda^F$$

In this case, any $\beta_i \in [0,1]$ is an equilibrium for the citizen-prosecutor subgame. Define β_i^* to be the probability that keeps the citizen indifferent between supporting the opposition and supporting the incumbent:

$$\Pr(G|1, \beta_i^*) = \mu \Rightarrow \beta_i^* = \frac{p\gamma_G - \mu\Pr(s=g)}{\mu\Pr(s=i) - p(1-\gamma_G)}$$

Then, there are three relevant cases:

1. $\beta_i < \beta_i^*$ which implies $\tau(1) = 1$.
2. $\beta_i = \beta_i^*$ which implies $\tau(1) \in [0,1]$.
3. $\beta_i > \beta_i^*$ which implies $\tau(1) = 0$.

For any of these cases, there are two possible equilibrium outcomes $\lambda = 0$ and $\lambda = \lambda^F$. Note, however, that whichever option is better must also be weakly preferred to the deviation $\lambda + t$ where t is arbitrarily small. This choice induces $a^*(s) = 1$ for all s and $\tau(1) = 0$. Thus, there is mixing on the path of play if two conditions are met (under some conditions, there might be mixing off-the-path). First, the incumbent's expected utility from $\lambda = \lambda^F$must be larger than her expected utility from $\lambda = 0$, i.e.,

$$(\Pr(g) + \Pr(i)\beta_i) [\alpha\psi + \tau(1, \beta_i)B] - K \left(\lambda^F\right) \geq \Pr(g) (\alpha\psi + B)$$

where $\tau(1, \beta_i)$ depends on β_i as explained in the three cases above. Second, the expected utility of λ^F also needs to be larger than the expected utility of deviating to $\lambda^F + t$:

$$(\Pr(g) + \Pr(i)\beta_i) [\alpha\psi + \tau(1, \beta_i)B] - K \left(\lambda^F\right) \geq \alpha\psi - K \left(\lambda^F + t\right)$$

Re-arranging this condition yields:

$$K \left(\lambda^F + t\right) - K \left(\lambda^F\right) \geq \alpha\psi \Pr(i)(1 - \beta_i) - \tau(1, \beta_i)B [\Pr(g) + \Pr(i)\beta_i]$$

The left-hand side of the preceding inequality converges to 0 as t becomes arbitrarily small (if K is continuous). Consequently, for a deviation *not* to be profitable, the right-hand side needs to be negative, i.e.,

$$\tau(1, \beta_i) > \frac{\alpha\psi}{B} \frac{\Pr(i)(1 - \beta_i)}{\Pr(g) + \Pr(i)\beta_i}.$$

In other words, the probability with which the citizen supports the incumbent needs to be sufficiently high. Given that the right-hand side of this inequality is positive, a situation in which $\beta_i > \beta_i^*$ is ruled out. However, depending on parameter values, it may be that $\beta_i < \beta_i^*$ or $\beta_i = \beta_i^*$. From the incumbent's perspective, the best strategy is $\beta_i = \beta_i^*$ and $\tau(1, \beta_i^*) = 1$ because this always persuades the citizen while making the outcomes $a = 1$ and $C = 1$ as likely as possible (this is the solution obtained in Kamenica and Gentzkow, 2011). However, this is equilibrium selection—the incumbent cannot induce the prosecutor to choose this particular probability.

Low Politicization Now consider the case when $b \in \left(0, \lambda^F\right)$. Indifference of the i-type prosecutor now requires:

$$\lambda + \tau(1)b = \psi\left[q - \Pr(G|i)\right]$$

Given that $b > 0$, the prosecutor may be kept indifferent by the citizen's strategy. Specifically:

$$\tau(1, \lambda) = \frac{\psi\left[q - \Pr(G|i)\right] - \lambda}{b}$$

For this to be a proper probability, two conditions have to be met: $\lambda \leq \lambda^F$ and $b \geq \lambda^F - \lambda$ (with strict inequalities for an interior probability). Note that τ is decreasing in λ. Intuitively, this is because the prosecutor's incentives to act increase when there is more interference; to maintain indifference, the citizen's probability of support for the incumbent must decrease. Thus, the citizen punishes interference without intrinsically caring about it.

For a level of interference that satisfies $\lambda \leq \lambda^F$ and $b \geq \lambda^F - \lambda$, the incumbent's expected utility is then:

$$\Pr(g)\left[\alpha\psi + \tau(1, \lambda)B\right] + \Pr(i) \cdot \beta_i\left[\alpha\psi + \tau(1, \lambda)B\right] - K(\lambda)$$

This is decreasing in interference λ. Define

$$\lambda^M \equiv \lambda^F - b.$$

The incumbent has three possible optimal choices:

1. $\lambda = 0$ which induces $\tau(1) = 1$ and $\beta_i = 0$ (the i-type cannot be made indifferent when $\lambda = 0$).
2. $\lambda = \lambda^M$ which induces $\tau(1, \lambda^M) = 1$ and β_i^*.
3. $\lambda = \lambda^F$ which induces $\tau(1) = 0$ and $\beta_i \in (\beta_i^*, 1]$.

Moreover, whichever choice is best among these three choices also needs to be weakly better than the deviation $\lambda = \lambda^F + t$ which induces $\tau(1) = 0$ and $\beta_i = 1$. By inspection, this means that the third option, $\lambda = \lambda^F$, can only be optimal if $\beta_i = 1$—otherwise, the incumbent would deviate to $\lambda^F + t$ for t small.

A comparison of these three candidates (0, λ^M, and λ^F) then reveals the optimal choice. Depending on the curvature of K and the size of B relative to $\alpha\psi$, each of these choices can be optimal.

High Politicization In this case, a pure strategy separating equilibrium for the citizen-prosecutor interaction does not exist, because a prosecutor who receives the innocent signal would pretend to have received the guilty signal in order to gain citizen support for the incumbent. However, there may be a semi-separating equilibrium in which the prosecutor mixes when receiving the innocent signal, and the citizen mixes when observing prosecutorial action. Specifically, $\lambda = 0$ is compatible with a fully mixed equilibrium in which the prosecutor chooses β_i^* and the citizen supports with probability $\tau(1, 0) = \frac{\lambda^F}{b}$.

The incumbent can also choose $\lambda = \lambda^F$ to receive:

$$\alpha\psi - K\left(\lambda^F\right).$$

For this to be optimal, the prosecutor must choose $\beta_i = 1$ since otherwise the incumbent deviates to $\lambda^F + t$. Nonintervention is optimal if:

$$\left[\Pr(g) + \Pr(i)\beta_i^*\right]\tau(1, 0)B \geq \alpha\psi\Pr(i)\left(1 - \beta_i^*\right) - K\left(\lambda^F\right).$$

Summary Examining mixed strategy equilibria yields several insights, although the pure strategy equilibrium analyzed in the main text seems substantively more plausible. First, with a highly politicized prosecutor, mixing re-establishes the possibility of (partial) citizen learning because it reduces the prosecutor's temptation to deviate after seeing the innocent signal, sometimes making the incumbent better off. Second, the incumbent can also be better off when even an apolitical prosecutor is allowed to mix after observing $s = i$, since this increases the probability of conviction while still persuading the citizen.

Third, when the citizen mixes to maintain a political prosecutor's indifference, she does so with a probability that is decreasing in interference, counterbalancing the prosecutor's increased incentives to act.

D Extension: Endogenous Effort

Although prosecutors may sometimes simply rely on information provided by third party reports (e.g., the police) when deciding whether to act, they often do exert costly effort to learn about a target's guilt. To account for this, we assume that the probability that the prosecutor receives a signal is equal to effort $e \in [0, 1]$ endogenously chosen at cost $C(e)$ increasing and convex. For simplicity, we consider a situation in which the signal is completely informative, i.e., $\gamma_G = 1$ and $\gamma_I = 0$. With probability $1 - e$, the prosecutor gets an uninformative signal, $s = \varnothing$.

Denote by $\Pr(G|s)$ the prosecutor's posterior having received the signal s. The prosecutor's decision to act, $a = 1$, is again given by:

$$\lambda \geq \psi \left[q - \Pr(G|s) \right].$$

However, there are now three values for $\Pr(G|s)$: 1 if $s = g$, 0 if $s = i$, and p if $s = \varnothing$. We can distinguish three corresponding strategies for the prosecutor:

(a) Always act: $a(s) = 1$ for all s.

(b) Act unless there is proof of innocence: $a(s) = 1$ if $s = g$ or if $s = \varnothing$.

(c) Act only if there is proof of guilt: $a(s) = 1$ if $s = g$ and $a = 0$ otherwise.

Which strategies are viable depend on the prosecutor's prior belief in the target's guilt p and his concern for convicting the innocent q. If $p \geq q$, he acts against the opponent even if he does not uncover additional information. By contrast, if $p < q$, he does not act unless he learns that the target is certainly guilty. We analyze each case in turn.

D.1 Act Unless There Is Evidence of Innocence

Suppose that $p \geq q$. Strategy (a), always to act, is optimal if the incumbent's offer λ is so high that it swamps the prosecutor's accuracy concerns. If the prosecutor chooses this strategy, the citizen learns nothing from the prosecutor and her posterior belief in the target's guilt is the same as her prior, p. Now consider strategy (b), act unless there is proof of innocence. This can be optimal if the incumbent's offer is sufficiently low. Denote by $\Pr_b(G|a)$ the citizen's posterior belief given this strategy. For an arbitrary effort level (in equilibrium, the citizen's beliefs about the prosecutor's effort will be correct), this is given by $\Pr_b(G|1) = \frac{p}{1-e(1-p)} > p$ if the prosecutor acts and by $\Pr_b(G|0) = 0$ if the prosecutor does *not* act. $\Pr_b(G|1) = \frac{p}{1-e(1-p)}$ is increasing in prosecutorial effort e.

How much effort the prosecutor exerts depends on the extent to which he expects new information to affect his behavior, given his strategy. In general, his maximization problem is:

$$\max_{e\in[0,1]} eV(s \neq \varnothing) + (1-e)V(s = \varnothing) - C(e)$$

where $V(s)$ is the prosecutor's utility from signal s. Differentiating and re-arranging yields that an interior solution is given by:

$$e^* = H\left(V(s \neq \varnothing) - V(s = \varnothing)\right) \tag{15}$$

where H is the inverse of $C'(e)$. Consequently, we need to find $V(s \neq \varnothing) - V(s = \varnothing)$. Suppose first that the prosecutor chooses to act regardless of the signal. In this case:

$$V(s \neq \varnothing) = p\left[-(1-\psi)(1-q) + \lambda\right] + (1-p)\left[-\psi q + \lambda\right]$$
$$V(s = \varnothing) = p\left[-(1-\psi)(1-q) + \lambda\right] + (1-p)\left[-\psi q + \lambda\right]$$

This means that $V(s \neq \varnothing) - V(s = \varnothing) = 0$, implying that effort is zero.

Now consider the case when the prosecutor acts unless he receives the innocent signal:

$$V(s \neq \varnothing) = p\left[-(1-\psi)(1-q)+\lambda\right] + (1-p)\left[0\right]$$
$$V(s = \varnothing) = p\left[-(1-\psi)(1-q)+\lambda\right] + (1-p)\left[-\psi q+\lambda\right]$$

This means that $V(s \neq \varnothing) - V(s = \varnothing) = (1-p)(\psi q - \lambda)$, implying

$$e_b^* = H\left((1-p)(\psi q - \lambda)\right)$$

which is decreasing in λ: interference essentially encourages the prosecutor to remain ignorant so that he can act in good conscience.

Turning to the incumbent's choice of interference, there are two relevant thresholds of citizen bias: p and $\Pr(G|1, \lambda = 0) = \frac{p}{1-e_b^*(0)(1-p)}$. Suppose first that $\mu < p$, i.e., the citizen supports the incumbent absent new information. Choosing $\lambda^F = \psi q$ induces no effort and prosecutor strategy (a), yielding the following expected utility:

$$\alpha\psi + B - K\left(\lambda^F\right)$$

Now consider a choice of $\lambda < \lambda^F$. In this case, the prosecutor exerts some effort and employs strategy (b). The incumbent's expected utility is:

$$\left[\Pr(g) + \Pr(\varnothing)\right](\alpha\psi + B) - K(\lambda) = \left[pe_b^* + 1 - e_b^*\right](\alpha\psi + B) - K(\lambda)$$

Here, the probability of the prosecutor acting, $pe_b^* + 1 - e_b^* = 1 - e_b^*(1-p)$ is increasing in λ because effort is decreasing in λ:

$$\frac{\partial e_b^*}{\partial \lambda} = -(1-p)H'\left((1-p)(\psi q - \lambda)\right) < 0$$

Consequently, the (locally) optimal level of interference is determined by the first-order condition:

$$-(1-p)\frac{\partial e_b^*}{\partial \lambda}(\alpha\psi + B) - K'(\lambda) = 0$$

Denote by λ^L the solution of this maximization problem. (For the remainder of this section, all locally optimal intervention levels will be denoted by λ^L—however, they may refer to different optimization problems.) Note that $\lambda^L > 0$ because $\frac{\partial e_b^*}{\partial \lambda}\big|_{\lambda=0} < 0$ and $K'(0) = 0$, which also rules out $\lambda = 0$ as an optimal choice.

The globally optimal choice is thus, either λ^L or λ^F, depending on the following inequality. The incumbent chooses full intervention if:

$$\alpha\psi + B - K\left(\lambda^F\right) \geq \left[pe_b^*(\lambda^L) + 1 - e_b^*(\lambda^L)\right](\alpha\psi + B) - K\left(\lambda^L\right)$$

or

$$\underbrace{B(1-p)e_b^*(\lambda^L)}_{\text{Informational Benefit}} \geq \overbrace{K\left(\lambda^F\right) - K\left(\lambda^L\right)}^{\text{Difference Costs}} - \alpha\psi \cdot \underbrace{(1-p)e_b^*(\lambda^L)}_{\text{Change Consequence}}$$

As in the baseline case, here the incumbent weighs the costs of interference net of the increase in the probability of inflicting the consequence against interference's informational implications. For the values of μ examined here, there is an informational benefit to interference, because it suppresses potentially damaging information. Notice that unlike the baseline case, here there is always *some* intervention, as the incumbent likes to decrease the prosecutor's effort.

Now suppose that $\mu \geq \Pr(G|1, \lambda = 0)$, i.e., the citizen never supports the incumbent even when the prosecutor acts under noninterference (i.e. even when the prosecutor's action is as informative as possible of guilt). As before, the incumbent can choose to fully intervene, $\lambda = \lambda^F$, or choose a lower level of intervention, yielding expected utility:

$$\alpha\psi\left[\Pr(g) + \Pr(\varnothing)\right] - K(\lambda)$$

Again, prosecutor action is increasing in interference, so a locally optimal level of intervention, λ^L, is given by the solution of the following first-order condition:

$$-(1-p)\alpha\psi\frac{\partial e_b^*}{\partial \lambda} - K(\lambda)$$

However, this level of intervention is not necessarily globally optimal. The incumbent may instead choose full intervention if:

$$\alpha\psi - K\left(\lambda^F\right) \geq \alpha\psi\left[e_b^*(\lambda^L) + 1 - e_b^*(\lambda^L)\right] - K(\lambda^L)$$

or

$$0 \geq K(\lambda^F) - K\left(\lambda^L\right) - \alpha\psi(1-p)e_b^*(\lambda^L)$$

Similar to the baseline analysis, intervention has no informational consequences for this region of citizen bias. Therefore, intervention depends only on whether its costs exceed its effect on the probability of inflicting the consequence. However, in contrast to the baseline case, there is again always some interference to depress effort.

Finally, suppose that $\mu \in (p, \Pr(G|1, \lambda = 0))$. Intervention now has three effects. As before, greater levels of intervention continuously increase the likelihood of prosecutorial action and the costs of intervention. Now, however, intervention also continuously decreases the likelihood of citizen support. This is because the citizen's posterior belief in the target's guilt after observing prosecutorial action, $\Pr_b(G|1)$, is increasing in effort, which means it is *decreasing* in interference. As a result, if the citizen's preferences μ are such that action by a sufficiently independent prosecutor would persuade her to drop the political opponent, at some point, increasing interference is counterproductive: coopting the prosecutor too much decreases the citizen's posterior beliefs upon observing $a = 1$ below μ, causing a discontinuous drop in the incumbent's utility. Then the level of intervention that solves $\Pr_b(G|1) = \mu$, i.e., that *just* persuades the citizen that the target is guilty when $a = 1$ is:[22]

$$\psi q - \frac{1}{1-p} \cdot C'\left(\frac{\mu - p}{\mu(1-p)}\right) \equiv \lambda_1^{BP}$$

In order to take all three effects into account, compute a locally optimal solution first,

[22]This pattern is similar to work on Bayesian Persuasion (see e.g., Kamenica and Gentzkow, 2011).

assuming it satisfies $\lambda_L < \lambda_1^{BP}$. The first-order condition is:

$$-(1-p)\frac{\partial e_b^*}{\partial \lambda}(\alpha\psi + B) - K'(\lambda) = 0$$

If $\lambda^L \geq \lambda_1^{BP}$, then the locally optimal choice is λ_1^{BP}. For simplicity, suppose that K is steep enough that λ^L is locally optimal. To determine global optimality, compare the incumbent's expected utility from λ^L to her expected utility from full intervention:

$$\alpha\psi - K\left(\lambda^F\right) \geq \left[e_b^*(\lambda^L)p + 1 - e_b^*(\lambda^L)\right](\alpha\psi + B) - K(\lambda^L)$$

or

$$-B\left[e_b^*(\lambda^L)p + 1 - e_b^*(\lambda^L)\right] \geq K\left(\lambda^F\right) - K\left(\lambda^L\right) - \alpha\psi(1-p)e_b^*(\lambda^L)$$

Similar to the baseline case, there is now an informational cost to intervening: if the incumbent fully interferes, the prosecutor has no incentive to exert effort, meaning he will never observe the guilty signal and persuade the citizen to drop the opponent. Here, as before, full intervention is least likely, but the incumbent still chooses some interference in order to decrease the likelihood that the prosecutor will exonerate the opponent.

Summarizing, when $p \geq q$, the broad patterns of the equilibrium analysis in the baseline case are similar. However, there are two important differences. First, there is always some level of interference in order to keep the prosecutor from exerting "too much" effort. Second, the incumbent's persuasion strategy is more sophisticated because the citizen's posterior depends on effort. Specifically, the incumbent takes into account that interference reduces the credibility of the prosecutor's action and makes sure that it does not cross a critical threshold (λ_1^{BP}).

D.2 Act Only If There Is Evidence of Guilt

Suppose that $p < q$, so that the prosecutor acts only if he receives the guilty signal. Then he plays either strategy (a) or (c). For the former case, effort is 0 in equilibrium by the analysis above. For the latter, optimal effort is determined by:

$$V(s \neq \varnothing) = p\left[-(1-\psi)(1-q)+\lambda\right]+(1-p)\left[0\right]$$
$$V(s = \varnothing) = p\left[-(1-q)+\lambda\right]+(1-p)\left[0\right]$$

Here, $V(s \neq \varnothing) - V(s = \varnothing) = p(\psi(1-q)+\lambda)$, implying $e_c^* = H\left(p(\psi(1-q)+\lambda)\right)$, which is increasing in λ.

For the citizen, two posteriors exist. If the prosecutor acts, she is now sure of guilt; however, if the prosecutor does not act, she updates negatively on the opponent's guilt, i.e., $\Pr(G|0,\lambda) = \frac{p(1-e(\lambda))}{1-pe(\lambda)} < p$. Moreover, the citizen's belief in guilt after prosecutorial inaction is decreasing in prosecutorial effort. Additionally, because prosecutorial effort is now increasing in interference, the citizen's posterior belief in guilt after observing inaction is now decreasing in interference. Thus, $\Pr(G|0,\lambda \to \lambda^F)$ is the lowest possible belief that the citizen may hold.

Suppose first that $\mu \leq \Pr(G|0,\lambda \to \lambda^F)$, i.e., the citizen supports the incumbent in all possible cases. Full intervention yields:

$$\alpha\psi + B - K(\lambda^F)$$

A lower intervention level yields:

$$\left[\Pr(g)\right]\alpha\psi + B - K(\lambda) = \left[e_c^*(\lambda)p\right]\alpha\psi + B - K(\lambda)$$

The locally optimal solution is given by:

$$p\frac{\partial e_c^*}{\partial\lambda}\alpha\psi - K'(\lambda) = 0$$

Call this solution again λ^L. The incumbent compares its expected utility with the expected utility of full intervention and chooses full intervention if:

$$0 \geq K\left(\lambda^F\right) - K\left(\lambda^L\right) - \alpha\psi\left[1 - e_c^*(\lambda^L)p\right]$$

As in the baseline, there are no informational consequences of interference for this range of μ, so the incumbent simply weighs the cost against the increased probability of inflicting the consequence. In contrast to the baseline (and the case $p \geq q$), however, the incumbent now always chooses some level of interference in order to *motivate* effort.

Now consider $\mu \geq p$. The expected utility of full intervention is $\alpha\psi - K(\lambda^F)$. By contrast, the expected utility of a lower level of intervention is:

$$\Pr(g)(\alpha\psi + B) - K(\lambda)$$

A locally optimal solution, λ^F, is:

$$p\frac{\partial e_c^*}{\partial \lambda}(\alpha\psi + B) - K'(\lambda) = 0$$

The incumbent chooses full intervention if:

$$-Bpe_c^*(\lambda^L) \geq K\left(\lambda^F\right) - K\left(\lambda^L\right) - \alpha\psi\left[1 - pe_c^*(\lambda^L)\right]$$

As before, since intervention precludes the release of potentially helpful information, there is a cost to interfering fully.

Finally, consider $\mu \in \left(\Pr(G|0, \lambda^F), p\right)$. Define λ_0^{BP} to be the solution to the following equation:

$$\Pr(G|0, \lambda_0^{BP}) = \mu$$

Solving yields the explicit solution:

$$\lambda_0^{BP} = p^{-1}C'\left(\frac{p - \mu}{p(1 - \mu)}\right) - \psi(1 - q)$$

This is the level of intervention that is sufficiently low to still persuade the citizen that the target might be guilty, even when $a = 0$.

Consider the following first-order condition:

$$p\frac{\partial e_c^*}{\partial \lambda}\alpha\psi - K'(\lambda) = 0$$

If the solution, λ^L, is smaller than λ_0^{BP}, then the incumbent compares the expected utility of λ^L to the expected utility of full interference, λ^F. Suppose K is sufficiently steep to make this inequality hold. Then, λ^F is optimal if:

$$\alpha\psi + B - K(\lambda^F) \geq \left[pe_c^*(\lambda^L)\right]\alpha\psi + B - K(\lambda^L)$$

or

$$0 \geq K\left(\lambda^F\right) - K\left(\lambda^L\right) - \alpha\psi\left[1 - pe_c^*(\lambda^L)\right]$$

Note that the incumbent receives B in *both* cases, but for different reasons: with full intervention (λ^F), no information is released, but given that the citizen currently favors the incumbent, the incumbent obtains the citizen's support. With partial interference (λ^L), there is some information release but the prosecutor's effort is so low that even if there is inaction, the citizen still believes it is possible that the target is guilty.

E Extension: Early vs. Late Interventions

Suppose that there are two periods ("early," E, and "late," L, in the incumbent's term). The court is informative with parameters ψ_θ, but can only produce a decision before the election when the investigation is initiated early—otherwise, a decision is reached after the citizen makes the support decision. The timing is as follows:

1. Incumbent chooses early level of intervention, $\lambda_E \geq 0$.
2. Prosecutor receives signal $s_E \in \{g, i\}$.
3. Prosecutor decides whether to act or wait, $a_E \in \{0, 1\}$.

4. If prosecutor acts:

   (a) Court produces a consequences with probability:

$$\Pr(C = 1|a_E = 1, \theta) = \psi_\theta$$

   (b) Citizen chooses to support incumbent or not.

5. If prosecutor waits ($a_E = 0$):

   (a) Incumbent chooses late level of intervention, $\lambda_L \geq 0$.

   (b) Prosecutor receives signal $s_L \in \{g, i\}$.

   (c) Prosecutor decides whether to act, $a_L \in \{0, 1\}$.

   (d) Citizen chooses to support incumbent or not.

   (e) Court produces a consequences with probability:

$$\Pr(C = 1|a, \theta) = \psi_\theta \cdot a_L$$

We assume that in each period, the probability of receiving a guilty signal conditional on the state is given by $\Pr(g|\theta) = \gamma_\theta$, with $\gamma_G < 1$ and $\gamma_I > 0$. We denote a posterior belief by $\Pr(G|s_E)$ and $\Pr(G|s_E, s_L)$. Similar to the baseline analysis, we assume that parameter values are such that given no intervention, the prosecutor is inclined to act when receiving a single guilty signal:

$$\Pr(G|g) > \frac{\psi_I q}{\psi_I q + (1-q)\psi_G} > \Pr(G|i) \quad \text{and} \quad \Pr(G|i, g) > \frac{\psi_I q}{\psi_I q + (1-q)\psi_G} \tag{16}$$

The prosecutor's payoffs are still given by $u_{C\theta}(q) + \lambda$. The interpretation of $a_E = 0$ is that the prosecutor *waits*: he temporarily does not act, but does not forfeit his ability to act in future (as in the single period model). We abstract away from any costs of waiting, e.g., we assume that there is no discounting, no cost to letting a potentially guilty person roam free, and no other cost or benefit to pursuing this case (as opposed to others) at a particular time.Finally, for simplicity, we assume that there are identical costs to the incumbent of intervening early or late.

As explained in the main text, we focus on the case where the incumbent intervenes for sure in the second period. By Expression 16, to ensure prosecutorial action in that period, she must target the type of prosecutor who received two innocent signals and offer the following:

$$\lambda_L^F \equiv \psi_I q - \Pr(G|i,i)\left[\psi_I q + (1-q)\psi_G\right]$$

The incumbent's expected utility of choosing this level of intervention is:

$$\alpha(\tilde{p}\psi_G + (1-\tilde{p})\psi_I) + \mathbb{1}(\tilde{p} \geq \mu)B - K\left(\lambda_L^F\right)$$

where $\tilde{p}$ is the (common) belief that the target is guilty at the beginning of the second period. In principle, this belief might differ from the prior if different types of prosecutors choose different actions in the first period, allowing both the incumbent and the citizen to learn about the target's likely guilt. However, if the incumbent chooses to fully interfere in period 2 and the prosecutor anticipates this behavior, both types of prosecutor choose to wait. To see this, note that the expected utility to type s_E of choosing $a_E = 1$ is:

$$\Pr(G|s_e)\left[-(1-\psi_G)(1-q)\right] + (1-\Pr(G|s_E))\left[-\psi_I q\right] + \lambda_E$$

or

$$-\psi_I q + \Pr(G|s_e)\left[\psi_I q - (1-\psi_G)(1-q)\right] + \lambda_E$$

By contrast, the expected utility to type s_E of choosing $a_E = 0$ is:

$$\mathbb{E}_{s_L}\left[\Pr(G|s_e,s_L)\left[-(1-\psi_G)(1-q)\right] + (1-\Pr(G|s_E,s_L))\left[-\psi_I q\right]\right] + \lambda_L^F$$

or

$$-\psi_I q + \mathbb{E}_{s_L}\left[\Pr(G|s_e,s_L)\left[\psi_I q - (1-\psi_G)(1-q)\right]\right] + \lambda_L^F$$

or, because averaging over the posterior yields the relevant prior:

$$-\psi_I q + \Pr(G|s_e)\left[\psi_I q - (1-\psi_G)(1-q)\right] + \lambda_L^F$$

Thus, both prosecutor types chooses to act today if:

$$\lambda_E \geq \lambda_L^F = \psi_I q - \Pr(G|i,i)\left[\psi_I q + (1-q)\psi_G\right]$$

Both types employ the same decision rule, rendering seperation impossible: if offered nothing today, they both choose $a_E = 0$; if offered at least as much today as offered tomorrow, they both choose $a_E = 1$.

As a consequence, when intervention is expected in the second period, the incumbent's equilibrium utility for the late period is:

$$V_{\text{Inc}}^F = \alpha\left(p\psi_G + (1-p)\psi_I\right) + \mathbb{1}(p \geq \mu)B - K\left(\lambda_L^F\right)$$

We now investigate the incumbent's decision in the first period. The expected utility of choosing $\lambda_E = \lambda_L^F$ and hence fully intervening today is:

$$\alpha\left[p\psi_G + (1-p)\psi_I\right] + \Pr(C=1|\lambda_E^F)\mathbb{1}\left(p^F(1) \geq \mu\right)B + \Pr(C=0|\lambda_E^F)\mathbb{1}\left(p^F(1) \geq \mu\right)B - K\left(\lambda_E^F\right)$$

where:

$$p^F(1) = \frac{p\psi_G}{p\psi_G + (1-p)\psi_I} \quad \text{and} \quad p^F(0) = \frac{p(1-\psi_G)}{p(1-\psi_G) + (1-p)(1-\psi_I)}$$

The expected utility of not intervening today is:

$$\alpha\left[p\psi_G + (1-p)\psi_I\right] + \mathbb{1}\left(p \geq \mu\right)B - K\left(\lambda_L^F\right)$$

The incumbent hence intervenes today if:

$$T_E \geq K(\lambda_E^F) - K(\lambda_L^F) = 0$$

where

$$T_E = \begin{cases} 0 & \text{if } \mu \leq p^F(0) \\ -B[1 - p\psi_G - (1-p)\psi_I] & \text{if } \mu \in \big(p^F(0), p\big] \\ B[p\psi_G + (1-p)\psi_I] & \text{if } \mu \in \big(p, p^F(1)\big] \\ 0 & \text{if } \mu > p^F(1) \end{cases}$$

The above analysis assumes that the incumbent intervenes in the second period. Here, we discuss the consequences of relaxing this assumption. To begin with, note that in a pure strategy equilibrium, the incumbent's (and citizen's) posterior belief at the beginning of the second period is either p (if both types choose $a_e = 0$) or $\Pr(G|i)$ (if type g chooses $a_E = 1$ but type i chooses $a_E = 0$). Consider the pooling situation in which both types of prosecutors choose $a_E = 0$, not permitting any learning. Then the posterior belief is p and the expected utility of not intervening is:

$$\alpha(p\gamma_G\psi_G + (1-p)\gamma_I\psi_I) + [p\gamma_G + (1-p)\gamma_I]\mathbb{1}(\Pr(G|g) \geq \mu)B +$$
$$[p(1-\gamma_G) + (1-p)(1-\gamma_I)]\mathbb{1}(\Pr(G|i) \geq \mu)B$$

Thus, the incumbent intervenes if

$$T_L \geq K(\lambda_L^F) - \alpha\Delta^\alpha$$

where

$$T_L = \begin{cases} 0 & \text{if } \mu \leq p^N(0) \\ B[1 - p\gamma_G - (1-p)\gamma_I] & \text{if } \mu \in \big(p^N(0), p\big] \\ -B[p\gamma_G + (1-p)\gamma_I] & \text{if } \mu \in \big(p, p^N(1)\big] \\ 0 & \text{if } \mu > p^N(1) \end{cases}$$

Thus, assuming that $\alpha\Delta^\alpha - K(\lambda_L^F) \geq B[p\gamma_G + (1-p)\gamma_I]$ renders choosing λ_L^F optimal for all popularity levels μ.

There is no equilibrium in which the prosecutor follows his signal in period 1, allowing learning, and the incumbent still intervenes in the second period if it is reached, i.e., if the prosecutor observes the innocent signal in the first period and first-period interference is not full ($\lambda_E < \lambda_L^F$). To see this, suppose there is. Then $\tilde{p} = \Pr(G|i)$. In the second period, the incumbent intervenes with a similar calculus as above, with $\Pr(G|i)$ replacing p. Suppose that parameter values are such that it is optimal to intervene at this point. But then the type of prosecutor who observed the guilty signal and acted in period 1 wishes to deviate to inaction whenever $\lambda_E < \lambda_L^F$. Thus, the separating equilibrium is infeasible.