Online Appendix

Political Interventions in the Administration of Justice

Carlo M. Horz* Hannah K. Simpson†

*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.

Contents

A Omitted Proofs **3**

B Robustness **11**

B.1 Partially Observed Interference 11

B.2 Citizen Punishment 14

B.3 Incumbent Protects Ally 15

C Additional Results **18**

C.1 Citizen Welfare 18

C.2 Mixed Strategy Equilibria 20

D Extension: Endogenous Effort **25**

D.1 Act Unless There Is Evidence of Innocence 26

D.2 Act Only If There Is Evidence of Guilt 30

E Extension: Early vs. Late Interventions **33**

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:

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

Moreover:

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

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:

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

□

Proof of Proposition 3:

Proof. Recall the threshold derived in the main text:

$$T_{MI} = B\left[1 - p\gamma_G - (1-p)\gamma_I\right] > 0$$

$$T_{MO} = B\left[-p\gamma_G - (1-p)\gamma_I\right] < 0$$

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

$$T_{IO} = B\left[-p\psi_G\gamma_G - (1-p)\psi_I\gamma_I\right] < 0$$

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 1 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 1: 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 in the main text.

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:

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

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:

$$Pr\left(\lambda \text{ observed}|\lambda\right) = \begin{cases} 0 & \text{if } \lambda = 0 \\ \varphi & \text{if } \lambda > 0 \end{cases} \tag{1}$$

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:

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

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:

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

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

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

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

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:

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

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

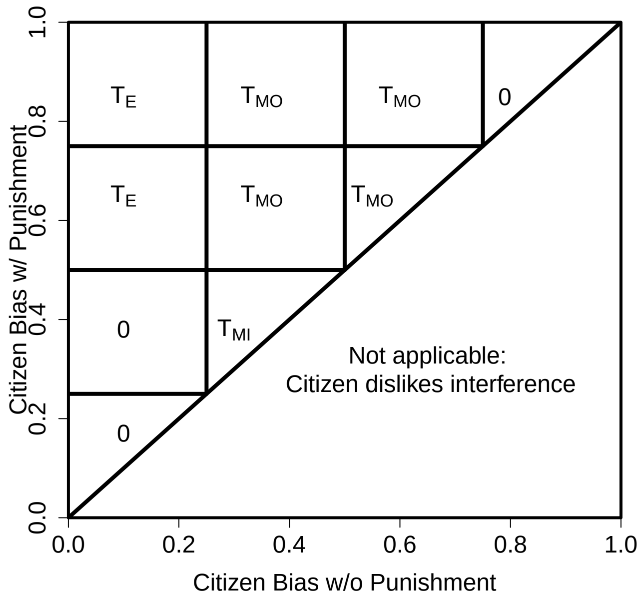

Figure 2: Overview of equilibrium thresholds with citizen punishment

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 2 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:

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

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:

$$U_V = u_{\theta C}(z) - M(\lambda) + rv_{\text{Inc}} + (1-r)[\mathbb{1}(\theta = G)v_O(G) + \mathbb{1}(\theta = I)v_O(I)] \qquad (2)$$

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

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

whereas that from supporting the opposition is

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

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 2 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:

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

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 3, 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

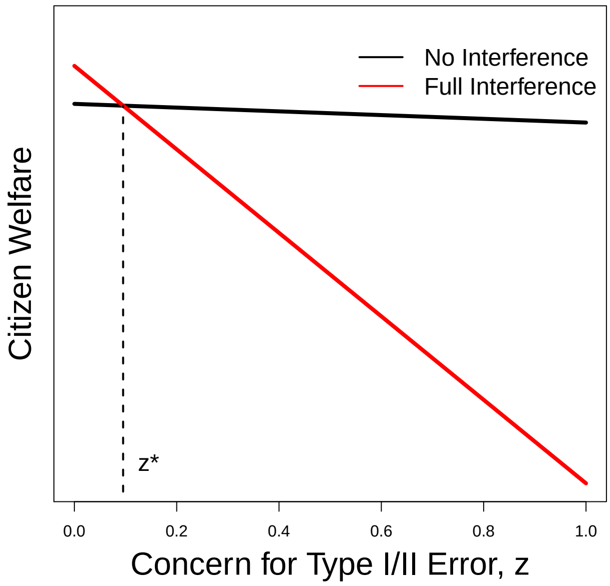

Figure 3: The citizen's welfare. Parameter values: $p = 0.3$, $\gamma_G = 0.76$, $\gamma_I = 0.25$, $\psi = 0.8$, $v_O(G) = -0.1$, $v_O(I) = 1.5$, $v_{\text{Inc}} = 0.6$, and $M(\lambda^F) = 0.01$.

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:

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

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.,

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

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 $\lambda^F + t$:

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

Re-arranging this condition yields:

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

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:

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

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:

$$e^* = H\left(V(s \neq \varnothing) - V(s = \varnothing)\right) \tag{3}$$

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:[1]

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

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

[1]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:

$$\Pr(G|g) > \frac{\psi_I q}{\psi_I q + (1-q)\psi_G} > \Pr(G|i) \quad \text{and} \quad \Pr(G|i, g) > \frac{\psi_I q}{\psi_I q + (1-q)\psi_G} \tag{4}$$

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 4, 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.

References

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