# Appendix

*Proof of Proposition 1.* Let $\mu(x) = P(\theta = 1|x)$ be the voter's posterior belief that the incumbent is aligned given the realized policy outcome $x_1 = x$. As effort is unobserved, let the voter have conjecture about the incumbent's effort choice, $\hat{a}_1(\theta)$. Formally, posterior beliefs can be expressed as

$$\mu(x) = \frac{\gamma\phi(\sqrt{\zeta}(x - \hat{a}_1(1)))}{\gamma\phi(\sqrt{\zeta}(x - \hat{a}_1(1))) + (1-\gamma)\phi(\sqrt{\zeta}(x - \hat{a}_1(0)))}.$$

The voter retains the incumbent iff $\mu(x) \geq \gamma$, which is equivalent to

$$x \geq \frac{\hat{a}_1 + \hat{a}_0}{2}.$$

Given $x = a + \varepsilon$, the incumbent leader survives iff $a + \varepsilon \geq \frac{\hat{a}_1 + \hat{a}_0}{2}$. Since $\varepsilon \sim N(0, \frac{1}{\zeta})$, the incumbent's reelection probability is equal to

$$\pi(a) = \Phi(\sqrt{\zeta}(a - \frac{\hat{a}_1 + \hat{a}_0}{2})).$$

The leader of type $\theta$ maximizes

$$\max_a \ \pi(a\Psi - \frac{\lambda_\theta}{2}a^2.$$

This leads to the first-order condition

$$-\lambda_\theta a + \sqrt{\zeta}\phi(\sqrt{\zeta}(a - \frac{\hat{a}_1 + \hat{a}_0}{2}))\Psi = 0.$$

Since beliefs are correct in equilibrium, $a_\theta = \hat{a}_\theta = a_\theta^*$, this simplifies to

$$-\lambda_\theta a_\theta^* + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^* + a_0^*}{2}))\Psi = 0.$$

Substituting in $\theta = 1$ and $\theta = 0$ yields the two equations in the proposition.

To show that this solution is a maximum, we ensure that the leader's utility is concave. The second-order condition is

$$-\lambda_\theta + \zeta^{3/2}(a - \frac{\hat{a}_1 + \hat{a}_0}{2})\phi(\sqrt{\zeta}(a - \frac{\hat{a}_1 + \hat{a}_0}{2}))\Psi.$$

Let $\eta = \sqrt{\zeta}(a - \frac{\hat{a}_1+\hat{a}_0}{2})$ so the second-order condition can be rewritten as

$$-\lambda_\theta + \zeta\eta\phi(\eta)\Psi.$$

The standard normal density tends to zero faster than any polynomial so $\eta\phi(\eta)$ is zero at $\eta = 0$ and approaches zero as $\eta \to \pm\infty$. The derivative of $\eta\phi(\eta)$ is $\phi(\eta) - \eta^2\phi(\eta)$ with critical points at $\eta = \pm 1$. Note that if $\eta = -1$ then the problem is globally concave. Hence the relevant constraint is at $\eta = 1$, where $\eta\phi(\eta) = \frac{1}{\sqrt{2\pi e}}$. Hence the leader's utility is concave iff

$$-\lambda_\theta + \frac{\zeta}{\sqrt{2\pi e}}\Psi < 0,$$

or $\zeta < \frac{\lambda_\theta\sqrt{2\pi e}}{\Psi}$. Hence a sufficient condition for both leaders to have concave utility functions is $\zeta < \frac{\lambda_1\sqrt{2\pi e}}{\Psi}$.

Furthermore, this equilibrium is unique because pooling cannot be an equilibrium. Since $\lambda_0 > \lambda_1$, efforts are always distinct unless both leaders were to pool on $a_\theta = 0$. But clearly the aligned leader has incentive to deviate from this strategy as it would increase the odds of reelection. $\square$

*Proof of Proposition 2.* Proof is analogous to that of Proposition 1 for the derivation of the optimal effort. The only difference is the derivation of the voter's policy cutoff, which is a function of conjectures about the leader's effort $\hat{a}_\theta$ as well as conjectures about the messages sent to the IO $\hat{p}_\theta$.

Denote $\mu(x, s)$ as the voter's posterior belief about the leader's type having observed IO report $s$ and signal $x$ of the leader's effort. Since the leader's true message $m$ and true effort $a$ are unobserved, the voter needs to have conjectures. Let $\hat{a}_\theta$ be the voter's conjecture about leader-type $\theta$'s effort, and let $\hat{p}_\theta = P(m = 1|\theta)$ be the voter's conjecture about the probability that leader-type $\theta$ sent message $m = 1$ to the IO. Then $\hat{m}_\theta = \hat{p}_\theta\phi(\sqrt{\tau}(s-1))+(1-\hat{p}_\theta)\phi(\sqrt{\tau}s)$ is the total probability that that IO's report is realized as the value $s$ given voter's conjectures. Then $\mu(x, s)$ can be expressed as

$$\mu(x, s) = \frac{\gamma\phi(\sqrt{\zeta}(x - \hat{a}_1))\hat{m}_1}{\gamma\phi(\sqrt{\zeta}(x - \hat{a}_1))\hat{m}_1 + (1 - \gamma)\phi(\sqrt{\zeta}(x - \hat{a}_0))\hat{m}_0},$$

such that it is optimal to retain the incumbent leader whenever

$$x \geq \frac{\hat{a}_1 + \hat{a}_0}{2} + \frac{\log(\frac{\hat{m}_0}{\hat{m}_1})}{\zeta(\hat{a}_1 - \hat{a}_0)} \equiv \hat{x}(\hat{a}, \hat{p}).$$

It is immediate that whenever $\hat{p}_1 = \hat{p}_0$ then $\hat{x}(\hat{a}, \hat{p}) = \frac{\hat{a}_1+\hat{a}_0}{2}$, as in the model without the IO. Optimal effort is thus identical to that characterized in Proposition 1.

Suppose $\hat{m}_1 \neq \hat{m}_0$. The first-order condition for leader-type $\theta$'s effort is

$$-\lambda_\theta a + \sqrt{\zeta}\phi(\sqrt{\zeta}(a - \frac{\hat{a}_1 + \hat{a}_0}{2} - \frac{\log(\frac{\hat{m}_0}{\hat{m}_1})}{\zeta(\hat{a}_1 - \hat{a}_0)}))\Psi = 0.$$

Since the normal density is log-concave, it is single peaked. Hence $\phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p})))$ is single peaked in $s$ such that there is a $s^{max}$ where $\frac{d}{ds}\phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p}))) > 0$ for $s < s^{max}$ and $\frac{d}{ds}\phi(\sqrt{\zeta}(a - \hat{x}(\hat{a}, \hat{p}))) < 0$ for $s > s^{max}$. As such optimal effort is single peaked in

$s$, $\frac{da_\theta^*}{ds}$ is nonmonotonic in $s$. Moreover, observe that $\lim_{s\to-\infty}\phi(\sqrt{\zeta}(a-\hat{x}(\hat{a},\hat{p})))=0$ and $\lim_{s\to\infty}\phi(\sqrt{\zeta}(a-\hat{x}(\hat{a},\hat{p})))=0$ such that as $s\to\pm\infty$, $a_\theta^*\to 0$.

Denote leader-type $\theta$'s optimal effort in the model without the IO as $\tilde{a}_\theta$. Therefore since $a_\theta^*$ is continuous in $s$ and $\tilde{a}_\theta>0$ there exists $\underline{s}_\theta$ such that $a_\theta^*=\tilde{a}_\theta$ when $\frac{da_\theta^*}{ds}>0$ and $\overline{s}_\theta$ such that $a_\theta^*=\tilde{a}_\theta$ when $\frac{da_\theta^*}{ds}<0$. $\square$

*Proof of Proposition 3.* The leader maximizes

$$\max_{m\in\{0,1\}}\int_{-\infty}^{\infty}\left[\pi(a_\theta^*(s))\Psi-\frac{\lambda_\theta}{2}a_\theta^{*^2}\right]\phi(\sqrt{\tau}(s-m))ds,$$

therefore choosing $m=1$ over $m=0$ whenever

$$\int_{-\infty}^{\infty}\left[\pi(a_\theta^*(s))\Psi-\frac{\lambda_\theta}{2}a_\theta^{*^2}\right]\phi(\sqrt{\tau}(s-1))ds\geq\int_{-\infty}^{\infty}\left[\pi(a_\theta^*(s))\Psi-\frac{\lambda_\theta}{2}a_\theta^{*^2}\right]\phi(\sqrt{\tau}s)ds,$$

which simplifies to

$$\int_{-\infty}^{\infty}\pi(a_\theta^*(s))\Big(\phi(\sqrt{\tau}(s-1))-\phi(\sqrt{\tau}s)\Big)ds\geq 0.$$

Define $\Delta_\theta(\hat{p}_1,\hat{p}_0)=\int_{-\infty}^{\infty}\pi(a_\theta^*(s))\Big(\phi(\sqrt{\tau}(s-1))-\phi(\sqrt{\tau}s)\Big)ds$ as the leader's difference in expected reelection probability from sending message $m=1$ versus $m=0$ when she is of type $\theta$. If $\hat{p}_1=\hat{p}_0$, then $\hat{x}(a^*,\hat{p})=\frac{a_1^*+a_0^*}{2}$, and $\pi(a_\theta^*;s)$ is constant in $s$ so $\Delta_\theta(\hat{p}_1,\hat{p}_0)$ is the difference of two densities integrated over their entire support, thus $\Delta_\theta(\hat{p}_1,\hat{p}_0)=0$. If $\Delta_\theta(\hat{p}_1,\hat{p}_0)=0$, it must be because $\hat{p}_1=\hat{p}_0$. Observe that $\pi(a^*;s)=0$ only if $s\to\pm\infty$, so for any finite $s$ $\pi(a^*;s)>0$. Moreover we are integrating over the entire space of $s$ so it must be that $\pi(a^*;s)$ is constant in $s$ and $\int_{-\infty}^{\infty}\Big(\phi(\sqrt{\tau}(s-1))-\phi(\sqrt{\tau}s)\Big)ds=0$, which occurs when $\hat{p}_1=\hat{p}_0$. Hence $\Delta_\theta(\hat{p}_1,\hat{p}_0)=0$ iff $\hat{p}_1=\hat{p}_0$.

Now we show that $\hat{p}_1 = \hat{p}_0$ must occur at an interior $p^* \in (0,1)$. For the aligned type,

$$\frac{\partial \Delta_1(\hat{p}_1, \hat{p}_0)}{\partial \hat{p}_1} = \int_{-\infty}^{\infty} \sqrt{\zeta}\phi(\sqrt{\zeta}(a_1^* - \hat{x}(a^*, \hat{p}))) \frac{1}{\zeta(a_1^* - a_0^*)\hat{m}_1} \Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)^2 ds > 0,$$

so increasing the voter's belief that the aligned type sends $m = 1$ increases the return from playing $m = 1$ versus $m = 0$. For the misaligned type,

$$\frac{\partial \Delta_0(\hat{p}_1, \hat{p}_0)}{\partial \hat{p}_0} = \int_{-\infty}^{\infty} -\sqrt{\zeta}\phi(\sqrt{\zeta}(a_0^* - \hat{x}(a^*, \hat{p}))) \frac{1}{\zeta(a_1^* - a_0^*)\hat{m}_0} \Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)^2 ds < 0.$$

From this we know that $\Delta_1(\hat{p}_1, \hat{p}_0) < 0$ if $\hat{p}_1 < \hat{p}_0$ and $\Delta_1(\hat{p}_1, \hat{p}_0) > 0$ if $\hat{p}_1 > \hat{p}_0$. Furthermore, $\Delta_0(\hat{p}_1, \hat{p}_0) > 0$ if $\hat{p}_1 > \hat{p}_0$ and $\Delta_1(\hat{p}_1, \hat{p}_0) < 0$ if $\hat{p}_1 < \hat{p}_0$. To see that $\hat{p}_1 = \hat{p}_0 = 1$ or $\hat{p}_1 = \hat{p}_0 = 0$ cannot be an equilibrium, observe that $\Delta_1(\hat{p}_1, 1) < 0$ for any $\hat{p}_1$, meaning the aligned type would deviate to $m = 0$. Similarly, $\Delta_1(\hat{p}_1, 0) > 0$ for any $\hat{p}_0$, meaning the misaligned type would deviate to $m = 1$. Thus the only equilibrium is $p_1^* = p_0^* = p^* \in (0,1)$. □

The model is also robust to several modifications. Suppose that there is a cost $k > 0$ associated with reporting $m = 1$. The function $\Delta_\theta(\hat{p}_1, \hat{p}_0)$ is now $\Delta_\theta(\hat{p}_1, \hat{p}_0) = \int_{-\infty}^{\infty} \pi(a_\theta^*(s))\Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)ds - k$, such that the core equilibrium result is robust, although the equilibrium mixing probability $p^*$ will shift with $k$.

Additionally consider the effects of reputation. Suppose that the leader incurs a reputational cost $c > 0$ if the IO reports a GPI lower than some threshold $\hat{s}$, which is to say that the IO releases a critical score of the leader's commitment to reforms. The probability of incurring the reputational cost upon sending message $m$ is thus $\Phi(\sqrt{\tau}(\hat{s} - m))$, and the function $\Delta_\theta(\hat{p}_1, \hat{p}_0)$ is now $\Delta_\theta(\hat{p}_1, \hat{p}_0) = \int_{-\infty}^{\infty} \pi(a_\theta^*(s))\Big(\phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s)\Big)ds - c(\Phi(\sqrt{\tau}(\hat{s} - 1)) + \Phi(\sqrt{\tau}\hat{s}))$. Results are not qualitatively different.