# Appendix

*Proof of Proposition 1.* Proceed by backward induction. In the second period, a leader of type $\theta$ maximizes

$$\max_{a_2(\theta)} \beta\Big(a_2(\theta) - \theta\frac{\lambda_M}{2}a_2(\theta)^2 - (1-\theta)\frac{\lambda_M + \lambda_S}{2}a_2(\theta)^2\Big).$$

The first-order condition is

$$\beta\Big(1 - \theta\lambda_M a_2(\theta) - (1-\theta)(\lambda_M + \lambda_S)a_2(\theta) = 0,$$

which has solution $a_2^*(1) = \frac{1}{\lambda_M}$ and $a_2^*(0) = \frac{1}{\lambda_M+\lambda_S}$. These choices are a maximum because the leader's utility function is globally concave, as the second-order condition is

$$-\theta\lambda_M - (1-\theta)(\lambda_M + \lambda_S) < 0.$$

Since $a_2^*(1) > a_2^*(0)$ and in particular $a_2^*(1)$ is the median voter's ideal point, the median voter wants to retain the incumbent leader when his posterior belief about the leader's honesty is greater than the prior. Moreover, since $x_1$ is FOSD-increasing in $a_1$, higher signals are on average more likely to signal honesty. Therefore the voter prefers to retain the incumbent whenever the signal $x_1$ is greater than some threshold $\hat{x}$. Let $\mu(x) = P(\theta = 1|x_1 = x)$ be the voter's posterior belief that the incumbent is honest 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 $x_1 = a_1 + \varepsilon_1$, the incumbent leader survives iff $a_1 + \varepsilon_1 \geq \frac{\hat{a}_1 + \hat{a}_0}{2}$. Since $\varepsilon_1 \sim N(0, \frac{1}{\zeta})$, the incumbent's reelection probability is equal to

$$\pi(a_1) = \Phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1) + \hat{a}_1(0)}{2})).$$

In the first period, the leader of type $\theta$ maximizes

$$\max_{a_1(\theta)} \beta\Big(a_1(\theta) - \theta\frac{\lambda_M}{2}a_1(\theta)^2 - (1-\theta)\frac{\lambda_M + \lambda_S}{2}a_1(\theta)^2\Big) + \pi(a_1(\theta))\Psi,$$

which leads to the first-order condition

$$\beta - \theta\lambda_M a_1(\theta) - (1-\theta)(\lambda_M + \lambda_S)a_1(\theta) + \sqrt{\zeta}\phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1) + \hat{a}_1(0)}{2}))\Psi = 0.$$

Since beliefs are correct in equilibrium, $a_1(\theta) = \hat{a}_1(\theta) = a_1^*(\theta)$, this simplifies to

$$\beta - \theta\lambda_M a_1(\theta) - (1-\theta)(\lambda_M + \lambda_S)a_1(\theta) + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^*(1) + a_1^*(0)}{2}))\Psi = 0.$$

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

$$-\theta\lambda_M - (1-\theta)(\lambda_M + \lambda_S) + \zeta^{3/2}(a_1 - \frac{\hat{a}_1(1) + \hat{a}_1(0)}{2})\phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1) + \hat{a}_1(0)}{2}))\Psi.$$

Let $\eta = \sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2})$ so the second-order condition can be rewritten as

$$-\lambda_M - (1-\theta)\lambda_S + \zeta\eta\phi(\eta)\Psi.$$

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

$$-\lambda_M - (1-\theta)\lambda_S + \frac{\zeta}{\sqrt{2\pi e}}\Psi < 0,$$

or $\zeta < \frac{\lambda_M + (1-\theta)\lambda_S\sqrt{2\pi e}}{\Psi}$. Hence a sufficient condition for both leaders to have concave utility functions is $\zeta < \frac{\lambda_M\sqrt{2\pi e}}{\Psi}$.

Furthermore, this equilibrium is unique because pooling cannot be an equilibrium. By way of contradiction, suppose that the voter believed $\hat{a}_1(\theta) = \hat{a}$ for any $\theta$. Then $\mu(x) = \gamma$ for any $x$, and the voter is indifferent between retaining and replacing the incumbent. Hence, depending on how ties are broken, the incumbent's reelection probability is either zero or 1. This means that the incumbent leader's maximization problem is equivalent to

$$\max_{a_1(\theta)} \beta\Big(a_1(\theta) - \theta\frac{\lambda_M}{2}a_1(\theta)^2 - (1-\theta)\frac{\lambda_M+\lambda_S}{2}a_1(\theta)^2\Big),$$

the solution to which is $a_1(1) = \frac{1}{\lambda_M}$ and $a_1(0) = \frac{1}{\lambda_M+\lambda_S}$ such that $a_1(1) \neq a_1(0)$. □

*Proof of Corollary 1.* From Proposition 1, leader's effort choices satisfy

$$\beta + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^*(1) - a_1^*(0)}{2}))\Psi = \beta\lambda_M a_1^*(1).$$
$$\beta + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{a_1^*(0) - a_1^*(1)}{2}))\Psi = \beta(\lambda_M+\lambda_S) a_1^*(0).$$

The LHS of these equations are the same, which implies that $\beta\lambda_M a_1^*(1) = \beta(\lambda_M + \lambda_S)a_1^*(0)$, or $a_1^*(1) = \frac{\lambda_M+\lambda_S}{\lambda_M} a_1^*(0)$ so $a_1^*(1) > a_1^*(0)$.

To see that $a_1^*(1) > \frac{1}{\lambda_M}$, substitute $a_1 = \frac{1}{\lambda_M}$ into the honest leader's first-order condition to get

$$\sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{1/\lambda_M - a_1^*(0)}{2}))\Psi > 0,$$

which holds for any $a_1^*(0)$. Since this first-order condition is positive, we must have $a_1^*(1) > \frac{1}{\lambda_M}$. Similarly, substituting $a_1 = \frac{1}{\lambda_M}$ into the captured leader's first-order condition yields

$$\beta(1 - \frac{\lambda_M + \lambda_S}{\lambda_M}) + \sqrt{\zeta}\phi(\sqrt{\zeta}(\frac{1/\lambda_M - a_1^*(1)}{2}))\Psi.$$

This expression can be either positive or negative. Note that the standard normal density takes a maximum value of $\frac{1}{\sqrt{2\pi}}$, and so a sufficient condition for the captured leader's equilibrium effort to be larger than $\frac{1}{\lambda_M}$ is

$$\beta(1 - \frac{\lambda_M + \lambda_S}{\lambda_M}) + \sqrt{\frac{\zeta}{2\pi}}\Psi > 0,$$

which occurs whenever $\lambda_S < \lambda_M + \sqrt{\frac{\zeta}{2\pi}}\frac{\Psi}{\beta}$. □

*Proof of Corollary 2.* Follows from $x_1$ FOSD-increasing in $a_1$ and $a_1^*(1) > a_1^*(0)$. □

*Proof of Corollary 3.* Define the Jacobian for type $\theta$ as

$$\mathbf{J}_\theta = \begin{bmatrix} \frac{\partial^2 v_\theta}{\partial a_\theta^2} & \frac{\partial^2 v_\theta}{\partial a_\theta \partial a_{\theta'}} \\ \frac{\partial^2 v_{\theta'}}{\partial a_\theta \partial a_{\theta'}} & \frac{\partial^2 v_{\theta'}}{\partial a_{\theta'}^2} \end{bmatrix}.$$

Observe that $\frac{\partial^2 v_1}{\partial a_1^2} < 0$ and $\frac{\partial^2 v_0}{\partial a_0^2} < 0$ at the equilibrium $(a_1^*, a_0^*)$ because they are maxima. Further observe that $\frac{\partial^2 v_1}{\partial a_1 \partial a_0} > 0$ and $\frac{\partial^2 v_0}{\partial a_1 \partial a_0} < 0$, hence $|\mathbf{J}_\theta| > 0$. Given this structure, the direct and indirect effects have the same sign; without loss of generality I simply consider the direct effects.

By monotone comparative statics, taking the cross-partial of the leader's utility with respect to parameters yields

$$\begin{aligned}
\frac{\partial^2 v_L(a_t;\theta)}{\partial a_1 \partial \Psi} &= \sqrt{\zeta}\phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2})) \geq 0. \\
\frac{\partial^2 v_L(a_t;\theta)}{\partial a_1 \partial \lambda_M} &= -a_1 \leq 0. \\
\frac{\partial^2 v_L(a_t;\theta)}{\partial a_1 \partial \lambda_S} &= -(1-\theta)a_1 \leq 0. \\
\frac{\partial^2 v_L(a_t;\theta)}{\partial a_1 \partial \zeta} &= \frac{\Psi}{2\sqrt{\zeta}}\phi(\sqrt{\zeta}(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2}))\Big(1-\zeta(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2})^2\Big).
\end{aligned}$$

These inequalities imply that effort $a^*_\theta$ is increasing in $\Psi$, decreasing in $\lambda_M$, and $a^*_0$ is decreasing in $\lambda_S$. Furthermore, while the direct effect $\frac{\partial a^*_1}{\partial \lambda_S} = 0$, the indirect effect from $a^*_0$ is such that $a^*_1$ is decreasing in $\lambda_S$ as well. Also observe that $\Big(1-\zeta(a_1 - \frac{\hat{a}_1(1)+\hat{a}_1(0)}{2})^2\Big) > 0$ is positive as $\zeta \to 0$ and decreasing in $\zeta$ so that the effect is inverse U-shaped. □

*Proof of Proposition 2.* Proof is analogous to that of Proposition 1. 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

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

Given this cutoff the rest of the proof is identical with an identical characterization of the optimal first period effort. □

*Proof of Corollary 4.* Follows from the fact that the IO's report $s$ has the MLRP in $m(\theta)$. □

*Proof of Corollary 5.* Recall that the voter's cutoff is defined as

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

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

$$\beta - \beta\theta\lambda_M a - \beta(1-\theta)(\lambda_M + \lambda_S)a + \sqrt{\zeta}\phi(\sqrt{\zeta}(a - \frac{\hat{a}_1 + \hat{a}_0}{2} - \frac{\log(\frac{\hat{m}_0}{\hat{m}_1})}{\zeta(\hat{a}_1 - \hat{a}_0)}))\Psi = 0.$$

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 $s \to \pm\infty$, $a^*_\theta \to \frac{1}{\lambda_M + (1-\theta)\lambda_S}$.

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 $\tilde{a}_\theta > \frac{1}{\lambda_M + (1-\theta)\lambda_S}$ 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 $\frac{da^*_\theta}{ds} < 0$. □

**Lemma 1.** *If $\hat{p}_1 \neq \hat{p}_0$, the voter's threshold $\hat{x}(\hat{a}, \hat{p})$:*

- *increases in $\hat{p}_1$ if $s < \frac{1}{2}$ and decreases in $\hat{p}_1$ if $s > \frac{1}{2}$;*

- *decreases in $\hat{p}_0$ if $s < \frac{1}{2}$ and increases in $\hat{p}_0$ if $s > \frac{1}{2}$.*

*Proof of Lemma 1.* The voter's threshold is

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

where $\hat{m}_\theta = \hat{p}_\theta \phi(\sqrt{\tau}(s-1)) + (1-\hat{p}_\theta)\phi(\sqrt{\tau}s)$. Observe that

$$\frac{\partial \hat{m}_\theta}{\partial \hat{p}_\theta} = \phi(\sqrt{\tau}(s-1)) - \phi(\sqrt{\tau}s),$$

which is negative if $s < \frac{1}{2}$ and positive if $s > \frac{1}{2}$.

Differentiating with respect to $\hat{p}_1$ yields

$$\frac{\partial \hat{x}(\hat{a}, \hat{p})}{\partial \hat{p}_1} = -\frac{1}{\zeta(\hat{a}_1 - \hat{a}_0)\hat{m}_1} \frac{\partial \hat{m}_1}{\partial \hat{p}_1},$$

such that $\frac{\partial \hat{x}(\hat{a},\hat{p})}{\partial \hat{p}_1} > 0$ if $s < \frac{1}{2}$ and $\frac{\partial \hat{x}(\hat{a},\hat{p})}{\partial \hat{p}_1} < 0$ if $s > \frac{1}{2}$.

Similarly, differentiating with respect to $\hat{p}_0$ yields

$$\frac{\partial \hat{x}(\hat{a}, \hat{p})}{\partial \hat{p}_0} = \frac{1}{\zeta(\hat{a}_1 - \hat{a}_0)\hat{m}_0} \frac{\partial \hat{m}_0}{\partial \hat{p}_0},$$

such that $\frac{\partial \hat{x}(\hat{a},\hat{p})}{\partial \hat{p}_0} < 0$ if $s < \frac{1}{2}$ and $\frac{\partial \hat{x}(\hat{a},\hat{p})}{\partial \hat{p}_0} > 0$ if $s > \frac{1}{2}$. □

*Proof of Proposition 3.* The leader maximizes

$$\max_{m \in \{0,1\}} \int_{-\infty}^{\infty} \Big[\beta\Big(a_\theta^* - \frac{\lambda_M + (1-\theta)\lambda_S}{2} a_\theta^{*^2}\Big) + \pi(a_\theta^*(s))\Psi\Big]\phi(\sqrt{\tau}(s-m))ds,$$

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

$$\int_{-\infty}^{\infty}\Big[\beta\Big(a_\theta^* - \frac{\lambda_M + (1-\theta)\lambda_S}{2}a_\theta^{*2}\Big) + \pi(a_\theta^*(s))\Psi\Big]\phi(\sqrt{\tau}(s-1))ds \geq \int_{-\infty}^{\infty}\Big[\beta\Big(a_\theta^* - \frac{\lambda_M + (1-\theta)\lambda_S}{2}a_\theta^{*2}\Big) + \pi(a_\theta^*(s))\Psi\Big]\phi(\sqrt{\tau}s)ds,$$

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 honest 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 honest type sends $m = 1$ increases the return from playing $m = 1$ versus $m = 0$. For the captured 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 honest type would deviate to $m = 0$. Similarly, $\Delta_1(\hat{p}_1, 0) > 0$ for any $\hat{p}_0$, meaning the captured type would deviate to $m = 1$. Thus the only equilibrium is $p_1^* = p_0^* = p \in (0, 1)$. $\square$