Creating Confusion

Supplementary Online Appendix

Chris Edmond* Yang K. Lu†

August 2020

This supplementary appendix is organized as follows. In Appendix C and Appendix D we provide additional details, results, and proofs for the two extensions of our benchmark model discussed in the main text. In Appendix E, we discuss a further extension of our benchmark model where the citizens apply different weights to the components of their loss function. In Appendix F we provide proofs of additional results hitherto omitted. In Appendix G we provide further details on the knife-edge case $c = 1$. In Appendix H we show that the expositional device of assuming that the coefficients in the citizens' strategy sum to one is without loss of generality.

*University of Melbourne. Email: ⟨cedmond@unimelb.edu.au⟩
†Hong Kong University of Science and Technology. Email: ⟨yanglu@ust.hk⟩

C Active media

In this appendix we provide further details on the extension with an *active media*, where citizens obtain their information from a collection of journalists with preferences that may not perfectly reflect the preferences of the citizens.

C.1 Journalists' best response

To construct the journalists' best response, we start from the optimal action $a(x_j)$ for an individual journalist with signal x_j

$$a(x_j) = (1-\lambda)\mathbb{E}[\,\theta\,|\,x_j] + \lambda\mathbb{E}[\,A(\theta)\,|\,x_j].$$

If other journalists use $a(x_j) = kx_j + (1-k)z$ and the politician uses $y(\theta) = (1-\delta)\theta + \delta z$, then the aggregate action is

$$A(\theta) = ky(\theta) + (1-k)z = k(1-\delta)\theta + (1-k(1-\delta))z \tag{C1}$$

Collecting terms then gives

$$a(x_j) = (1-\lambda(1-k(1-\delta)))\,\mathbb{E}[\,\theta\,|\,x_j] + \lambda(1-k(1-\delta))\,z; \tag{C2}$$

which is a weighted average of the posterior and prior expectations.

The posterior expectation continues to be

$$\mathbb{E}[\,\theta\,|\,x_j] = \frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}x_j + \left(1 - \frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}\right)z \tag{C3}$$

Plugging this formula back into (C2) and matching coefficients we get the fixed-point condition

$$k = (1-\lambda(1-k(1-\delta)))\frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}$$

which has the unique solution

$$k(\delta) := \frac{(1-\delta)\alpha}{(1-\delta)^2\alpha + 1} \tag{C4}$$

where $\alpha := (1-\lambda)\alpha_x/\alpha_z$. In this notation, $k^*_{nm} = k(0)$.

C.2 Politician's welfare

The politician's value function continues to be

$$v(k) := \max_{\delta\in[0,1]} V(\delta, k) \tag{C5}$$

where $V(\delta, k)$ denotes the politician's ex-ante expected utility if they choose manipulation δ and the journalists have response coefficient k. This is again

$$V(\delta, k) = \frac{1}{\alpha_z}(B(\delta, k) - C(\delta)) + \frac{1}{\alpha_x}k^2 \tag{C6}$$

where as in our benchmark model, $B(\delta, k) := (k\delta + 1 - k)^2$ and $C(\delta) := c\delta^2$. In this notion, $v^* = v(k^*)$.

Let $v_{nm}(k)$ denote the politician's value function without manipulation

$$v_{nm}(k) := V(0, k) \leq \max_{\delta\in[0,1]} V(\delta, k) =: v(k) \tag{C7}$$

In this notation, $v^*_{nm} = v_{nm}(k^*_{nm})$.

When does manipulation backfire?

Supplementary Proposition 1.

(i) For each $\lambda < -1/2$ and $c < 1$, there exists a cutoff signal precision $\underline{\alpha}_x^*$ such that for all $\alpha_x < \underline{\alpha}_x^*$ the politician's manipulation backfires, $v^* < v_{nm}^*$.

(ii) For each $\lambda > +1/2$ and $c > 1$, there exists a cutoff signal precision $\overline{\alpha}_x^* > \underline{\alpha}_x^*$ such that for all $\alpha_x > \overline{\alpha}_x^*$ the politician's manipulation backfires, $v^* < v_{nm}^*$.

Proof. See Appendix F.2. □

To understand why backfiring can occur, notice that the manipulation technology has both direct and indirect effects on the politician's payoff. The direct effect benefits the politician by making the journalists' signals noisier than they would be absent manipulation. The indirect effect causes the journalists' equilibrium response coefficient to change from k_{nm}^* to k^*, which may or may not benefit the politician.

Backfiring occurs when the change from k_{nm}^* to k^* moves against the politician's interest by a sufficiently large amount. If $\lambda < 0$ the politician prefers higher k^* and backfiring will occur when journalists are sufficiently less responsive to their signals than they would be absent manipulation, i.e., when k^* is sufficiently smaller than k_{nm}^*. If $\lambda > 0$ the politician prefers lower k^* and backfiring will occur when journalists are sufficiently more responsive to their signals than they would be absent manipulation, i.e., when k^* is sufficiently larger than k_{nm}^*.

To see this, we decompose the change in the politician's payoff as

$$v^* - v_{nm}^* = (v(k^*) - v_{nm}(k^*)) + (v_{nm}(k^*) - v_{nm}(k_{nm}^*)) \tag{C8}$$

Since $v(k) \geq v_{nm}(k)$ for all k, the first term in the decomposition (C8) is not negative. So to obtain backfiring the second term $v_{nm}(k^*) - v_{nm}(k_{nm}^*)$ must be *sufficiently negative*. Now observe that this second term is a comparison of the function $v_{nm}(k)$ at two different points, k^* and k_{nm}^*, where $v_{nm}(k) := V(0, k)$ is given by[1]

$$v_{nm}(k) = \frac{1}{\alpha_z}(1-k)^2 + \frac{1}{\alpha_x}k^2. \tag{C9}$$

This quadratic in k decreases from $v_{nm}(0) = 1/\alpha_z$ till it reaches its global minimum at $k_{min} := \alpha_x/(\alpha_x + \alpha_z)$ and then increases to $v_{nm}(1) = 1/\alpha_x$. Now suppose the journalists' actions are strategic substitutes, $\lambda < 0$. Then $k_{nm}^* > k_{min}$ and so $v_{nm}(k)$ is strictly increasing on $(k_{nm}^*, 1)$. So if $\lambda < 0$ a necessary condition for $v_{nm}(k^*) - v_{nm}(k_{nm}^*) < 0$ is that $k^* < k_{nm}^*$. Similarly, if the journalists' actions are strategic complements, $\lambda > 0$, then $k_{nm}^* < k_{min}$ and so $v_{nm}(k)$ is strictly decreasing on $(0, k_{nm}^*)$. So if $\lambda > 0$ a necessary condition for $v_{nm}(k^*) - v_{nm}(k_{nm}^*) < 0$ is that $k^* > k_{nm}^*$.

Conditions on the primitives. We now establish conditions on the primitives sufficient to ensure that the gap between $v_{nm}(k^*)$ and $v_{nm}(k_{nm}^*)$ is indeed large enough that the politician's manipulation backfires. To do this we use:

Remark 1. Journalists are less responsive to their signals with manipulation

$$k^*(\alpha, c) < k_{nm}^*(\alpha) \qquad \text{if and only if} \qquad c < c_{nm}^*(\alpha) \tag{C10}$$

where

$$c_{nm}^*(\alpha) = \begin{cases} \dfrac{\alpha}{\alpha - 1}\left(\dfrac{\alpha}{\alpha+1}\right)^2 & \text{if } \alpha > 1 \\ +\infty & \text{if } \alpha \leq 1 \end{cases} \tag{C11}$$

Proof. From Lemma 1, if $\alpha \leq 1$ then $k(\delta)$ is decreasing in δ. Any $c < +\infty$ implies $\delta^*(\alpha, c) > 0$ and hence $k(\delta^*) < k(0)$. Recall that $k(0) = \alpha/(\alpha + 1) =: k_{nm}^*(\alpha)$. Therefore, if $\alpha \leq 1$, it is always the case that $k^*(\alpha, c) < k_{nm}^*(\alpha)$. With $\alpha > 1$, $k(\delta)$ is first increasing and then decreasing in δ. We then need to find

[1] This expression for $v_{nm}(k)$ can also be obtained as the limit of $v(k)$ from (**??**) as $c \to \infty$.

combinations of (α, c) that give $k^*(\alpha, c) = k^*_{nm}(\alpha)$. To do so, first determine the equilibrium δ^* that equates $k(\delta; \alpha)$ and $k^*_{nm}(\alpha)$, namely

$$\delta^*_{nm}(\alpha) = \frac{\alpha - 1}{\alpha}, \qquad \alpha > 1 \tag{C12}$$

Then solve for c that equates $\delta(k^*_{nm}(\alpha); c)$ and $\delta^*_{nm}(\alpha)$, namely

$$c = \frac{\alpha}{\alpha - 1}\left(\frac{\alpha}{\alpha + 1}\right)^2 =: c^*_{nm}(\alpha) \tag{C13}$$

(with $c^*_{nm}(\alpha) = +\infty$ for $\alpha \leq 1$). We now show that $k^*(\alpha, c) < k^*_{nm}(\alpha)$ iff $c < c^*_{nm}(\alpha)$. Observe that

$$\delta^*_{nm}(\alpha) = \frac{\alpha - 1}{\alpha} > \hat{\delta}(\alpha) \tag{C14}$$

where $\hat{\delta}(\alpha)$ is the critical point from Lemma 1. Hence $k(\delta; \alpha)$ is decreasing in δ for any $\delta \geq \delta^*_{nm}(\alpha)$. Now observe that $k(\delta^*_{nm}(\alpha); \alpha) = k^*_{nm}(\alpha)$ so that $k^*(\alpha, c) < k^*_{nm}(\alpha)$ iff $\delta^*(\alpha, c) > \delta^*_{nm}(\alpha)$. From Lemma 4 we know that $\delta^*(\alpha, c)$ is strictly decreasing in c hence any $c < c^*_{nm}(\alpha)$ is equivalent to $\delta^*(\alpha, c) > \delta^*_{nm}(\alpha)$. $\square$

In other words, if the composite parameter $\alpha \leq 1$ then we know that $k^* < k^*_{nm}$ regardless of c but if $\alpha > 1$ then the journalists' k^* is less than k^*_{nm} only if c is low enough.[2]

Now observe from (C9) that $v_{nm}(k)$ is a linear combination of the terms $(1 - k)^2$ and k^2 with the relative importance of the k^2 term decreasing in α_x. As α_x decreases, the function $v_{nm}(k)$ behaves more like the increasing k^2 term so that if $\lambda < 0$ and $k^* < k^*_{nm}$ then the second term in the decomposition $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ becomes more and more negative, eventually becoming negative enough that the net result is for the politician to be worse off. Similarly, as α_x increases, the function $v_{nm}(k)$ behaves more like the decreasing $(1 - k)^2$ term so that if $\lambda > 0$ and $k^* > k^*_{nm}$ the second term in the decomposition $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ becomes more and more negative, eventually becoming negative enough that the net result is that the politician is again worse off.

When does manipulation benefit? Although information manipulation can backfire on the politician, there are nonetheless clear situations where the politician benefits from information manipulation.

Supplementary Proposition 2. The politician benefits from manipulation, $v^* > v^*_{nm}$, if either:

(i) The journalists' actions are strategic substitutes, $\lambda \leq 0$, and the costs of manipulation are sufficiently high, $c > c^*_{nm}(\alpha)$, or

(ii) The journalists' actions are strategic complements, $\lambda \geq 0$, and the costs of manipulation are sufficiently low, $c < c^*_{nm}(\alpha)$.

Proof. Recall the decomposition (C8) above. Since $v(k) \geq v_{nm}(k)$ for all k, the first term is not negative, so for the politician to gain it is sufficient that the second term $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ is positive. When the journalists' actions are strategic substitutes, $\lambda < 0$, $v_{nm}(k)$ is strictly increasing on $(k^*_{nm}, 1)$ and hence $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ is positive if $k^* > k^*_{nm}$. From Remark 2 we know that $k^* > k^*_{nm}$ if and only if $c > c^*_{nm}(\alpha)$. Similarly, when the journalists' actions are strategic substitutes, $\lambda > 0$, $v_{nm}(k)$ is strictly decreasing on $(0, k^*_{nm})$ and hence $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ is positive if $k^* < k^*_{nm}$, which from Remark 2 happens if and only if $c < c^*_{nm}(\alpha)$. $\square$

These sufficient conditions guarantee that the introduction of the manipulation technology changes the journalists' equilibrium response coefficient from k^*_{nm} to k^* in a direction that benefits the politician, i.e., increasing to $k^* > k^*_{nm}$ if $\lambda < 0$ or decreasing to $k^* < k^*_{nm}$ if $\lambda > 0$. Notice that in the knife-edge special case with no interactions among journalists, $\lambda = 0$, the politician benefits from manipulation regardless of c.

Figure 8 illustrates both benefits from manipulation and backfiring in the same figure. The top row shows the politician's benefit from manipulation $v^* - v^*_{nm}$ as a function of the intrinsic precision α_x for the

[2]The function $c^*_{nm}(\alpha)$ is at first steeply decreasing in α, crosses $c^*_{nm}(\alpha) = 1$ and then reaches a minimum before increasing again, approaching $c = 1$ from below as $\alpha \to \infty$. So in the limit as $\alpha \to \infty$, the question of whether or not the equilibrium k^* is less than k^*_{nm} reduces to whether or not c is more or less than 1.

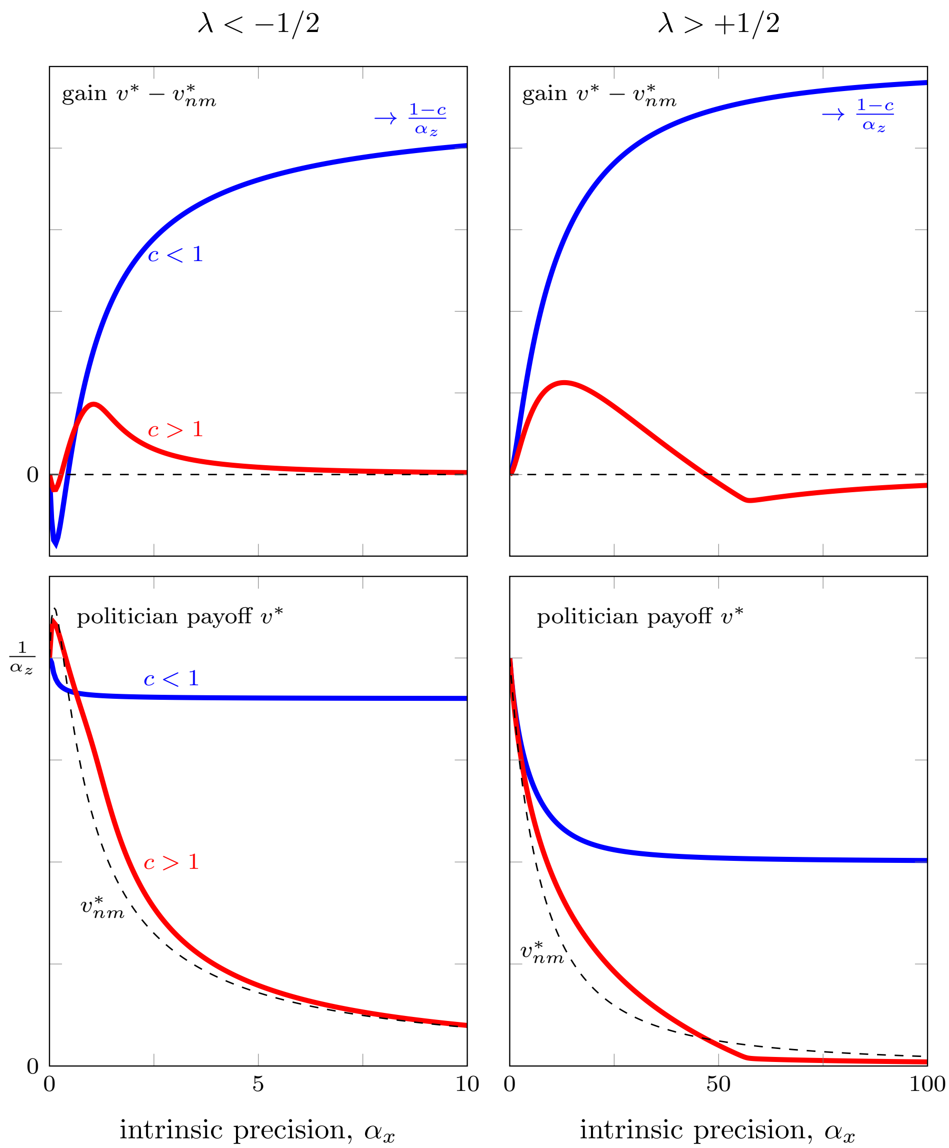


Figure 1: Politician benefits most when c is low and α_x is high.

Politician's benefit from manipulation $v^* - v^*_{nm}$ (top row) and payoff v^* (bottom row) as functions of the intrinsic precision α_x for various costs of manipulation c when the journalists' actions are strong strategic substitutes $\lambda < -1/2$ (left column) or strong strategic complements $\lambda > 1/2$ (right column). The politician's payoff absent manipulation v^*_{nm} asymptotes to zero as $\alpha_x \to \infty$. If $c > 1$ the politician's payoff with manipulation v^* also asymptotes to zero but if $c < 1$ then v^* asymptotes to $(1-c)/\alpha_z > 0$ so that the politician benefits. The politician benefits the most when when c is low and α_x is high. In the left column we use $\lambda < -1$ to highlight that for this parameter setting v^* and v^*_{nm} need not be monotonic in α_x.

case of low costs of manipulation, $c < 1$ (in blue), and the case of high costs of manipulation, $c > 1$ (in red). The bottom row shows the underlying levels v^* for $c < 1$ (in blue) and $c > 1$ (in red) along with the politician's welfare v^*_{nm} in the absence of manipulation (dashed black). The left column shows the results when the journalists' actions are strong strategic substitutes, $\lambda < -1/2$. The right column shows the results when the journalists' actions are strong strategic complements, $\lambda > +1/2$.

A striking feature of Figure 8 is that the politician gains the most from manipulation when c is low and α_x is high, regardless of λ.

C.3 Journalists' and citizens' welfare

Journalists. We first define the journalists' loss function

$$l_{\mathcal{J}}(\delta) := \min_{k\in[0,1]} L_{\mathcal{J}}(k,\delta) \tag{C15}$$

where $L_{\mathcal{J}}(k,\delta)$ denotes the journalists' ex-ante expected loss, i.e., the expectation of (**??**) in the main text with respect to the prior that θ is normally distributed with mean z and precision α_z, if they choose k when the politician has manipulation δ. This works out to be

$$L_{\mathcal{J}}(k,\delta) = \frac{1-\lambda}{\alpha_z}B(\delta,k) + \frac{1}{\alpha_x}k^2 \tag{C16}$$

where again $B(\delta,k) := (k\delta + 1 - k)^2$ denotes the politician's benefit from manipulation. Evaluating at the journalists' best response $k(\delta)$ and collecting terms gives

$$l_{\mathcal{J}}(\delta) = L_{\mathcal{J}}(k(\delta),\delta) = \frac{1}{\alpha_x}\left(\frac{k(\delta)}{1-\delta}\right) = \left(\frac{1}{1+\alpha(1-\delta)^2}\right)\frac{1-\lambda}{\alpha_z}. \tag{C17}$$

Citizens. The citizens evaluate outcomes according to the loss

$$\int_0^1 (a_j-\theta)^2\,dj = (A-\theta)^2 + \int_0^1 (a_j - A)^2\,dj \tag{C18}$$

So the citizens are at their bliss point if the journalists all produce $a_j = \theta$.

Now let $L_{\mathcal{C}}(k,\delta)$ denote the citizens' ex ante expected loss, i.e., the expectation of (C18) with respect to the prior that θ is normally distributed with mean z and precision α_z, if the *journalists* choose k when the politician has manipulation δ. This works out to be

$$L_{\mathcal{C}}(k,\delta) = \frac{1}{\alpha_z}B(\delta,k) + \frac{1}{\alpha_x}k^2 \tag{C19}$$

as in equation (**??**) in the main text.

Wedge between journalists' and citizens' losses. Comparing (C19) and (C16) we see that

$$L_{\mathcal{C}}(k,\delta) = L_{\mathcal{J}}(k,\delta) + \frac{\lambda}{\alpha_z}B(\delta,k) \tag{C20}$$

In the special case $\lambda = 0$, where the journalists care only about accurate reporting with no interactions amongst themselves, the citizens' loss and the journalists' loss *coincide*. More generally, since $B(\delta,k) \geq 0$, the citizens' loss is larger than the journalists' loss whenever $\lambda > 0$ and is less than the journalists' loss whenever $\lambda < 0$. Intuitively, an incentive to coordinate, $\lambda > 0$, means that individual journalists respond more to their common prior z than its underlying precision warrants. Therefore, from the citizens' point of view, the journalists are excessively responsive to their prior and hence *under-responsive* to the information contained in their signals. For example, if $\lambda \to 1$ the journalists can be quite content when they are producing similar reports, $a_i \approx A$, even if those reports are far from θ and hence very unsatisfactory from the citizens' point of view.

Effects of manipulation on journalists and citizens. Evaluating $L_{\mathcal{J}}(k,\delta)$ at the journalists' best response $k(\delta)$ we can then write

$$l_{\mathcal{C}}(\delta) := L_{\mathcal{C}}(k(\delta),\delta) = l_{\mathcal{J}}(\delta) + \frac{\lambda\alpha_z}{(1-\lambda)^2}\, l_{\mathcal{J}}(\delta)^2 \tag{C21}$$

The comparison of the citizens' equilibrium loss with and without manipulation is then reduced to comparing $l^*_{\mathcal{C}} = l_{\mathcal{C}}(\delta^*)$ and $l^*_{\mathcal{C},nm} = l_{\mathcal{C}}(0)$. Similarly, the comparison of the journalists' equilibrium loss with and without manipulation is reduced to comparing $l^*_{\mathcal{J}} = l_{\mathcal{J}}(\delta^*)$ and $l^*_{\mathcal{J},nm} = l_{\mathcal{J}}(0)$. Our main result here is:

Supplementary Proposition 3.

(i) The journalists are worse off with manipulation, $l^*_{\mathcal{J}} > l^*_{\mathcal{J},nm}$.

(ii) The citizens are worse off with manipulation, $l^*_{\mathcal{C}} > l^*_{\mathcal{C},nm}$, if $\lambda > -1$.

(iii) The citizens are *better off* with manipulation, $l^*_{\mathcal{C}} < l^*_{\mathcal{C},nm}$, if $\lambda < -1$ and $\alpha_x < \widehat{\alpha}^{**}_x$.

Proof. See Appendix F.2. □

So the journalists are always worse off with manipulation. Whether the citizens are worse off or not depends on the strategic interactions among the journalists. If the journalists' actions are not strong strategic substitutes, $\lambda > -1$, the citizens are also unambiguously worse off with manipulation. But if the journalists' actions are strong strategic substitutes, $\lambda < -1$, and if in addition the intrinsic precision of journalists' signals is low enough, $\alpha_x < \widehat{\alpha}^{**}_x$, then, perhaps surprisingly, the citizens are in fact *better off* with manipulation. To understand this, first notice that when $\lambda < -1$, the journalists have a strong incentive to differentiate themselves from one another and their response k to their idiosyncratic signals is, from the citizens' point of view, more than is warranted by the underlying precision of their signals. This is especially problematic for the citizens when the signals are imprecise, i.e., when α_x is very low. By reducing k, the politician's manipulation then "corrects" for this, which makes the citizens better off than they would be absent manipulation.[3]

Effects of α_x on journalists' loss. Notice from the journalists' loss function (C17) the strategic interaction term $(1-\lambda)$ and the prior precision α_z simply scale the whole loss. Similar to the citizens' loss (C17) in the benchmark model, we can measure the equilibrium payoffs for the journalists, $l^*_{\mathcal{J}} = l_{\mathcal{J}}(\delta^*)$, by their *indirect utility* evaluated at the equilibrium manipulation:

$$u^*(\alpha,c) = \alpha(1-\delta^*(\alpha,c))^2. \tag{C22}$$

This term is identical to (**??**) in the main text except that the composite precision parameter $\alpha := (1-\lambda)\alpha_x/\alpha_z$ now incorporates the effect of the strategic interactions. The welfare results on $u^*(\alpha,c)$ in Proposition 3 and Remark 1 of the main text therefore also apply to the equilibrium payoffs of the journalists in the extended model. In the following corollary, we restate the welfare results in terms of the journalists' loss and the intrinsic signal precision α_x:

Remark 2. The journalists' loss $l^*_{\mathcal{J}}$ is strictly decreasing in α_x if and only if $\alpha_x < \alpha^{**}_x$. For $c > 1$ the critical point $\alpha^{**}_x = +\infty$.

Proof. The journalists' loss $l^*_{\mathcal{J}}$ is proportional to $(1+u^*(\alpha,c))^{-1}$. Using part (i) and (ii) of Proposition 3 and the definition of the composite precision parameter $\alpha := (1-\lambda)\alpha_x/\alpha_z$, we obtain $\alpha^{**}_x = (\alpha_z/(1-\lambda))\alpha^*(c)$ where $\alpha^*(c)$ is the critical value in part (ii) of Proposition 3. □

[3] The region of the parameter space where the citizens are better off with manipulation is in a sense quite small. The critical point turns out to be

$$\widehat{\alpha}^{**}_x = -\left(\frac{1+\lambda}{(1-\lambda)^2}\right)\alpha_z, \qquad \lambda < -1$$

This is maximized at $\lambda = -3$ for which $\widehat{\alpha}^{**}_x = \alpha_z/8$. Even allowing the value of λ most favorable to this scenario, it only occurs if the intrinsic signal precision α_x is less than one-eighth of the prior precision α_z.

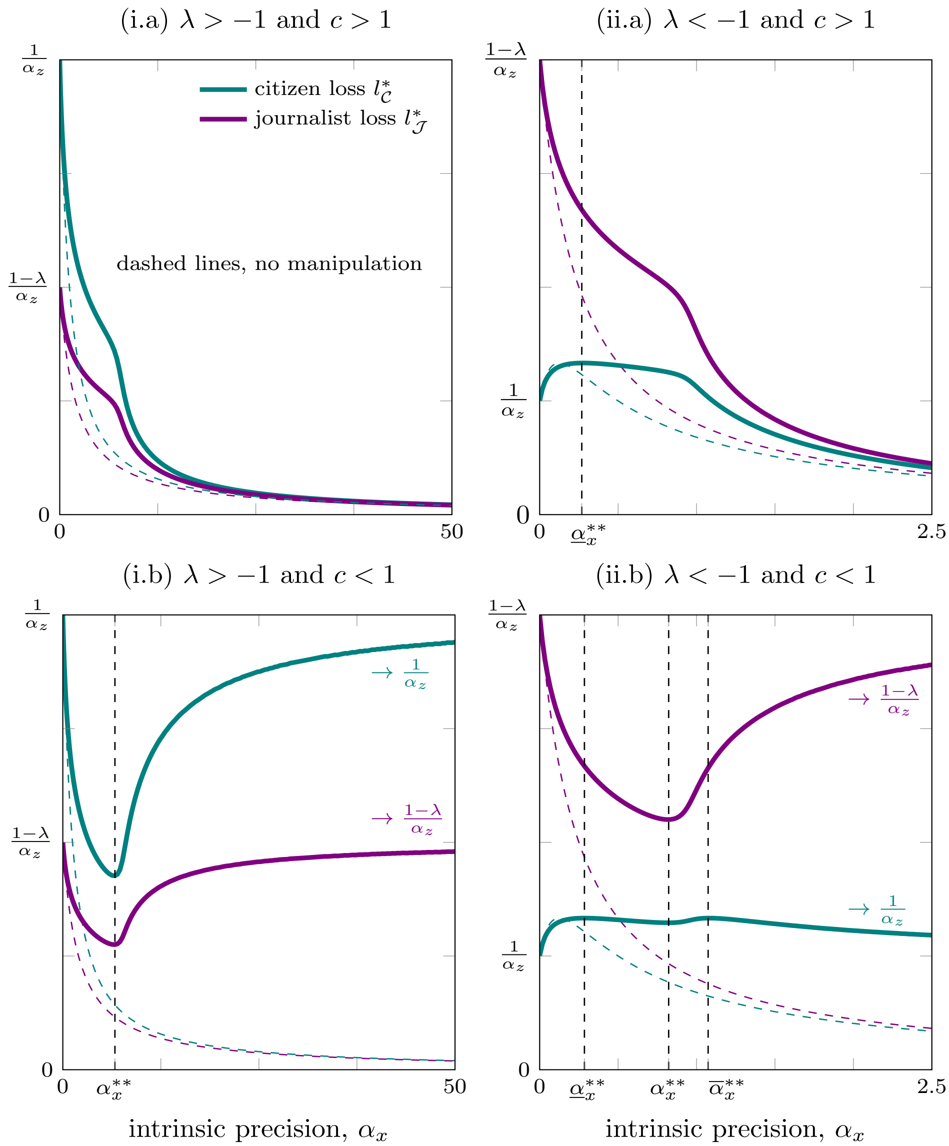


Figure 2: Citizens and journalists lose most when c is low and α_x is high.

Citizens' loss $l^*_\mathcal{C}$ and journalists' loss $l^*_\mathcal{J}$ as functions of α_x for $c > 1$ (top row) and $c < 1$ (bottom row) and for $\lambda > -1$ (left column) and $\lambda < -1$ (right column). If $\lambda > -1$ both loss functions move in the same direction in response to α_x. If $c > 1$ both loss functions are strictly decreasing (top left). If $c < 1$ both loss functions are $\cup$-shaped with critical point α_x^{**} (bottom left). If $\lambda < -1$ the loss functions move in the same direction only between the critical points $\underline{\alpha}_x^{**}$ and $\overline{\alpha}_x^{**}$ (right column). If $c < 1$ the citizens' loss asymptotes to $1/\alpha_z$ and the journalists' loss asymptotes to $(1-\lambda)/\alpha_z$. For the left column we use $\lambda > 0$ which implies that the journalists' loss is less than the citizens' loss. The colored dashed lines show the corresponding loss functions absent manipulation. If $\lambda < -1$ then for α_x sufficiently small the citizens are better off with manipulation.

Effects of α_x on citizens' loss. Evaluating the expression for the citizens' loss in (C21) at the equilibrium manipulation δ^* gives

$$l_{\mathcal{C}}^* = l_{\mathcal{J}}^* + \frac{\lambda \alpha_z}{(1-\lambda)^2} \, l_{\mathcal{J}}^{*\,2} \tag{C23}$$

Hence the effects of α_x on the citizens' equilibrium loss are given by the total derivative

$$\frac{dl_{\mathcal{C}}^*}{d\alpha_x} = \frac{dl_{\mathcal{J}}^*}{d\alpha_x}\left[1 + \frac{2\lambda\alpha_z}{(1-\lambda)^2} \, l_{\mathcal{J}}^*\right] \tag{C24}$$

This expression is convenient because all the effects of α_x enter $l_{\mathcal{C}}^*$ only through $l_{\mathcal{J}}^*$. This gives:

Supplementary Proposition 4. The citizens' loss $l_{\mathcal{C}}^*$ and the journalists' loss $l_{\mathcal{J}}^*$ move in the same direction in response to changes in α_x if and only if either (i) $\lambda > -1$, or (ii) $\lambda < -1$ and $\alpha_x \in (\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$. For $c > 1$, $\overline{\alpha}_x^{**} = +\infty$.

Proof. See Appendix F.2. □

Figure 2 illustrates the effects of α_x on the journalists' and citizens' loss. The left and right columns show the cases $\lambda > -1$ and $\lambda < -1$ respectively. The top and bottom rows show the cases $c > 1$ and $c < 1$ respectively. Each panel shows the loss of the citizens $l_{\mathcal{C}}^*$ and the journalists $l_{\mathcal{J}}^*$ as functions of α_x. The dashed lines demarcate the critical points α_x^{**} and $\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**}$. As with the journalists' loss, the limit of the citizens' loss as $\alpha_x \to \infty$ is sensitive to the costs of manipulation c. If $c < 1$, as $\alpha_x \to \infty$ the citizens's loss $l_{\mathcal{C}}^*$ asymptotes to the same loss $1/\alpha_z$ the citizens would have if $\alpha_x = 0$. If $c > 1$, the citizens's loss $l_{\mathcal{C}}^*$ asymptotes to zero, the same limit of the citizens' loss without manipulation, $l_{\mathcal{C},nm}^*$.

D Heterogeneous priors and manipulation of the signal variance

In this appendix we provide further details on the extension where citizens have heterogeneous priors and where the politician can manipulate the signal variance.

Setup. Given that the citizens have linear strategies $a(x_i, z_i) = kx_i + (1-k)z_i$ the politican's problem is now to choose y and γ to maximize

$$\begin{aligned} V(y,\gamma) &= \int_0^1 (k(y+\varepsilon_i) + (1-k)(z+\eta_i) - \theta)^2 \, di - c(y-\theta)^2 - c_\gamma(\gamma-1)^2 \\ &= (ky + (1-k)z - \theta)^2 + \gamma\sigma_x^2 k^2 + \sigma_\eta^2(1-k)^2 - c(y-\theta)^2 - c_\gamma(\gamma-1)^2 \end{aligned}$$

The first order condition for the variance manipulation factor γ can be written

$$\gamma(k) = 1 + \frac{\sigma_x^2}{2c_\gamma} k^2 \tag{D1}$$

And since the objective is separable in y and γ the first order condition for the signal mean y is as in the benchmark model

$$y(\theta) = (1-\delta)\theta + \delta z, \qquad \delta(k) = \frac{k-k^2}{c-k^2} \tag{D2}$$

where the second order condition again requires $c - k^2 \geq 0$. The optimal action for the citizens is $a_i = \mathbb{E}[\theta \,|\, x_i, z_i]$. The citizens have signals $x_i = y + \varepsilon_i = (1-\delta)\theta + \delta z + \varepsilon_i$ and prior $z_i = z + \eta_i = \theta + \varepsilon_z + \eta_i$. Using the properties of the bivariate normal distribution, conditional on x_i, z_i the citizens posterior for θ is normal with expected value

$$\mathbb{E}[\theta \,|\, x_i, z_i] = \frac{(1-\delta)\sigma_z^2 + \sigma_\eta^2}{(1-\delta)^2\sigma_z^2 + \sigma_\eta^2 + \gamma\sigma_x^2} \, x_i + \left(1 - \frac{(1-\delta)\sigma_z^2 + \sigma_\eta^2}{(1-\delta)^2\sigma_z^2 + \sigma_\eta^2 + \gamma\sigma_x^2}\right) z_i$$

Hence indeed the citizens have strategies of the form $a(x_i, z_i) = kx_i + (1-k)z_i$ where

$$k(\delta,\gamma) = \frac{(1-\delta)\sigma_z^2 + \sigma_\eta^2}{(1-\delta)^2\sigma_z^2 + \sigma_\eta^2 + \gamma\sigma_x^2} \tag{D3}$$

Equilibrium.

Supplementary Proposition 5. There is a unique equilibrium, that is, a unique triple k^*, δ^*, γ^* simultaneously satisfying the three best response functions (D1), (D2), and (D3)

Proof. Plugging the expressions for $\gamma(k)$ and $\delta(k)$ into (D3), we can write the equilibrium problem as solving the following fixed point problem in k

$$L(k) = R(k) \tag{D4}$$

analogous to (??), where now

$$L(k) := (k-1)\frac{\sigma_\eta^2}{\sigma_z^2} + k\left(1 + \frac{\sigma_x^2}{2c_\gamma}k^2\right)\frac{\sigma_x^2}{\sigma_z^2} \tag{D5}$$

and where

$$R(k) := c\frac{(c-k)(1-k)}{(c-k^2)^2} \tag{D6}$$

The curve $R(k)$ on the RHS is exactly the same as in (??) from the proof of Proposition 1 and is strictly decreasing in k with limits $R(0) = 1$ and $R(\min(\sqrt{c}, 1)) =: \underline{R}(c)$. The curve $L(k)$ on the LHS is a generalization of its counterpart in (??) and nests our benchmark model as a special case. In particular, when $\sigma_\eta^2 = 0$ and $c_\gamma \to +\infty$ the LHS reduces to

$$L(k) = k\,\frac{\sigma_x^2}{\sigma_z^2}, \qquad \{\sigma_\eta^2 = 0, \quad c_\gamma \to \infty\}$$

which, recognizing $\sigma_x^2/\sigma_z^2 = \alpha_z/\alpha_x = 1/\alpha$ is the same $L(k) = k/\alpha$ as in the benchmark model (??). Here the LHS is strictly increasing in k with limits $L(0) = -\sigma_\eta^2/\sigma_z^2 < 0$ and $L(\min(\sqrt{c}, 1)) := \overline{L}(c)$.

If $c \geq 1$, we clearly have $\overline{L}(c) = L(1) > 0$ and $\underline{R}(c) = R(1) = 0$, so by the intermediate value theorem, there is a unique $k^* \in [0, 1]$ solving $L(k^*) = R(k^*)$.

If $c < 1$, as $k \to \sqrt{c}$ on the LHS we have finite limit

$$\overline{L}(c) = L(\sqrt{c}) = (\sqrt{c}-1)\frac{\sigma_\eta^2}{\sigma_z^2} + \sqrt{c}\left(1 + \frac{\sigma_x^2}{2c_\gamma}c\right)\frac{\sigma_x^2}{\sigma_z^2} \tag{D7}$$

whereas on the RHS we have

$$\underline{R}(c) = R(\sqrt{c}) = c(c-\sqrt{c})(1-c)\lim_{k\to\sqrt{c}}\frac{1}{(c-k^2)^2} = -\infty \tag{D8}$$

since $c < \sqrt{c} < 1$. Hence by the intermediate value theorem there is a unique $k^* \in [0, \sqrt{c}]$ such that $L(k^*) = R(k^*)$. The equilibrium γ^* and δ^* are in turn determined by the best response functions $\gamma(k)$ and $\delta(k)$ evaluated at k^*. □

Comparative statics.

Supplementary Proposition 6. The equilibrium signal variance ${\sigma_x^2}^* = \gamma^*\sigma_x^2$ is:

(i) Increasing in the intrinsic signal variance σ_x^2, and

(ii) Increasing in the prior dispersion σ_η^2.

Proof. For part (i), from the best response (D1), the dervivative of ${\sigma_x^2}^* = \gamma^*\sigma_x^2$ is

$$\frac{d}{d\sigma_x^2}\,{\sigma_x^2}^* = \frac{d}{d\sigma_x^2}\left(\sigma_x^2 + \frac{1}{2c_\gamma}\left(k{\sigma_x^2}^*\right)^2\right) = 1 + \frac{k\sigma_x^2}{c_\gamma} + \frac{k\sigma_x^4}{c_\gamma}\frac{dk}{d\sigma_x^2}$$

which is positive if and only if

$$\frac{dk}{d\sigma_x^2} > -\frac{1 + \frac{k^2\sigma_x^2}{c_\gamma}}{\frac{k\sigma_x^4}{c_\gamma}} = -\frac{k + \frac{k^3\sigma_x^2}{c_\gamma}}{\frac{k^2\sigma_x^4}{c_\gamma}} \tag{D9}$$

The equilibrium k^* is the solution to the fixed point problem (D4), which can be written:

$$H\left(k,\sigma_x^2\right) := k\left(\frac{c-k}{c-k^2}\right)^2\sigma_z^2 + k\left(\sigma_\eta^2+\sigma_x^2\right) + \frac{k^3}{2c_\gamma}\left(\sigma_x^2\right)^2 - \frac{c-k}{c-k^2}\sigma_z^2 - \sigma_\eta^2 = 0 \tag{D10}$$

Using the implicit function theorem:

$$\frac{dk}{d\sigma_x^2} = -\frac{\frac{\partial H}{\partial\sigma_x^2}}{\frac{\partial H}{\partial k}} = -\frac{k + k^3\frac{\sigma_x^2}{c_\gamma}}{\frac{3}{2}k^2\frac{\sigma_x^4}{c_\gamma} + \left(\sigma_\eta^2+\sigma_x^2\right) + \sigma_z^2\left[\frac{\partial}{\partial k}\left(k\left(\frac{c-k}{c-k^2}\right)^2 - \frac{c-k}{c-k^2}\right)\right]} \tag{D11}$$

Comparing the denominators of the RHS of inequalities (D9) and (D11), we find that a sufficient condition for (D9) to hold is

$$\frac{\partial}{\partial k}\left(k\left(\frac{c-k}{c-k^2}\right)^2 - \frac{c-k}{c-k^2}\right) > 0 \tag{D12}$$

The derivative works out to be

$$\left(\frac{c-k}{c-k^2}\right)^2 + \frac{\left(2kc-k^2-c\right)^2}{\left(c-k^2\right)^3} \tag{D13}$$

which is strictly positive since from the second order condition of the politician's maximization, $c-k^2>0$.

For part (ii), from the best response (D1), we have $\sigma_x^{2\,*} = \gamma^*\sigma_x^2$ is increasing in σ_η^2 if and only if k^* is increasing in σ_η^2. Recall that k^* is determined in the fixed point problem (D4). Then observe that $R(k)$ is independent of $\left(\sigma_x^2,\sigma_\eta^2\right)$ and $L(k)$ is decreasing in σ_η^2. Combining with the fact that $R'(k)<0$ and $L'(k)>0$, we can conclude that k^* is increasing in σ_η^2. □

"Regime changes" in the amount of manipulation. First we define the mapping:

$$F\left(k,\sigma_z^2,\sigma_x^2,\sigma_\eta^2,c_\gamma\right) := \frac{\sigma_z^2}{\sigma_z^2 + \sigma_x^4 k^3(1+k)/c_\gamma + \sigma_\eta^2(1+k)} \tag{D14}$$

To facilitate some comparisons, in some of the expressions below we return to precision notation $\alpha_x = 1/\sigma_x^2$ and $\alpha_z = 1/\sigma_z^2$ with relative precision $\alpha = \alpha_x/\alpha_z$ etc. With this notation in mind we then have a result analogous to Proposition 2 in the main text:

Supplementary Proposition 7.

(i) For each $\alpha < 2+\sqrt{4+2/(c_\gamma\alpha_z)}$, the politician's equilibrium manipulation δ^* is smoothly decreasing in c with

$$\left.\frac{d\delta^*}{dc}\right|_{c=1} = -\frac{k^*}{(1-k^*)(1+k^*)^2 + (1-k^*)^2k^*F\left(k^*,\sigma_z^2,\sigma_x^2,\sigma_\eta^2,c_\gamma\right)} < 0 \tag{D15}$$

This derivative approaches $-\infty$ as $\alpha\to 2+\sqrt{4+2/(c_\gamma\alpha_z)}$.

(ii) For each $\alpha > 2+\sqrt{4+2/(c_\gamma\alpha_z)}$, the politician's manipulation jumps discontinuously from $\overline{\delta}(\alpha,\alpha_x)$ as $c\to 1^-$ to $\underline{\delta}(\alpha,\alpha_x)$ as $c\to 1^+$ where

$$\underline{\delta}(\alpha,\alpha_x),\, \overline{\delta}(\alpha,\alpha_x) = \frac{1}{2}\left(1\pm\sqrt{1-(4/\alpha)-2/(c_\gamma\alpha\alpha_x)}\right).$$

The size of the jump $\overline{\delta}(\alpha,\alpha_x) - \underline{\delta}(\alpha,\alpha_x)$ is increasing in α_x and independent of σ_η^2.

(iii) For any $c>1$, the politician's equilibrium manipulation δ^* is bounded above by $1/2$ and can be made arbitrarily close to zero by making α_x large enough.

Proof. For part (i), applying the implicit function theorem to $\delta^* = \delta\left(k^*\left(\delta^*\right),c\right)$, we obtain

$$\frac{d\delta^*}{dc} = \left(\frac{1}{1-\delta'(k^*)\,k'(\delta^*)}\right)\frac{\partial\delta(k^*,c)}{\partial c} \tag{D16}$$

just as in equation (**??**), but now we have

$$k'\left(\delta^*\right)\Big|_{c=1} = -\frac{\left(1-k^*\right)\sigma_z^2/(1+k^2)}{\left(\sigma_x^2\right)^2\left(k^*\right)^2/c_\gamma + \sigma_z^2/\left(k^* + \left(k^*\right)^2\right) + \sigma_\eta^2/k^*}$$

$$\delta'\left(k^*\right)\Big|_{c=1} = \frac{1}{\left(1+k^*\right)^2}$$

$$\frac{\partial\delta\left(k^*, c\right)}{\partial c}\Big|_{c=1} = -\frac{k^*}{\left(1-k^*\right)\left(1+k^*\right)^2}.$$

Substituting these expressions in (D16) and using the definition of $F\left(k, \sigma_z^2, \sigma_x^2, \sigma_\eta^2, c_\gamma\right)$ in (D14) and simplifying then gives the expression in (D15).

Notice that $F\left(k^*, \sigma_z^2, \sigma_x^2, \sigma_\eta^2, c_\gamma\right) = 1$ if $c_\gamma = \infty$ and $\sigma_\eta^2 = 0$, which is our benchmark case where

$$\frac{d\delta^*}{dc}\Big|_{c=1} = -\frac{k^*}{\left(1-k^*\right)\left(1+k^*\right)^2 + \left(1-k^*\right)^2 k^*}$$

This derivative approaches $-\infty$ as $k^* \to 1$. In turn, as in the main text, k^* is increasing in α and approaches 1 as α becomes sufficiently large.

For part (ii), $\underline{\delta}\left(\alpha, \alpha_x\right)$ and $\overline{\delta}\left(\alpha, \alpha_x\right)$ are the roots of

$$k\left(\delta, 1\right) = \frac{\left(1-\delta\right)\sigma_z^2 + \sigma_\eta^2}{\left(1-\delta\right)^2\sigma_z^2 + \sigma_\eta^2 + \sigma_x^2 + \frac{1}{2c_\gamma}\left(\sigma_x^2\right)^2} = 1$$

The roots exist when $\alpha \geq 2 + \sqrt{4 + 2/\left(c_\gamma\alpha_z\right)}$. In the knife-edge case $\alpha = 2 + \sqrt{4 + 2/\left(c_\gamma\alpha_z\right)}$ exactly, the two roots are the same and are equal to 1/2 so that the unique equilibrium is $(k^* = 1, \delta^* = 1/2)$.

For part (iii), the proof is the same as the part (iii) of Proposition 2. □

E Weights on components of citizens' loss function

In this appendix, we discuss a further extension to our benchmark model that allows the citizens to have different weights on the $(a_i - \theta)^2$ and $(A - \theta)^2$ terms in their loss function (in other words, allows the citizens to weigh these *discord* and *disinformation* terms differently to the policitican). In particular, suppose each citizen seeks to minimize the expected loss

$$l_i = (a_i - \theta)^2 + \omega(A - \theta)^2 \tag{E1}$$

The weight ω does not affect their optimal action so we still have the usual

$$a_i = kx_i + (1-k)z \tag{E2}$$

where k is given by

$$k(\delta) = \frac{(1-\delta)\alpha}{(1-\delta)^2\alpha + 1}, \qquad \alpha := \alpha_x/\alpha_z \tag{E3}$$

The weight ω plays no role in determining the equilibrium outcomes k^*, δ^* but it does affect how those outcomes are evaluated. The citizens' ex ante expected loss is now

$$L(k, \delta; \omega) = \frac{1+\omega}{\alpha_z}B(k, \delta) + \frac{1}{\alpha_x}k^2 \tag{E4}$$

where as usual $B(k, \delta) = (k\delta + 1 - k)^2$. Notice that $\omega = 0$ is our benchmark model. Now write this as

$$L(k, \delta; \omega) = L(k, \delta; 0) + \frac{\omega}{\alpha_z}B(k, \delta). \tag{E5}$$

And recall that at the best response

$$B(k(\delta), \delta) = (1 - k(\delta)(1-\delta))^2 = \left(\frac{k(\delta)}{\alpha(1-\delta)}\right)^2 \tag{E6}$$

so that we can write

$$L(k, \delta; 0) = \frac{1}{\alpha_x}\frac{k(\delta)}{1-\delta}. \tag{E7}$$

which then implies

$$B(k(\delta), \delta) = \left(\frac{\alpha_x}{\alpha} L(k, \delta; 0)\right)^2 = (\alpha_z L(k, \delta; 0))^2 \tag{E8}$$

so that we can then write

$$L(k, \delta; \omega) = L(k, \delta; 0) + \omega\alpha_z L(k, \delta; 0)^2 \tag{E9}$$

Now let $l^*(\omega) := L(k^*, \delta^*; \omega)$. In this notation $l^*(0)$ is the citizens' loss in our benchmark model. We then have the following welfare result analogous to Supplementary Proposition 4 in Appendix C above.

REMARK 3. $l^*(\omega)$ and $l^*(0)$ move in the same direction in response to changes in α_x if and only if either (i) $\omega > -1/2$, or (ii) $\omega < -1/2$ and $\alpha_x \in (\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$. For $c > 1$, $\overline{\alpha}_x^{**} = +\infty$.

Proof. From (E9) we have

$$\frac{dl^*(\omega)}{d\alpha_x} = (1 + 2\omega\alpha_z l^*(0)) \times \frac{dl^*(0)}{d\alpha_x}$$

So the two derivatives share the same sign if and only if

$$1 + 2\omega\alpha_z l^*(0) > 0 \tag{E10}$$

Clearly $\omega \geq 0$ suffices for the inequality above. When $\omega < 0$, the inequality above can be written as

$$l^*(0) < -\frac{1}{2\omega\alpha_z} =: l_{crit} \tag{E11}$$

We know from Proposition 3 and Remark 1 that the maximum of $l^*(0)$ is $l^*_{max} = 1/\alpha_z$. If $l^*_{max} < l_{crit}$, i.e., if $\omega > -1/2$, the inequality (E11) holds. Alternatively, if $l^*_{max} < l_{crit}$, i.e., if $\omega < -1/2$, there exists a subset of α_x such that the inequality (E11) does not hold. For any $c > 1$, $l^*(0)$ is strictly decreasing in α_x with $\lim_{\alpha_x \to 0^+} l^*(0) = l^*_{max}$ and $\lim_{\alpha_x \to \infty} l^*(0) = 0$. For any $c < 1$, $l^*(0)$ is strictly decreasing in α_x if and only if $\alpha_x < \alpha_x^{**}$, and $\lim_{\alpha_x \to 0^+} l^*(0) = \lim_{\alpha_x \to \infty} l^*(0) = l^*_{max}$. Using the same argument as in the proof of Supplementary Proposition 4, we can conclude that conditional on $\omega < -1/2$, the inequality (E11) holds if and only if $\alpha_x \in (\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$. □

F Omitted proofs

In this appendix we provide proofs of results hitherto omitted. We first state and prove two supplementary lemmas used in proof of Proposition 3 in the main text. We then provide proofs of the supplemantary propositions in the extension with active media in Appendix C above.

F.1 Further details from proof of Proposition 3 in main text

SUPPLEMENTARY LEMMA 1. The total derivative of the journalists' equilibrium loss l^* with respect to α is strictly positive if and only if

$$F(k^*) := k^{*4} - 2k^{*3} + 2ck^* - c^2 > 0 \tag{F1}$$

Proof. Recall that $l^* = l(\delta^*; \alpha)$ where

$$l(\delta; \alpha) = \frac{1}{(1-\delta)^2\alpha + 1} \tag{F2}$$

From this we obtain

$$\frac{dl^*}{d\alpha} > 0 \qquad \Leftrightarrow \qquad (1-\delta^*) - 2\alpha\frac{d\delta^*}{d\alpha} < 0 \tag{F3}$$

Equivalently, if and only if

$$\frac{d\delta^*}{d\alpha} > \frac{1}{2\alpha}(1-\delta^*) > 0 \tag{F4}$$

Now recall that in equilibrium the politician's manipulation depends on α_x only via the journalists' response coefficient, $\delta^*(\alpha) = \delta(k^*(\alpha))$, so that

$$\frac{d\delta^*}{d\alpha} = \delta'(k^*)\frac{dk^*}{d\alpha} \tag{F5}$$

So we can write condition (F4) as

$$\delta'(k^*)\frac{dk^*}{d\alpha} > \frac{1}{2\alpha}(1-\delta^*) > 0 \tag{F6}$$

Applying the implicit function theorem to the equilibrium condition (**??**) from the main text we have

$$\frac{dk^*}{d\alpha} = \frac{\frac{1}{\alpha^2}k^*}{\frac{1}{\alpha} - R'(k^*)} > 0 \tag{F7}$$

where $R(k)$ is defined in (**??**) in the main text. Plugging this into (F6) and simplifying we get the equivalent condition

$$\frac{1}{\alpha}\left(\delta'(k^*)k^* - \frac{1}{2}(1-\delta^*)\right) > -\frac{1}{2}(1-\delta^*)R'(k^*) \tag{F8}$$

Now observe from (**??**) in the main text that

$$\delta'(k^*)k^* - \frac{1}{2}(1-\delta^*) = \frac{1}{2}\left(\frac{1}{c-k^{*2}}\right)^2\left(k^{*3} - 3ck^{*2} + 3ck^* - c^2\right) \tag{F9}$$

and that using the formula for $R'(k)$ given in (**??**) in the main text we can calculate that

$$\frac{1}{2}(1-\delta^*)R'(k^*) = \frac{1}{2}\left(\frac{1}{c-k^{*2}}\right)^2 R(k^*)\frac{1}{1-k^*}P(k^*) \tag{F10}$$

where $P(k)$ is also defined in (**??**). Plugging these calculations back into (F8) gives

$$\frac{1}{\alpha}\left(\frac{1}{2}\left(\frac{1}{c-k^{*2}}\right)^2\left(k^{*3} - 3ck^{*2} + 3ck^* - c^2\right)\right) > -\frac{1}{2}\left(\frac{1}{c-k^{*2}}\right)^2 R(k^*)\frac{1}{1-k^*}P(k^*) \tag{F11}$$

Canceling common terms gives the condition

$$\frac{1}{\alpha}\left(k^{*3} - 3ck^{*2} + 3ck^* - c^2\right) > -R(k^*)\frac{1}{1-k^*}P(k^*) \tag{F12}$$

Using the equilibrium condition $L(k^*) = R(k^*)$ from (**??**) gives

$$\frac{1}{\alpha}\left(k^{*3} - 3ck^{*2} + 3ck^* - c^2\right) > -\frac{1}{\alpha}\frac{k^*}{1-k^*}P(k^*) \tag{F13}$$

Using the definition of $P(k)$ and canceling more common terms gives the condition

$$F(k^*) := k^{*4} - 2k^{*3} + 2ck^* - c^2 > 0 \tag{F14}$$

□

Supplementary Lemma 2. Define

$$F(k) := k^4 - 2k^3 + 2ck - c^2 \tag{F15}$$

(i) If $c > 1$, then $F(k) < 0$;

(ii) If $c < 1$, there is an interval $(\underline{k}, \overline{k})$ with $0 < \underline{k} < \overline{k} < 1$ such that $F(k) > 0$ for $k \in (\underline{k}, \overline{k})$ and $F(k) \leq 0$ otherwise. Moreover, the cutoffs are on either side of c so that $0 < \underline{k} < c < \overline{k} < 1$.

Proof. Write $F(k) = J(k;c) - G(k)$ where $J(k;c) := 2ck - c^2$ and $G(k) := 2k^3 - k^4$. Observe that $G(0) = 0$, $G(1) = 1$, $G(k) < k$ for all k; $G'(k) = 2k^2(3-2k) \geq 0$ with $G'(0) = 0$ and $G'(1) = 2$; and $G''(k) = 12k(1-k) \geq 0$ so that $G'(k) \leq G'(1) = 2$ for all k. Further observe that $J(0;c) = -c^2 < 0$, $J(1;c) = 2c - c^2 \leq 1$ (with equality if $c = 1$) and $J'(k;c) = 2c > 0$ for all k so that $J(k;c) \leq J(1;c) = 2c - c^2 \leq 1$ for all k, c. These imply $F(0) = J(0;c) - G(0) = -c^2 < 0$ and $F(1) = J(1;c) - G(1) = 2c - c^2 - 1 \leq 0$ (with equality if $c = 1$); $F'(k) = J'(k;c) - G'(k) = 2c - G'(k)$ and $F''(k) = -G''(k) \leq 0$. Since $G'(k) \leq 2$ we have

$$F'(k) = J'(k;c) - G'(k) = 2c - G'(k) \geq 2c - 2 = 2(c-1) \tag{F16}$$

For part (i) $c > 1$. Then $F'(k) \geq 2(c-1) > 0$ so $F(k)$ is strictly increasing from $F(0) = -c^2 < 0$ to $F(1) = 2c - c^2 - 1 < 0$ so that $F(k) < 0$ for all k.

For part (ii) $c < 1$. Then since $G'(k)$ is monotone increasing from $G'(0) = 0$ to $G'(1) = 2$ there is a unique critical point $\tilde{k}$ such that

$$F'(\tilde{k}) = 0 \qquad \Leftrightarrow \qquad 2c = G'(\tilde{k}) \tag{F17}$$

Since $F''(k) \leq 0$, this critical point maximizes $F(k)$ hence

$$F(k) \leq \max_{k \in [0,1]} F(k) = F(\tilde{k}) \tag{F18}$$

and observe that if we take $k = c < 1$ (which is feasible since here $c < 1$) then we have

$$F(c) = J(c;c) - G(c) = 2c^2 - c^2 - G(c) = c^2 - 2c^3 + c^4 = c^2(1 - 2c + c^2) > 0 \tag{F19}$$

so that indeed

$$F(\tilde{k}) \geq F(c) > 0 \tag{F20}$$

Hence for $c < 1$ there exist k such that $F(k) > 0$. More precisely, the function $F(k)$ increases from $F(0) = -c^2 < 0$ to a lower cutoff $\underline{k} \in (0, \tilde{k})$ defined by $F(\underline{k}) = 0$. The function $F(k)$ keeps increasing until it reaches the critical point $\tilde{k}$ at which $F'(\tilde{k}) = 0$ and $F(\tilde{k}) > 0$. From there $F(k)$ decreases, crossing zero again at a higher cutoff $\overline{k} \in (\tilde{k}, 1)$ defined by $F(\overline{k}) = 0$ and keeps decreasing until $F(1) = 2c - c^2 - 1 < 0$ (since $c < 1$).

So for $c < 1$ there is an interval $(\underline{k}, \overline{k})$ with $0 < \underline{k} < \overline{k} < 1$ such that $F(k) > 0$ for $k \in (\underline{k}, \overline{k})$ and $F(k) \leq 0$ otherwise. For $c < 1$ these critical points are defined by the roots of $F(k;c) = 0$. Observe that since $F(c) > 0$ yet $\underline{k}$ is the first k for which $F(k) = 0$ it must be the case that $\underline{k} < c$. Likewise since $F(\overline{k}) = 0$ it must also be the case that $\overline{k} > c$. In short, the cutoffs are on either side of c so that $0 < \underline{k} < c < \overline{k} < 1$. □

F.2 Proofs of additional results from extension with active media

Proof of Supplementary Proposition 1

Using the fixed point condition (**??**) with the redefined $\alpha = (1-\lambda)\alpha_x/\alpha_z$, we can write

$$v^* = \frac{1}{(1-\lambda)\alpha_x}\left\{k^* - \lambda k^{*2} + \frac{k^{*2}(1-k^*)^2}{c - k^*}\right\} \tag{F21}$$

Using the analogous condition for k^*_{nm}, we can write

$$v^*_{nm} = \frac{1}{(1-\lambda)\alpha_x}\left\{k^*_{nm} - \lambda k^{*2}_{nm}\right\} \tag{F22}$$

Hence the politician's manipulation backfires, $v^* < v^*_{nm}$, if and only if

$$g(k^*) < f(k^*_{nm}) - f(k^*) \tag{F23}$$

where

$$f(k) := k - \lambda k^2, \qquad g(k) := \frac{k^2(1-k)^2}{c-k} \geq 0 \tag{F24}$$

For part (i) suppose that $\lambda < 0$. We know from (C8) and (C9) that a necessary condition for the politician's manipulation to backfire is $k < k^*_{nm}$. We can rewrite the inequality in (F23) as

$$\frac{k^{*2}(1-k^*)^2}{c-k^*} < (k^*_{nm} - k)(1 - \lambda(k^*_{nm} + k^*)) \tag{F25}$$

Using the fixed point conditions (**??**) for both k^* and k^*_{nm}, we can rewrite the key condition (F25) as

$$\lambda < \frac{1}{c-k^*}\frac{K_1K_2}{K_3K_4} \tag{F26}$$

where

$$\begin{aligned}
K_1 &:= 4ck^{*2} - c^2 - k^{*3} - 2ck^{*3} - k^{*4} + k^{*5} \\
K_2 &:= c(c-k^*)(1-k^*) + k^*(c-k^{*2})^2 > 0 \\
K_3 &:= k^{*3} - 2ck^* + c > 0 \\
K_4 &:= (1+k^*)(c-k^{*2})^2 + c(c-k^*)(1-k^*) > 0
\end{aligned}$$

Now consider taking $\alpha_x \to 0$ for fixed $\lambda < 0$ such that $k^* \to 0$. We then have the following limits

$$K_1 \to -c^2, \quad K_2 \to +c^2, \quad K_3 \to c, \quad K_4 \to 2c^2$$

So in the limit the RHS of (F26) is

$$\frac{1}{c-k^*}\frac{K_1K_2}{K_3K_4} \to \frac{1}{(c-0)}\frac{(-c^2)(c^2)}{(c)(2c^2)} = -\frac{1}{2} \tag{F27}$$

Hence for any $\lambda < -1/2$ we can find α_x sufficiently close to zero such that (F26) is satisfied and in turn the politician's manipulation backfires, $v^* < v^*_{nm}$.

For part (ii), suppose that $\lambda > 0$. We know from (C8) and (C9) that the necessary condition for the politician's manipulation to backfire is $k^* > k^*_{nm}$. We can rewrite the inequality in (F23) as

$$\frac{k^{*2}(1-k^*)^2}{k^* - k^*_{nm}} < (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) \tag{F28}$$

Using the fixed point conditions (**??**) for both k^* and k^*_{nm}, we can rewrite this key condition as

$$\frac{k^{*2}(1-k^*)}{\frac{k^*_{nm}}{k^*}\left(c\frac{(c-k^*)}{(c-k^{*2})^2}\right) - 1} < (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) \tag{F29}$$

Observe that if, in addition, $c > 1$ and $\lambda > 1/2$, then the RHS of (F29) converges to a strictly positive constant

$$\lim_{\alpha_x \to \infty} (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) = (\lambda 2 - 1)(c - 1) > 0 \tag{F30}$$

(since $k^* \to 1$ if $c > 1$). But the LHS of (F29) converges to zero

$$\lim_{\alpha_x \to \infty} \frac{k^{*2}(1-k^*)}{\frac{k^*_{nm}}{k^*} c\frac{c-k^*}{(c-k^{*2})^2} - 1} = \frac{0^+}{\frac{c}{(c-1)} - 1} = 0^+ \tag{F31}$$

Therefore, if $k^* > k^*_{nm}$, $c > 1$ and $\lambda > 1/2$ then there exists α^*_x such that for $\alpha_x > \alpha^*_x$ the LHS of (F29) is strictly less than the RHS of (F29) so that the politician's manipulation backfires, $v^* < v^*_{nm}$.

Finally, we know from Remark 2 that $k^* < k^*_{nm}$ if and only if $c < c^*_{nm}(\alpha)$. Also observe that $c < 1$ is sufficient for $c < c^*_{nm}(\alpha)$ if $1 < c^*_{nm}(\alpha)$. From (C11) we have $1 < c^*_{nm}(\alpha)$ if $\alpha < 1$, or if $\alpha > 1$ and $\alpha < (1+\sqrt{5})/2$. Since $\alpha = (1-\lambda)\alpha_x/\alpha_z$, the critical point $\underline{\alpha}^*_x$ must be

$$\underline{\alpha}^*_x < \left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{\alpha_z}{1-\lambda}\right) \tag{F32}$$

Likewise, $c > 1$ is sufficient condition for $c > c^*_{nm}(\alpha)$ if $1 > c^*_{nm}(\alpha)$, and we need $\alpha > (1+\sqrt{5})/2$ to ensure that $1 > c^*_{nm}(\alpha)$. Since $\alpha = (1-\lambda)\alpha_x/\alpha_z$, the critical point $\overline{\alpha}^*_x$ must be

$$\overline{\alpha}^*_x > \left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{\alpha_z}{1-\lambda}\right) \tag{F33}$$

□

Proof of Supplementary Proposition 3.

For part(i), $l_{\mathcal{J}}(\delta)$ is increasing in δ from equation (C17). Hence, $l_{\mathcal{J}}(\delta^*) > l_{\mathcal{J}}(0)$ whenever $\delta^* > 0$. We then have that the journalists are unambiguously worse off when the politician can manipulate. For part (ii) use $l_{\mathcal{C}}^* = l_{\mathcal{C}}(\delta^*)$ and equation (C21) to write

$$l_{\mathcal{C}}^* = l_{\mathcal{J}}(\delta^*) + \frac{\lambda\alpha_z}{(1-\lambda)^2}\, l_{\mathcal{J}}(\delta^*)^2 \tag{F34}$$

and likewise

$$l_{\mathcal{C},nm}^* = l_{\mathcal{J}}(0) + \frac{\lambda\alpha_z}{(1-\lambda)^2}\, l_{\mathcal{J}}(0)^2 \tag{F35}$$

Differencing these expressions we can write

$$l_{\mathcal{C}}^* - l_{\mathcal{C},nm}^* = \Big(l_{\mathcal{J}}(\delta^*) - l_{\mathcal{J}}(0)\Big) \times \left[1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(l_{\mathcal{J}}(\delta^*) + l_{\mathcal{J}}(0)\Big)\right] \tag{F36}$$

Now write the term in square brackets on the LHS as $\Delta(\delta^*)$ where $\Delta(\delta)$ is the function

$$\Delta(\delta) := 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(l_{\mathcal{J}}(\delta) + l_{\mathcal{J}}(0)\Big) \tag{F37}$$

From part (i) we know $l_{\mathcal{J}}(\delta^*) > l_{\mathcal{J}}(0)$ so $l_{\mathcal{C}}^* > l_{\mathcal{C},nm}^*$ if and only if $\Delta(\delta^*) > 0$. Since $\alpha_z > 0$ and $l_{\mathcal{J}}(\delta^*) > l_{\mathcal{J}}(0) > 0$, a sufficient condition for $\Delta(\delta^*) > 0$ is that $\lambda > 0$. To prove part (ii) we need to show that any $\lambda > -1$ is also sufficient. To see this, observe that since $l_{\mathcal{J}}(\delta)$ is strictly increasing in δ, for $\lambda < 0$ we also know that $\Delta(\delta)$ is strictly decreasing in δ which in turn implies $\Delta(\delta) \geq \Delta(1)$. Hence if $\lambda < 0$ a sufficient condition for $\Delta(\delta^*) > 0$ is that $\Delta(1) > 0$. Calculating $\Delta(1)$ gives

$$\begin{aligned}\Delta(1) &= 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(l_{\mathcal{J}}(1) + l_{\mathcal{J}}(0)\Big)\\ &= 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(\frac{1-\lambda}{\alpha_z} + \frac{1-\lambda}{(1-\lambda)\alpha_x + \alpha_z}\Big)\end{aligned}$$

where the second equality follows from the expression for $l_{\mathcal{J}}(\delta)$ in equation (C17) evaluated at $\delta = 1$ and $\delta = 0$. Simplifying further

$$\Delta(1) = 1 + \frac{\lambda}{1-\lambda}\Big(1 + \frac{1}{1+\alpha}\Big) \tag{F38}$$

where $\alpha := (1-\lambda)\alpha_x/\alpha_z > 0$. So for $\lambda < 0$ a sufficient condition for $\Delta(1) > 0$ and hence $\Delta(\delta^*) > 0$ is

$$\frac{\lambda}{1-\lambda}\Big(1 + \frac{1}{1+\alpha}\Big) > -1 \tag{F39}$$

or equivalently

$$1 + \alpha > -\lambda \tag{F40}$$

Since $\alpha > 0$ a sufficient condition for this is $\lambda > -1$. To summarize, any $\lambda > -1$ is sufficient for $\Delta(\delta^*) > 0$ and hence sufficient for $l_{\mathcal{C}}^* > l_{\mathcal{C},nm}^*$. For part (iii) we then know that $\lambda < -1$ is necessary for $l_{\mathcal{C}}^* < l_{\mathcal{C},nm}^*$. Recall that $\Delta(\delta)$ is strictly decreasing in δ, i.e., $\Delta(\delta) \leq \Delta(0)$, for $\lambda < 0$. Hence for $\lambda < -1$ a sufficient condition for $\Delta(\delta^*) < 0$ is that $\Delta(0) < 0$. Calculating $\Delta(0)$ gives

$$\begin{aligned}\Delta(0) &= 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(l_{\mathcal{J}}(0) + l_{\mathcal{J}}(0)\Big)\\ &= 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(\frac{2(1-\lambda)}{(1-\lambda)\alpha_x + \alpha_z}\Big)\\ &= 1 + \frac{\lambda}{1-\lambda}\Big(\frac{2}{1+\alpha}\Big)\end{aligned}$$

So for $\lambda < -1$ a sufficient condition for $\Delta(0) < 0$ and hence $\Delta(\delta^*) < 0$ is

$$\alpha < \frac{\lambda+1}{\lambda-1} = -\left(\frac{1+\lambda}{1-\lambda}\right), \qquad \lambda < -1 \tag{F41}$$

Since $\alpha := (1-\lambda)\alpha_x/\alpha_z > 0$ we rewrite this as

$$\alpha_x < \widehat{\alpha}_x^{**} := -\left(\frac{1+\lambda}{(1-\lambda)^2}\right)\alpha_z, \qquad \lambda < -1 \tag{F42}$$

To summarize, for each $\lambda < -1$ there is a critical point $\widehat{\alpha}_x^{**}$ such that together $\lambda < -1$ and $\alpha_x < \widehat{\alpha}_x^{**}$ are sufficient for $\Delta(\delta^*) < \Delta(0) < 0$ and hence sufficient for $l_{\mathcal{C}}^* < l_{\mathcal{C},nm}^*$. □

Proof of Supplementary Proposition 4.

From equation (C24) we see that the derivative of $l_{\mathcal{C}}^*$ with respect to α_x and the derivative of $l_{\mathcal{J}}^*$ with respect to α_x have the same sign if and only if

$$1 + \frac{2\lambda\alpha_z}{(1-\lambda)^2}\, l_{\mathcal{J}}^* > 0 \tag{F43}$$

Write this key term $T(\delta^*) > 0$ where

$$T(\delta) := 1 + \frac{2\lambda\alpha_z}{(1-\lambda)^2}\, l_{\mathcal{J}}(\delta) \tag{F44}$$

Clearly $\lambda \geq 0$ suffices for $T(\delta^*) > 0$. When $\lambda < 0$, $T(\delta^*) > 0$ if and only if

$$l_{\mathcal{J}}^* < -\frac{(1-\lambda)^2}{2\lambda\alpha_z} := l_{crit} \tag{F45}$$

From Proposition 3 and Remark 1 we know that the maximum of $l_{\mathcal{J}}^*$ is $l_{max}^* = (1-\lambda)/\alpha_z$. If $l_{max}^* < l_{crit}$, i.e., if $\lambda \in (-1, 0)$, the inequality (F45) holds and therefore $T(\delta^*) > 0$. Alternatively, if $l_{max}^* > l_{crit}$, i.e., if $\lambda \in (-\infty, -1)$, there exists a subset of α_x such that the inequality (F45) does not hold and in turn $T(\delta^*) < 0$.

We now determine the set of α_x such that (F45) does not hold, conditional on $\lambda < -1$. For any $c > 1$ we know from Proposition 3 and Remark 1 that $l_{\mathcal{J}}^*$ is strictly decreasing in α_x with $\lim_{\alpha_x \to 0^+} l_{\mathcal{J}}^* = l_{max}^*$ and $\lim_{\alpha_x \to \infty} l_{\mathcal{J}}^* = 0$. Hence for each $\lambda < -1$ and $c > 1$ there is a unique critical point $\underline{\alpha}_x^{**} > 0$ such that $T(\delta^*) > 0$ if and only if $\alpha_x > \underline{\alpha}_x^{**}$. Similarly, for any $c < 1$ we again know from Proposition 3 and Remark 1 that $l_{\mathcal{J}}^*$ is strictly decreasing in α_x if and only if $\alpha_x < \alpha_x^{**}$ and $\lim_{\alpha_x \to 0^+} l_{\mathcal{J}}^* = \lim_{\alpha_x \to \infty} l_{\mathcal{J}}^* = l_{max}^*$. Let l_{min}^* denote the journalists' loss at the $\alpha_x = \alpha_x^{**}$ that achieves the minimum. For any $c < 1$ and any fixed loss $l \in (l_{min}^*, l_{max}^*)$ there are two critical points $\underline{\alpha}_x(l) < \alpha_x^{**} < \overline{\alpha}_x(l)$ such that $l_{\mathcal{J}}^* < l$ if and only if $\alpha_x \in (\underline{\alpha}_x(l), \overline{\alpha}_x(l))$. Then for each $\lambda < -1$ and $c < 1$ there are two possibilities, either $l_{crit} \in (l_{min}^*, l_{max}^*)$ or $l_{crit} \leq l_{min}^*$. For the interior cases $l_{crit} \in (l_{min}^*, l_{max}^*)$ we define the critical points by $\underline{\alpha}_x^{**} := \underline{\alpha}_x(l_{crit})$ and $\overline{\alpha}_x^{**} := \overline{\alpha}_x(l_{crit})$. For the boundary case $l_{crit} \leq l_{min}^*$ we define the critical points by $\underline{\alpha}_x^{**} = \overline{\alpha}_x^{**} = +\infty$. Given these critical points, we have $T(\delta^*) > 0$ if and only if $\alpha_x \in (\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$. □

G Knife-edge case $c = 1$

In this appendix we discuss the technicalities that arise when the costs of manipulation $c = 1$ exactly.

Preliminaries. There is no issue with $c = 1$ if the relative precision $\alpha \leq 4$. The issues with $c = 1$ arise only if $\alpha > 4$. To see this, first recall from Lemma 1 that if $\alpha > 1$ the citizens' best response $k(\delta; \alpha)$ is increasing in δ on the interval $[0, \hat{\delta}(\alpha)]$ and obtains its maximum at $\delta = \hat{\delta}(\alpha) = 1 - 1/\sqrt{\alpha} \in (0, 1)$. At the maximum, the citizens' best response takes on the value $k(\hat{\delta}(\alpha); \alpha) = \sqrt{\alpha}/2$. Hence for $\alpha > 4$ the maximum value exceeds 1. Moreover, by continuity of the best response in δ if $\alpha > 4$ there is an interval of δ such that $k(\delta; \alpha) > 1$. The boundaries of this interval $(\underline{\delta}(\alpha), \overline{\delta}(\alpha))$ are given by the roots of $k(\delta; \alpha) = 1$, which work out to be

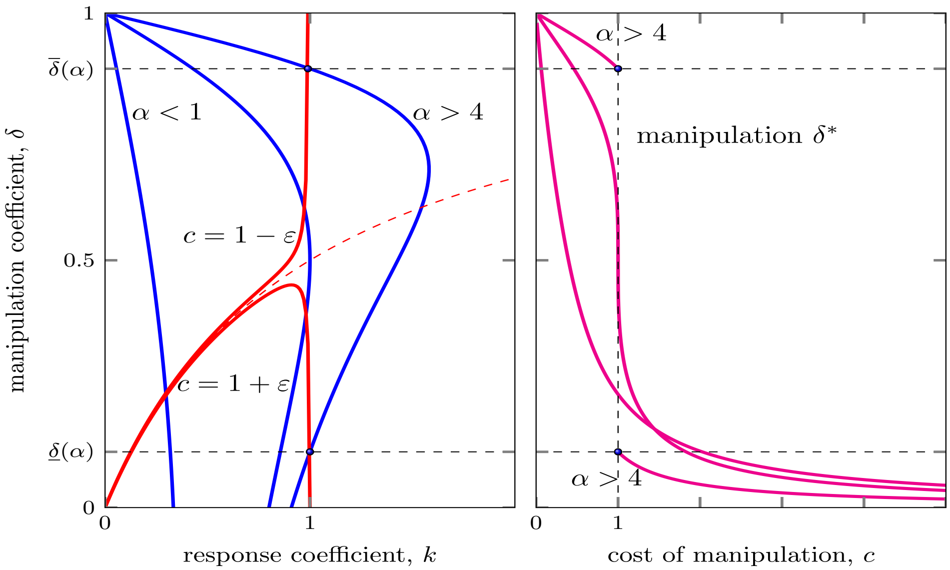
$$\underline{\delta}(\alpha)\,,\, \overline{\delta}(\alpha) = \frac{1}{2}\left(1 \pm \sqrt{1 - (4/\alpha)}\right), \qquad \alpha \geq 4 \tag{G1}$$

Observe that this interval is symmetric and centred on $1/2$ with a width of

$$\overline{\delta}(\alpha) - \underline{\delta}(\alpha) = \sqrt{1 - (4/\alpha)} \geq 0, \qquad \alpha \geq 4 \tag{G2}$$

If $\alpha = 4$, we have $\underline{\delta}(4) = \overline{\delta}(4) = 1/2$ but as α increases the width of the interval $(\underline{\delta}(\alpha), \overline{\delta}(\alpha))$ expands around $1/2$ with the boundaries $\underline{\delta}(\alpha) \to 0^+$ and $\overline{\delta}(\alpha) \to 1^-$ as $\alpha \to \infty$. Now recall from Proposition 1 that only $k \in [0, \min(c, 1)]$ and $\delta \in [0, 1]$ are candidates for an equilibrium. So if $\alpha > 4$ then none of the values of $\delta \in (\underline{\delta}(\alpha), \overline{\delta}(\alpha))$ are candidates for an equilibrium.

Costs of manipulation, $c \neq 1$. Now consider the politician's best response $\delta(k; c)$ parameterized by $c \neq 1$ and suppose $\alpha > 4$. When $c \neq 1$, the politician's objective depends on δ over the entire support $k \in [0, \min(c, 1)]$. From Proposition 1, there is a unique intersection between the politician's and the citizens' best responses. As illustrated below, if $c < 1$ the politician's best response $\delta(k; c < 1)$ lies above $\delta(k; 1) = k/(1 + k)$ and hence the equilibrium point k^*, δ^* must be on the "upper branch" of $k(\delta; \alpha)$ where $\delta^* > \overline{\delta}(\alpha)$. But for the same value of α and instead $c > 1$ the equilibrium point k^*, δ^* must be on the "lower branch" of $k(\delta; \alpha)$ where $\delta^* < \underline{\delta}(\alpha)$ because the politician's best response $\delta(k; c > 1)$ lies below $\delta(k; 1) = k/(1 + k)$.



Discontinuity at $c = 1$ and jump in the amount of manipulation δ^*

The left panel shows the citizens' best response $k(\delta; \alpha)$ for $\alpha < 1$, $\alpha = 4$ and $\alpha > 4$ (blue) and the politician's best response $\delta(k; c)$ for $c = 1 - \varepsilon$, $c = 1$, and $c = 1 + \varepsilon$ (red). For $\alpha > 4$, in the limit as $c \to 1^-$ the equilibrium is at $k^* = 1, \delta^* = \overline{\delta}(\alpha)$ but in the limit as $c \to 1^+$ the equilibrium is at $k^* = 1, \delta^* = \underline{\delta}(\alpha)$. For $\alpha > 4$ and $c = 1$ exactly both of these are equilibria because for this knife-edge special case the politician is indifferent between $\underline{\delta}(\alpha)$ and $\overline{\delta}(\alpha)$. The right panel shows the equilibrium manipulation δ^* as a function of c for $\alpha < 1$, $\alpha = 4$ and $\alpha > 4$. For $\alpha \leq 4$, the manipulation δ^* is continuous in c. But for $\alpha > 4$ the manipulation jumps discontinuously at $c = 1$. In the limit as $\alpha \to \infty$ the boundaries $\underline{\delta}(\alpha) \to 0^+$ and $\overline{\delta}(\alpha) \to 1^+$ so that the manipulation jumps by the maximum possible amount, from $\delta^* = 0$ if $c < 1$ to $\delta^* = 1$ if $c > 1$.

Summary. In brief, when $\alpha > 4$ for each $c < 1$ the equilibrium $\delta^* > \overline{\delta}(\alpha)$ with $\delta^* \to \overline{\delta}(\alpha)$ from above as $c \to 1^-$ and for each $c > 1$ the equilibrium $\delta^* < \underline{\delta}(\alpha)$ with $\delta^* \to \overline{\delta}(\alpha)$ from below as $c \to 1^+$.

Knife-edge case. Now consider the case $c = 1$ exactly. The key part of the politician's objective becomes

$$B(\delta, k) - C(\delta) = (k^2 - 1)\delta^2 + 2k(1 - k)\delta + (1 - k)^2 \tag{G3}$$

When $k \neq 1$, the politician's best response is $\delta(k; 1) = k/(1 + k)$, which is increasing in k and approaches $1/2$ as $k \to 1$. But when $k = 1$, the politician's objective is independent of δ and in turn the politician is indifferent in the choice of δ. The equilibrium $(k^* = 1, \delta^*)$ is thus entirely determined by the citizens' best response. If $\alpha < 4$, the citizens' best response $k(\delta; \alpha) < 1$ so that $k^* = 1$ is never an equilibrium. If $\alpha = 4$, there is a unique equilibrium determined by the maximum of the citizens' best response $(k^* = 1, \delta^* = 1/2)$. If $\alpha > 4$, there are two equilibria corresponding to the two roots of $k(\delta; \alpha) = 1$: namely $(k^* = 1, \delta^* = \underline{\delta}(\alpha))$ and $(k^* = 1, \delta^* = \overline{\delta}(\alpha))$.

Further intuition for large changes in manipulation near $c = 1$. Now consider the sensitivity of the equilibrium amount of manipulation to changes in c near $c = 1$. Recall that, taking the citizens' k as given, the politician chooses manipulation δ to maximize

$$V(\delta, k) = \frac{1}{\alpha_z}\left(B(\delta, k) - C(\delta)\right) + \frac{1}{\alpha_x}k^2 \tag{G4}$$

with benefit $B(\delta, k) = (k\delta + 1 - k)^2$ and costs of manipulation $C(\delta) = c\delta^2$.

Now consider an environment where the citizens are inclined to be very responsive to their signals, $\alpha \to \infty$ so that $k \to \min(c, 1)$. First, suppose that $c > 1$ so that $k \to 1$. Then the relevant part of the politician's objective simplifies to

$$B(\delta, 1) - C(\delta) = (1 - c)\delta^2 \tag{G5}$$

so that for any $c > 1$ the politician will choose $\delta = 0$. Next, suppose instead that $c < 1$ so that $k \to c$. In this case the relevant part of the politician's objective simplifies to

$$B(\delta, c) - C(\delta) = -c(1 - c)\delta^2 + 2c(1 - c)\delta + (1 - c)^2 \tag{G6}$$

so that for any $c < 1$ the politician will choose $\delta = 1$. In short, as $\alpha \to \infty$, the politician's manipulation is a *step function* in c, with $\delta = 1$ for all $c < 1$ and $\delta = 0$ for all $c > 1$.

What is the meaning of $c = 1$**?** So given that the amount of manipulation can be extremely sensitive to c near $c = 1$, what does $c = 1$ mean? Recall that in the politician's objective (5) the gross benefit $\int_0^1 (a_i - \theta)^2\, di$ has a coefficient normalized to 1. If instead we had written the objective with $b \int_0^1 (a_i - \theta)^2\, di$ for some $b > 0$ then throughout the analysis the relevant parameter would be the cost/benefit ratio c/b and the critical point would be where the cost/benefit ratio is $c/b = 1$. In this parameterization, the politician's equilibrium manipulation is extremely sensitive to changes in either c or b in the vicinity of $c/b = 1$. With α high and costs and benefits *evenly poised*, a small decrease in b or small increase in c would lead to a large reduction in manipulation.

H Coefficients sum to one

In this appendix we show that writing the citizens' linear strategy as $a_i = kx_i + (1 - k)z$ is without loss of generality. Suppose that the citizens' linear strategy is

$$a_i = \beta_0 + \beta_1 x_i + \beta_2 z$$

for some coefficients $\beta_0, \beta_1, \beta_2$. We will show that in any linear equilibrium $\beta_0 = 0$ and $\beta_1 + \beta_2 = 1$.

The politician's problem is then to choose y to maximize

$$\begin{aligned} V(y) &= \int_0^1 \left(\beta_0 + \beta_1(y + \varepsilon_i) + \beta_2 z - \theta\right)^2 \, di - c(y - \theta)^2 \\ &= (\beta_0 + \beta_1 y + \beta_2 z - \theta)^2 + \frac{1}{\alpha_x}\beta_1^2 - c(y - \theta)^2 \end{aligned}$$

The solution to this problem is

$$y = \gamma_0 + \gamma_1\theta + \gamma_2 z$$

where

$$\gamma_0 = \frac{\beta_0\beta_1}{c - \beta_1^2} \tag{H1}$$

$$\gamma_1 = \frac{c - \beta_1}{c - \beta_1^2} \tag{H2}$$

$$\gamma_2 = \frac{\beta_1\beta_2}{c - \beta_1^2} \tag{H3}$$

But if the politician has the strategy $y = \gamma_0 + \gamma_1\theta + \gamma_2 z$, the citizens' posterior expectation of θ, and hence their action a_i, is given by

$$\begin{aligned} a_i = \mathbb{E}[\theta \,|\, x_i] &= \frac{\gamma_1\alpha_x}{\gamma_1^2\alpha_x + \alpha_z}\left(\frac{1}{\gamma_1}(x_i - \gamma_2 z) - \frac{\gamma_0}{\gamma_1}\right) + \frac{\alpha_z}{\gamma_1^2\alpha_x + \alpha_z}z \\ &= \frac{\gamma_1\alpha_x}{\gamma_1^2\alpha_x + \alpha_z}x_i + \frac{\alpha_z - \gamma_1\alpha_x\gamma_2}{\gamma_1^2\alpha_x + \alpha_z}z - \frac{\gamma_1\alpha_x}{\gamma_1^2\alpha_x + \alpha_z}\gamma_0 \end{aligned}$$

Matching coefficients with $a_i = \beta_0 + \beta_1 x_i + \beta_2 z$ we then have

$$\beta_0 = -\frac{\gamma_1\alpha_x}{\gamma_1^2\alpha_x + \alpha_z}\gamma_0 \tag{H4}$$

$$\beta_1 = \frac{\gamma_1\alpha_x}{\gamma_1^2\alpha_x + \alpha_z} \tag{H5}$$

$$\beta_2 = \frac{\alpha_z - \gamma_1\alpha_x\gamma_2}{\gamma_1^2\alpha_x + \alpha_z} \tag{H6}$$

First observe that equations (H1) and (H4) together imply that the intercepts are $\beta_0 = \gamma_0 = 0$. Then observe from (H2)-(H3) and (H5)-(H6) that $\gamma_1 + \gamma_2 = 1$ implies $\beta_1 + \beta_2 = 1$ and vice-versa. So indeed the citizens' strategy takes the form $a_i = kx_i + (1-k)z$ where $k = \beta_1$ and the politician's strategy takes the form $y = (1-\delta)\theta + \delta z$ where $\delta = \gamma_2$. Hence from (H3) and (H5) we can write

$$\delta = \frac{k - k^2}{c - k^2}, \qquad k = \frac{(1-\delta)\alpha}{(1-\delta)^2\alpha + 1}$$

where $\alpha := \alpha_x/\alpha_z$. These are the same as the best response formulas equations (**??**) and (**??**) in the main text and from Proposition 1 we know that there is a unique pair k^*, δ^* satisfying these conditions.