Media Censorship Backfire—Online Appendices

Appendix A Model setup

A.1 Citizen belief updating

A.1.1 Posterior beliefs of the citizens

Since the signals perfectly match the states, $Pr(s = \beta|\omega = B) = Pr(s = \phi|\omega = N) = 1$ and $Pr(s = \beta|\omega = N) = Pr(s = \phi|\omega = B) = 0$. Applying the Bayes' rule, citizen i's posterior belief of the bad state upon encountering bad news is calculated as follows:

$$\begin{aligned}
&\rho_\beta = Pr(\omega = B|r = \beta) \\
&= \tfrac{Pr(r=\beta|\omega=B)}{Pr(r=\beta)} \\
&= \tfrac{Pr(\omega=B)[Pr(s=\beta|\omega=B)Pr(r=\beta|s=\beta)+Pr(s=\phi|\omega=B)Pr(r=\beta|s=\phi)]}{Pr(\omega=B)[Pr(s=\beta|\omega=B)Pr(r=\beta|s=\beta)+Pr(s=\phi|\omega=B)Pr(r=\beta|s=\phi)]+Pr(\omega=N)[Pr(s=\beta|\omega=N)Pr(r=\beta|s=\beta)+Pr(s=\phi|\omega=N)Pr(r=\beta|s=\phi)]} \\
&= \tfrac{\rho_o[1*\sigma_\beta+0*\sigma_\phi]}{\rho_o\sigma_\beta+(1-\rho_o)[0*\sigma_\beta+1*\sigma_\phi]} \\
&= \tfrac{\rho_o\sigma_\beta}{\rho_o\sigma_\beta+(1-\rho_o)\sigma_\phi}.
\end{aligned}$$

And citizen i's posterior belief of the bad state after receiving good news is given by

$$\begin{aligned}
&\rho_\phi = Pr(\omega = B|r = \phi) \\
&= \tfrac{Pr(r=\phi|\omega=B)}{Pr(r=\phi)} \\
&= \tfrac{Pr(\omega=B)[Pr(s=\beta|\omega=B)Pr(r=\phi|s=\beta)+Pr(s=\phi|\omega=B)Pr(r=\phi|s=\phi)]}{Pr(\omega=B)[Pr(s=\beta|\omega=B)Pr(r=\phi|s=\beta)+Pr(s=\phi|\omega=B)Pr(r=\phi|s=\phi)]+Pr(\omega=N)[Pr(s=\beta|\omega=N)Pr(r=\phi|s=\beta)+Pr(s=\phi|\omega=N)Pr(r=\phi|s=\phi)]} \\
&= \tfrac{\rho_o[1*(1-\sigma_\beta)+0*(1-\sigma_\phi)]}{\rho_o[1*(1-\sigma_\beta)+0*(1-\sigma_\phi)]+(1-\rho_f o)[0*(1-\sigma_\beta)+1*(1-\sigma_\phi)]} \\
&= \tfrac{\rho_o(1-\sigma_\beta)}{\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)}.
\end{aligned}$$

A.1.2 News informativeness

The news is informative if, compared with the prior, the citizen's posterior belief of the bad state increases for bad news, i.e., $\rho_\beta > \rho_o$, and decreases for good news, i.e., $\rho_\phi < \rho_o$.

$$\rho_\beta - \rho_o = \frac{\rho_o\sigma_\beta}{\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi} - \rho_o = \frac{\rho_o\sigma_\beta - \rho_o[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]}{\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi} = \frac{\rho_o(1-\rho_o)(\sigma_\beta - \sigma_\phi)}{\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi}. \tag{A. 1}$$

Given the assumption $\sigma_\beta > \sigma_\phi$, it follows that $\rho_\beta > \rho_o$.

$$\begin{aligned}\rho_\phi - \rho_o &= \frac{\rho_o(1-\sigma_\beta)}{\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)} - \rho_o \\ &= \frac{\rho_o(1-\sigma_\beta) - \rho_o[\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)]}{\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)} \\ &= \frac{\rho_o(1-\rho_o)(\sigma_\phi-\sigma_\beta)}{\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)}. \end{aligned} \tag{A. 2}$$

Given the assumption $\sigma_\beta > \sigma_\phi$, it follows that $\rho_\phi < \rho_o$.

News informativeness decreases in media bias in either direction if $\frac{\partial \rho_\beta}{\partial \sigma_\beta} > 0$, $\frac{\partial \rho_\phi}{\partial \sigma_\beta} < 0$, $\frac{\partial \rho_\beta}{\partial \sigma_\phi} < 0$, and $\frac{\partial \rho_\phi}{\partial \sigma_\phi} > 0$.

$$\frac{\partial \rho_\beta}{\partial \sigma_\beta} = \frac{\rho_o - \rho_o^2\sigma_\beta}{[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2} = \frac{\rho_o(1-\rho_o\sigma_\beta)}{[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2} > 0, \tag{A. 3}$$

$$\begin{aligned}\frac{\partial \rho_\phi}{\partial \sigma_\beta} &= \frac{-\rho_o[\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)] - \rho_o(1-\sigma_\beta)(-\rho_o)}{[\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)]^2} \\ &= \frac{-\rho_o(1-\rho_o)(1-\sigma_\phi)}{[\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)]^2} < 0, \end{aligned} \tag{A. 4}$$

$$\frac{\partial \rho_\beta}{\partial \sigma_\phi} = \frac{0 - [(1-\rho_o)\rho_o\sigma_\beta]}{[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2} = \frac{-\rho_o(1-\rho_o)\sigma_\beta}{[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2} < 0, \tag{A. 5}$$

$$\frac{\partial \rho_\phi}{\partial \sigma_\phi} = \frac{0 - [-(1-\rho_o)\rho_o(1-\sigma_\beta)]}{[\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)]^2} = \frac{\rho_o(1-\rho_o)(1-\sigma_\beta)}{[\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)]^2} > 0. \tag{A. 6}$$

A.1.3 Definitions of the indifferent citizens

A.1.3.1 Indifferent citizens of high concern

For a citizen of high concern, denoted as $\alpha \in [\alpha^{oo}, \alpha^H]$, the expected gain from the news

is

$$S_H(\alpha;\sigma) = Pr(r=\beta)EU(0;\rho_\beta,\alpha) + Pr(r=\phi)EU(1;\rho_\phi,\alpha) - EU(1;\rho_o,\alpha). \quad \text{(A. 7)}$$

The citizen is indifferent between subscribing or not if $S_H(\alpha;\sigma) = p$, which is equivalent to

$$[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi](-\alpha\eta) + [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi][-\frac{b\rho_o(1-\sigma_\beta)}{\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)}] + b\rho_o = p. \quad \text{(A. 8)}$$

Simplify the equation and obtain

$$(-\alpha\eta)[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi] - b\rho_o(1-\sigma_\beta) + b\rho_o = p. \quad \text{(A. 9)}$$

Solve for the α in equation (A. 9) and define it as the indifferent citizen with high concern,

$$\alpha_p^H \equiv max\{\alpha^{oo}, \frac{b\rho_o\sigma_\beta - p}{\eta[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]}\}. \quad \text{(A. 10)}$$

Assume that $p < b\rho_o\sigma_\beta$, so that $\alpha_p^H > 0$.

Regarding how the indifferent citizen's position changes with media bias towards good news, there is

$$\frac{\partial\alpha_p^H}{\partial\sigma_\beta} = \frac{b\rho_o[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi] - \rho_o(b\rho_o\sigma_\beta - p)}{\eta[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2} = \frac{\rho_o[b(1-\rho_o)\sigma_\phi - p]}{\eta[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2}. \quad \text{(A. 11)}$$

Assume that $p < b(1-\rho_o)\sigma_\phi$, and then $\frac{\partial\alpha^H}{\partial\sigma_\beta} > 0$.

As to how the indifferent citizen's position changes with media bias towards bad news,

there is

$$\frac{\partial \alpha_p^H}{\partial \sigma_\phi} = \frac{-(1-\rho_o)(b\rho_o\sigma_\beta - p)}{\eta[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2} < 0. \tag{A. 12}$$

Overall, if p is not excessively high, the cut point α_p^H decreases in media bias towards either good news or bad news.

A.1.3.2 Indifferent citizen with low concern

For a citizen of low concern, denoted as $\alpha \in [\alpha^L, \alpha^{oo})$, the expected gain from news is

$$S_L(\alpha;\sigma) = Pr(r=\beta)EU(0;\rho_\beta,\alpha) + Pr(r=\phi)EU(1;\rho_\phi,\alpha) - EU(0;\rho_o,\alpha). \tag{A. 13}$$

The citizen is indifferent between subscribing or not if $S_H(\alpha;\sigma) = p$, which equals

$$[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi](-\alpha\eta) + [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi][-\frac{b\rho_o(1-\sigma_\beta)}{\rho_o(1-\sigma_\beta)+(1-\rho_o)(1-\sigma_\phi)}] + \alpha\eta = p. \tag{A. 14}$$

Simplify the equation and obtain

$$\alpha\eta[1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi] - b\rho_o(1-\sigma_\beta) = p. \tag{A. 15}$$

Solve for the α in (A. 15) and define it as the indifferent citizen with low concern,

$$\alpha_p^L \equiv min\{\frac{p + b\rho_o(1-\sigma_\beta)}{\eta[1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]}, \alpha^{oo}\}. \tag{A. 16}$$

Regarding how the indifferent citizen's position changes with media bias towards good news, there is

$$\frac{\partial \alpha_p^L}{\partial \sigma_\beta} = \frac{-b\rho_o[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]+\rho_o[p+b\rho_o(1-\sigma_\beta)])}{\eta[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]^2} = \frac{\rho_o[p-b(1-\rho_o)(1-\sigma_\phi)]}{\eta[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]^2}. \tag{A. 17}$$

Assume that $p < b(1-\rho_o)(1-\sigma_\phi)$, and then $\frac{\partial \alpha_p^L}{\partial \sigma_\beta} < 0$.

Concerning how the indifferent citizen's position changes with media bias towards bad news, there is

$$\frac{\partial \alpha_p^L}{\partial \sigma_\phi} = \frac{(1-\rho_o)[p+b\rho_o(1-\sigma_\beta)]}{\eta[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]^2} > 0. \tag{A. 18}$$

Overall, if p is not excessively high, the cut point α_p^L increases in media bias towards either good news or bad news.

Appendix B Media bias manipulation

B.1 Media bias manipulation backfire I

When the outlet is market-driven, manipulating media bias backfires if the government's benefits from citizen compliance increase in the threshold $\bar{\sigma}_\beta$, i.e., $\frac{\partial h(\cdot)}{\partial \bar{\sigma}_\beta} > 0$. With a slight abuse of the notation, $\frac{\partial h(\cdot)}{\partial \bar{\sigma}_\beta} > 0$ is equal to $\frac{\partial h(\cdot)}{\partial \sigma_\beta} > 0$, since in equilibrium the market-driven

outlet exhibits a bias towards good news $\sigma_\beta^* = \bar{\sigma}_\beta$ (Lemma **??**).

$$\begin{aligned}\frac{\partial h(\cdot)}{\partial \bar{\sigma}_\beta} = \frac{\partial h(\cdot)}{\partial \sigma_\beta} &= [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]'[1-F(\alpha_p^H)] + [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi][1-F(\alpha_p^H)]' \\ &\quad + [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]'[1-F(\alpha_p^L)] + [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi][1-F(\alpha_p^L)]' \\ &= -\rho_o[F(\alpha_p^H) - F(\alpha_p^L)] - [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]f(\alpha_p^H)\frac{\partial \alpha_p^H}{\partial \sigma_\beta} \\ &\quad - [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]f(\alpha_p^L)\frac{\partial \alpha_p^L}{\partial \sigma_\beta}. \end{aligned} \tag{A. 19}$$

When $\frac{\partial h(\cdot)}{\partial \sigma_\beta} > 0$, there is

$$-[1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]f(\alpha_p^L)\frac{\partial \alpha_p^L}{\partial \sigma_\beta} > \rho_o[F(\alpha_p^H) - F(\alpha_p^L)] + [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]f(\alpha_p^H)\frac{\partial \alpha_p^H}{\partial \sigma_\beta}, \tag{A. 20}$$

which is equivalent to

$$-Pr(r=\phi)f(\alpha_p^L)\frac{\partial \alpha_p^L}{\partial \sigma_\beta} > \rho_o[F(\alpha_p^H) - F(\alpha_p^L)] + Pr(r=\beta)f(\alpha_p^H)\frac{\partial \alpha_p^H}{\partial \sigma_\beta}. \tag{A. 21}$$

The optimal level of censorship for the government is achieved when

$$\frac{\partial \pi_G}{\partial \sigma_\beta} = \psi\frac{\partial h(\cdot)}{\partial \sigma_\beta} - \frac{\partial z_1}{\partial \sigma_\beta} = 0, \tag{A. 22}$$

which is equivalent to

$$\psi\{-\rho_o[F(\alpha_p^H)-F(\alpha_p^L)]-\frac{\rho_o\sigma_\beta f(\alpha_p^H)(-\rho_o p)}{\eta(\rho_o\sigma_\beta)^2}-\frac{(1-\rho_o\sigma_\beta)f(\alpha_p^L)\rho_o[p-b(1-\rho_o)]}{\eta(1-\rho_o\sigma_\beta)^2}\}-\frac{\partial z_1}{\partial\sigma_\beta}=0$$
$$\Rightarrow \sigma_\beta(1-\rho_o\sigma_\beta)\{\frac{\frac{\partial z_1}{\partial\sigma_\beta}}{\psi}+\rho_o[F(\alpha_p^H)-F(\alpha_p^L)]\}-\frac{f(\alpha_p^H)p(1-\rho_o\sigma_\beta)}{\eta}+\frac{f(\alpha_p^L)[p-b(1-\rho_o)]\sigma_\beta}{\eta}=0$$
$$\Rightarrow \rho_o\{\frac{\frac{\partial z_1}{\partial\sigma_\beta}}{\psi}+\rho_o[F(\alpha_p^H)-F(\alpha_p^L)]\}\sigma_\beta^2-\{\{\frac{\frac{\partial z_1}{\partial\sigma_\beta}}{\psi}+\rho_o[F(\alpha_p^H)-F(\alpha_p^L)]\}$$
$$-\frac{f(\alpha_p^H)p\rho_o}{\eta}-\frac{f(\alpha_p^L)[p-b(1-\rho_o)]}{\eta}\}\sigma_\beta+\frac{f(\alpha_p^H)p}{\eta}=0. \quad \text{(A. 23)}$$

This is a quadratic equation in terms of σ_β, and the solution is

$$\bar{\sigma}_\beta^*=\sigma_\beta^*=min\{max\{max\{\frac{-B-\sqrt{B^2-4AC}}{2A},0\},max\{\frac{-B+\sqrt{B^2+4AC}}{2A},0\}\},1\}, \quad \text{(A. 24)}$$

where $A=\rho_o\{\frac{\frac{\partial z_1}{\partial\sigma_\beta}}{\psi}+\rho_o[F(\alpha_p^H)-F(\alpha_p^L)]\}$, $B=\{\frac{\frac{\partial z_1}{\partial\sigma_\beta}}{\psi}+\rho_o[F(\alpha_p^H)-F(\alpha_p^L)]\}-\frac{f(\alpha_p^H)p\rho_o}{\eta}-\frac{f(\alpha_p^L)[p-b(1-\rho_o)]}{\eta}$, and $C=\frac{f(\alpha_p^H)p}{\eta}$.

B.2 Liberal media outlet under media bias manipulation

Without censorship, it is evident that the liberal outlet in unbiased toward good news, i.e., $\hat{\sigma}_\beta=1$. The optimal level of bias towards bad news for the outlet, denoted as $\hat{\sigma}_\phi$, is achieved when

$$\frac{\partial\pi_{M_L}}{\partial\sigma_\phi}=v[\frac{\partial F(\alpha_p^H)}{\partial\sigma_\phi}-\frac{\partial F(\alpha_p^L)}{\partial\sigma_\phi}]+\gamma \quad \text{(A. 25)}$$
$$=v\{\frac{F(\alpha_p^H)[[-(b\rho_o\sigma_\beta-p)(1-\rho_o)]}{\eta[\rho_o\sigma_\beta+(1-\rho_o)\sigma_\phi]^2}-\frac{F(\alpha_p^L)[p+b\rho_o(1-\sigma_\beta)](1-\rho_o)}{\eta[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]^2}]+\gamma=0.$$

Equation (A. 25) can be transformed into a quadratic equation involving σ_ϕ. While this equation may have multiple solutions for σ_ϕ, extracting meaningful implications from

them can be overly complicated. The focus of interest here is how the optimal value $\hat{\sigma}_\phi$ changes with σ_β. Applying the Implicit Function Theorem,

$$\frac{\partial \hat{\sigma}_\phi}{\partial \sigma_\beta} = \frac{f(\alpha_p^H)\{b\rho_o[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2 - 2(b\rho_o\sigma_\beta - p)[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi][\rho_o + (1-\rho_o)\frac{\partial \sigma_\phi}{\partial \sigma_\beta}]\}}{[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^4} + \frac{f(\alpha_p^L)\{-b\rho_o[1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]^2 - 2[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi][1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi][-\rho_o - (1-\rho_o)\frac{\partial \sigma_\phi}{\partial \sigma_\beta}]\}}{[1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]^4}$$

$$= 0, \tag{A. 26}$$

and it is equal to

$$\frac{\partial \hat{\sigma}_\phi}{\partial \sigma_\beta} = \frac{b\rho_o[f(\alpha_p^H)Pr(r=\phi)^2 - f(\alpha_p^L)Pr(r=\beta)^2]}{2\eta(1-\rho_o)[\alpha_p^H f(\alpha_p^H)Pr(r=\phi)^2 - \alpha_p^L f(\alpha_p^L)Pr(r=\beta)^2]}. \tag{A. 27}$$

The sign of $\frac{\partial \hat{\sigma}_\phi}{\partial \sigma_\beta}$ can be positive, negative, or zero.

B.3 Media bias manipulation backfire II

The government has four potential strategies to manipulate the liberal outlet's bias, which will be discussed in order below, as needed.

Case 1: Not manipulating media bias, i.e., $\bar{\sigma}_\beta = 1$ and $\bar{\sigma}_\phi = 1$.

The government does not implement censorship when the expected payoff from citizen compliance, denoted as $h(\cdot)$, rises with media bias towards bad news, and declines with

media bias towards good news.

$$\begin{aligned}\frac{\partial h(\cdot)}{\partial \sigma_\phi} &= [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]'[1-F(\alpha_p^H)] + [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi][1-F(\alpha_p^H)]' \\ &\quad + [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]'[1-F(\alpha_p^L)] + [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi][1-F(\alpha_p^L)]' \\ &= -(1-\rho_o)[F(\alpha_p^H) - F(\alpha_p^L)] - [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]f(\alpha_p^H)\frac{\partial \alpha_p^H}{\partial \sigma_\phi} \\ &\quad - [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]f(\alpha_p^L)\frac{\partial \alpha_p^L}{\partial \sigma_\phi}. \end{aligned} \tag{A. 28}$$

Compelling the liberal outlet to be less biased towards bad news backfires, when $\frac{\partial h(\cdot)}{\partial \sigma_\phi} > 0$, i.e.,

$$-[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]f(\alpha_p^L)\frac{\partial \alpha_p^L}{\partial \sigma_\phi} > (1-\rho_o)[F(\alpha_p^H) - F(\alpha_p^L)] + [1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]f(\alpha_p^H)\frac{\partial \alpha_p^H}{\partial \sigma_\phi}, \tag{A. 29}$$

which is equal to

$$-Pr(r=\beta)f(\alpha_p^H)\frac{\partial \alpha_p^H}{\partial \sigma_\phi} > (1-\rho_o)[F(\alpha_p^H) - F(\alpha_p^L)] + Pr(r=\phi)f(\alpha_p^L)\frac{\partial \alpha_p^L}{\partial \sigma_\phi}. \tag{A. 30}$$

Condition (A. 30) encapsulates the phenomenon of media bias manipulation backfire II.

If the government imposes a threshold on media bias towards good news, the liberal outlet has flexibility to adjust its bias towards bad news in this case. As the government

does not manipulate σ_β, there must be

$$\begin{aligned}\frac{\partial h(\sigma_\beta,\sigma_\phi(\sigma_\beta))}{\partial\sigma_\beta} &= [\rho_o\sigma_\beta+(1-\rho_o)\sigma_\phi]'[1-F(\alpha_p^H)]+[\rho_o\sigma_\beta+(1-\rho_o)\sigma_\phi][1-F(\alpha_p^H)]' \\ &\quad +[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]'[1-F(\alpha_p^L)]+[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi][1-F(\alpha_p^L)]' \\ &= -[\rho_o+(1-\rho_o)\frac{\partial\sigma_\phi}{\partial\sigma_\beta}][F(\alpha_p^H)-F(\alpha_p^L)]-[\rho_o\sigma_\beta+(1-\rho_o)\sigma_\phi]f(\alpha_p^H)\frac{\partial\alpha_p^H(\sigma_\beta,\sigma_\phi(\sigma_\beta))}{\partial\sigma_\beta} \\ &\quad -[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]f(\alpha_p^L)\frac{\partial\alpha_p^L(\sigma_\beta,\sigma_\phi(\sigma_\beta))}{\partial\sigma_\beta}>0, \qquad \text{(A. 31)}\end{aligned}$$

where $\frac{\partial\sigma_\phi}{\partial\sigma_\beta}$ is given by (A. 27), and

$$\begin{aligned}\frac{\partial\alpha_p^H(\sigma_\beta,\sigma_\phi(\sigma_\beta))}{\partial\sigma_\beta} &= \frac{b\rho_o[\rho_o\sigma_\beta+(1-\rho_o)\sigma_\phi]-(b\rho_o\sigma_\beta-p)[\rho_{o+}(1-\rho_o)\frac{\partial\sigma_\phi}{\partial\sigma_\beta}]}{\eta[\rho_o\sigma_\beta+(1-\rho_o)\sigma_\phi]^2} \\ &= \frac{\partial\alpha_p^H}{\partial\sigma_\beta}-\frac{(b\rho_o\sigma_\beta-p)(1-\rho_o)\frac{\partial\sigma_\phi}{\partial\sigma_\beta}}{\eta[\rho_o\sigma_\beta+(1-\rho_o)\sigma_\phi]^2}, \qquad \text{(A. 32)}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\alpha_p^L(\sigma_\beta,\sigma_\phi(\sigma_\beta))}{\partial\sigma_\beta} &= \frac{-b\rho_o[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]+\rho_o[p+b\rho_o(1-\sigma_\beta)]+[p+b\rho_o(1-\sigma_\beta)](1-\rho_o)\frac{\partial\sigma_\phi}{\partial\sigma_\beta}}{\eta[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]^2} \\ &= \frac{\partial\alpha_p^L}{\partial\sigma_\beta}+\frac{[p+b\rho_o(1-\sigma_\beta)](1-\rho_o)\frac{\partial\sigma_\phi}{\partial\sigma_\beta}}{\eta[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]^2}. \qquad \text{(A. 33)}\end{aligned}$$

Comparing the differences between equation (A. 19) and equation (A. 31), it is obvious that

$$\frac{\partial h(\sigma_\beta,\sigma_\phi(\sigma_\beta))}{\partial\sigma_\beta} = \overbrace{\frac{\partial h(\sigma_\beta)}{\partial\sigma_\beta}}^{\text{effect of change in }\sigma_\beta} + \overbrace{\frac{\partial h(\sigma_\phi(\sigma_\beta))}{\partial\sigma_\beta}}^{\text{effect of change in }\sigma_\phi\text{ due to change in }\sigma_\beta}, \qquad \text{(A. 34)}$$

where $\frac{\partial h(\sigma_\beta)}{\partial \sigma_\beta}$ is given by (A. 19) and

$$\begin{aligned}\frac{\partial h(\sigma_\phi(\sigma_\beta))}{\partial \sigma_\beta} = & -(1-\rho_o)\frac{\partial \sigma_\phi}{\partial \sigma_\beta}[F(\alpha_p^H)-F(\alpha_p^L)] - f(\alpha_p^H)\frac{-(b\rho_o\sigma_\beta - p)(1-\rho_o)\frac{\partial \sigma_\phi}{\partial \sigma_\beta}}{\eta[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]} \\ & - f(\alpha_p^L)\frac{[p+b\rho_o(1-\sigma_\beta)](1-\rho_o)\frac{\partial \sigma_\phi}{\partial \sigma_\beta}}{\eta[1-\rho_o\sigma_\beta-(1-\rho_o)\sigma_\phi]} \\ = & -(1-\rho_o)\frac{\partial \sigma_\phi}{\partial \sigma_\beta}\{[F(\alpha_p^H)-F(\alpha_p^L)] - f(\alpha_p^H)\alpha_p^H - f(\alpha_p^L)\alpha_p^L\}. \end{aligned} \tag{A. 35}$$

It is evident that the government's optimal levels of media bias manipulation are $\bar{\sigma}_\beta^* = 1$ and $\bar{\sigma}_\phi^* = 1$ in *Case 1*.

Case 2: Solely manipulating media bias towards good news, i.e., $\bar{\sigma}_\beta \in [0,1)$ and $\bar{\sigma}_\phi = 1$.

When the government only manipulates the liberal outlet's bias towards good news, the threshold $\bar{\sigma}_\phi^* = 1$, and the threshold $\bar{\sigma}_\beta^*$ solves

$$\frac{\partial \pi_G}{\partial \sigma_\beta} = \psi\frac{\partial h(\sigma_\beta, \sigma_\phi(\sigma_\beta))}{\partial \sigma_\beta} - \frac{\partial z_1}{\partial \sigma_\beta} = 0. \tag{A. 36}$$

Equation (A. 36) can be transformed into a sextic equation involving σ_β. While σ_β^* does exist, an analytic solution is not readily available.

Case 3: Solely manipulating media bias towards bad news, i.e., $\bar{\sigma}_\beta = 1$ and $\bar{\sigma}_\phi \in [0, \hat{\sigma}_\phi)$.

The government does not manipulate the liberal outlet's bias towards good news, which means $\bar{\sigma}_\beta^* = 1$. Regarding media bias towards bad news, if $\sigma_\phi^* = \bar{\sigma}_\phi$, then $\bar{\sigma}_\phi^*$ solves

$$\frac{\partial \pi_G}{\partial \sigma_\phi} = \psi\frac{\partial h(\cdot)}{\partial \sigma_\phi} - \frac{\partial z_2}{\partial \sigma_\phi} = 0. \tag{A. 37}$$

Equation (A. 37) can be transformed into a quadratic equation in terms of σ_ϕ, and the

solution is

$$\bar{\sigma}_\phi^* = \sigma_\phi^* = min\{max\{max\{\frac{-B-\sqrt{B^2-4AC}}{2A}, 0\}, max\{\frac{-B+\sqrt{B^2+4AC}}{2A}, 0\}\}, 1\}, \tag{A. 38}$$

where

$A = (1-\rho_o)^2\{\frac{\frac{\partial z_2}{\partial \sigma_\phi}}{\psi} + [F(\alpha_p^H) - F(\alpha_p^L)]\}$,

$B = (1-\rho_o)(2\eta\rho_o\sigma_\beta + 1)\{\frac{\frac{\partial z_2}{\partial \sigma_\phi}}{\psi} + [F(\alpha_p^H) - F(\alpha_p^L)]\} + \frac{(1-\rho_o)^2}{\eta}\{f(\alpha_p^H)(b\rho_o\sigma_\beta - p) - f(\alpha_p^L)[p + b\rho_o(1-\sigma_\beta)]\}$,

$C = \rho_o\sigma_\beta(\rho_o\sigma_\beta + 1)\{\frac{\frac{\partial z_2}{\partial \sigma_\phi}}{\psi} + [F(\alpha_p^H) - F(\alpha_p^L)]\} - \frac{(1-\rho_o)}{\eta}\{f(\alpha_p^H)(1-\rho_o\sigma_\beta)(b\rho_o\sigma_\beta - p) - f(\alpha_p^L)\rho_o\sigma_\beta[p + b\rho_o(1-\sigma_\beta)]\}$.

If $\sigma_\phi^* < \bar{\sigma}_\phi$, then the optimal threshold $\bar{\sigma}_\phi^*$ depends on how the solutions(s) of the first order condition $\frac{\partial \pi_{M_L}}{\partial \sigma_\phi} = 0$ look like.

Appendix C News cost manipulation

C.1 News cost manipulation backfire I

When the outlet is market-driven, manipulating the citizens' cost of accessing news backfires if

$$\begin{aligned}\frac{\partial h(\cdot)}{\partial \theta} &= Pr(r=\beta)[-f(\alpha_\theta^H)\frac{\partial \alpha_\theta^H}{\partial \theta}] + Pr(r=\phi)[-f(\alpha_\theta^L)\frac{\partial \alpha_\theta^L}{\partial \theta}] \\ &= Pr(r=\beta)[\frac{f(\alpha_\theta^H)}{\eta Pr(r=\beta)}] + Pr(r=\phi)[\frac{-f(\alpha_\theta^L)}{\eta Pr(r=\phi)}] \\ &= \frac{f(\alpha_\theta^H) - f(\alpha_\theta^L)}{\eta} < 0, \end{aligned} \tag{A. 39}$$

which is equivalent to $f(\alpha_\theta^H) < f(\alpha_\theta^L)$.

C.2 News cost manipulation backfire II

Without censorship, the liberal outlet is unbiased towards good news, i.e., $\hat{\sigma}_\beta = 1$. However, it can be biased towards bad news when its ideological intensity γ is sufficiently large. The liberal outlet's optimal level of bias towards bad news satisfies the first order condition,

$$\begin{aligned}\frac{\partial \pi_{M_L}}{\partial \sigma_\phi} &= v[\frac{\partial F(\alpha_\theta^H)}{\partial \sigma_\phi} - \frac{\partial F(\alpha_\theta^L)}{\partial \sigma_\phi}] + \gamma \qquad \text{(A. 40)}\\ &= \frac{v(1-\rho_o)}{\eta}\{\frac{f(\alpha_\theta^H)[-(b\rho_o\sigma_\beta - p - \theta)]}{[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2} - \frac{f(\alpha_\theta^L)[b\rho_o(1-\sigma_\beta) + p + \theta]}{[1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]^2}\} + \gamma = 0.\end{aligned}$$

The focus of interest is how the optimal level of bias towards bad news changes with the increased cost of news access. Applying the Implicit Function Theorem to equation (A. 40),

$$\begin{aligned}&\frac{f(\alpha_\theta^H)\{[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2 + 2(b\rho_o\sigma_\beta - p - \theta)[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi](1-\rho_o)\frac{\partial \sigma_\phi}{\partial \theta}\}}{[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^4} -\\ &\frac{f(\alpha_\theta^L)\{[1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]^2 + 2[b\rho_o(1-\sigma_\beta) + p + \theta][1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi][(1-\rho_o)\frac{\partial \sigma_\phi}{\partial \theta}]\}}{[1-\rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]^4}\\ &= 0, \qquad \text{(A. 41)}\end{aligned}$$

which is equivalent to

$$\frac{\partial \hat{\sigma}_\phi}{\partial \theta} = -\frac{f(\alpha_\theta^H)Pr(r=\beta)Pr(r=\phi)^3 - f(\alpha_\theta^L)Pr(r=\phi)Pr(r=\beta)^3}{2(1-\rho_o)[f(\alpha_\theta^H)(b\rho_o\sigma_\beta - p - \theta)Pr(r=\phi)^3 - f(\alpha_\theta^L)[b\rho_o(1-\sigma_\beta) + p + \theta]Pr(r=\beta)^3]}. \qquad \text{(A. 42)}$$

The sign of $\frac{\partial \hat{\sigma}_\phi}{\partial \theta}$ can be positive, negative, or zero. It means that $\hat{\sigma}_\phi$ may increase, decrease or remain unaffected by θ, depending on the parameter values.

The government has two potential strategies to manipulate the citizen's cost of news

access.

Case 1: Not manipulating news cost, i.e., $\theta = 0$.

In this case, manipulating the citizen's cost of accessing news backfires, i.e.,

$$\begin{aligned}\frac{\partial h(\theta, \sigma_\phi(\theta))}{\partial \theta} &= [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]'[1 - F(\alpha_\theta^H)] + [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi][1 - F(\alpha_\theta^H)]' \\ &\quad + [1 - \rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]'[1 - F(\alpha_\theta^L)] + [1 - \rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi][1 - F(\alpha_\theta^L)]' \\ &= -(1-\rho_o)\frac{\partial \sigma_\phi}{\partial \theta}[F(\alpha_\theta^H) - F(\alpha_\theta^L)] + [\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi][-f(\alpha_\theta^H)\frac{\partial \alpha_\theta^H(\theta, \sigma_\phi(\theta))}{\partial \theta}] \\ &\quad + [1 - \rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi][-f(\alpha_\theta^L)\frac{\partial \alpha_\theta^L(\theta, \sigma_\phi(\theta))}{\partial \theta}] < 0, \qquad \text{(A. 43)}\end{aligned}$$

where

$$\frac{\partial \alpha_\theta^H(\theta, \sigma_\phi(\theta))}{\partial \theta} = \frac{-[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi] - (b\rho_o\sigma_\beta - p - \theta)(1-\rho_o)\frac{\partial \sigma_\phi}{\partial \theta}}{\eta[\rho_o\sigma_\beta + (1-\rho_o)\sigma_\phi]^2}, \qquad \text{(A. 44)}$$

$$\frac{\partial \alpha_\theta^L(\theta, \sigma_\phi(\theta))}{\partial \theta} = \frac{[1 - \rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi] + [b\rho_o(1-\sigma_\beta) + p + \theta](1-\rho_o)\frac{\partial \sigma_\phi}{\partial \theta}}{\eta[1 - \rho_o\sigma_\beta - (1-\rho_o)\sigma_\phi]^2}, \qquad \text{(A. 45)}$$

in which $\frac{\partial \sigma_\phi}{\partial \theta}$ is given by (A. 42).

Comparing (A. 39) and (A. 43), it is evident that

$$\frac{\partial h(\theta, \sigma_\phi(\theta))}{\partial \theta} = \overbrace{\frac{\partial h(\theta)}{\partial \theta}}^{\text{effect of change in } \theta} + \overbrace{\frac{\partial h(\sigma_\phi(\theta))}{\partial \theta}}^{\text{effect of change in } \sigma_\phi \text{ due to change in } \theta}, \qquad \text{(A. 46)}$$

where

$$\frac{\partial h(\sigma_\phi(\theta))}{\partial \theta} = -(1-\rho_o)\frac{\partial \sigma_\phi}{\partial \theta}\{[F(\alpha_\theta^H) - F(\alpha_\theta^L)] - \frac{f(\alpha_\theta^H)\alpha_\theta^H - f(\alpha_\theta^L)\alpha_\theta^L}{\eta}\}. \qquad \text{(A. 47)}$$

Case 2: Increasing the cost of news access, i.e., $\theta \in (0, 1]$.

In this case, manipulating the citizen's cost of news access does not backfire (condition (A. 43) does not hold), i.e., $\frac{\partial h(\theta, \sigma_\phi(\theta))}{\partial \theta} > 0$.

The government's optimal level of news cost manipulation is achieved when

$$\frac{\partial \pi_G}{\partial \theta} = \frac{\partial h(\theta, \sigma_\phi(\theta))}{\partial \theta} - \tau = 0. \tag{A. 48}$$

The solution is

$$\theta^* = min\{max\{\frac{A}{B}, 0\}, 1\}, \tag{A. 49}$$

where

$$\begin{aligned} A = 2\eta\{f(\alpha_\theta^H)(b\rho_o\sigma_\beta - p)Pr(r=\phi)^3 - f(\alpha_\theta^L)[b\rho_o(1-\sigma_\beta)+p)]Pr(r=\beta)^3\}\{\tau \\ - \frac{\psi}{\eta}[f(\alpha_\theta^H) - f(\alpha_\theta^L)]\} - \psi[f(\alpha_\theta^H)Pr(r=\beta)Pr(r=\phi)^3 - f(\alpha_\theta^L)Pr(r=\phi)Pr(r=\beta)^3] \\ \{\eta Pr(r=\beta)Pr(r=\phi) - f(\alpha_\theta^H)Pr(r=\phi)(b\rho_o\sigma_\beta - p) + f(\alpha_\theta^L)Pr(r=\beta)[b\rho_o(1-\sigma_\beta)+p]\}, \end{aligned} \tag{A. 50}$$

$$\begin{aligned} B = 2\eta\{f(\alpha_\theta^H)Pr(r=\phi)^3 - f(\alpha_\theta^L)Pr(r=\beta)^3\}\{\tau - \frac{\psi}{\eta}[f(\alpha_\theta^H) - f(\alpha_\theta^L)]\}+ \\ \psi[f(\alpha_\theta^H)Pr(r=\beta)Pr(r=\phi)^3 - f(\alpha_\theta^L)Pr(r=\phi)Pr(r=\beta)^3] \\ [f(\alpha_\theta^H)Pr(r=\phi) + f(\alpha_\theta^L)Pr(r=\beta)]. \end{aligned} \tag{A. 51}$$

Appendix D Censorship and citizen welfare

D.1 Media bias manipulation and citizen welfare

When the government manipulates media biases and the outlet is market-driven, the effect of media bias on the aggregate welfare of citizens is as follows.

$$\frac{\partial W_I}{\partial \sigma_\beta} = \frac{\partial \alpha_p^H}{\partial \sigma_\beta} S_L(\alpha_p^H; \sigma) f(\alpha_p^H) - \frac{\partial \alpha_p^L}{\partial \sigma_\beta} S_H(\alpha_p^L; \sigma) f(\alpha_p^L). \tag{A. 52}$$

Given the definitions of α_p^H and α_p^L, it follows that $S_H(\alpha_p^H; \sigma) = S_L(\alpha_p^L; \sigma) = p > 0$. As probabilities, $f(\alpha_p^H)$ and $f(\alpha_p^L)$ fall within the range of zero to one. Moreover, there are $\frac{\partial \alpha_p^H}{\partial \sigma_\beta} > 0$ (A. 11) and $\frac{\partial \alpha_p^L}{\partial \sigma_\beta} < 0$ (A. 17). Consequently, the overall derivative $\frac{\partial W_I}{\partial \sigma_\beta} > 0$.

For a liberal outlet, the government has four potential strategies for manipulating media bias. First, if the government chooses not to censor, it is evident that the citizens' welfare remains unaffected. Second, if the government solely manipulates media bias towards good news, considering the liberal outlet's adjustment of bias towards bad news, then

$$\frac{\partial W_I}{\partial \sigma_\beta} = \frac{\partial \alpha_p^H(\sigma_\beta, \sigma_\phi(\sigma_\beta))}{\partial \sigma_\beta} S_L(\alpha_p^H; \sigma) f(\alpha_p^H) - \frac{\partial \alpha_p^L(\sigma_\beta, \sigma_\phi(\sigma_\beta))}{\partial \sigma_\beta} S_H(\alpha_p^L; \sigma) f(\alpha_p^L). \tag{A. 53}$$

The signs of $\frac{\partial \alpha_p^H(\sigma_\beta, \sigma_\phi(\sigma_\beta))}{\partial \sigma_\beta}$ (given by (A. 32)) and $\frac{\partial \alpha_p^L(\sigma_\beta, \sigma_\phi(\sigma_\beta))}{\partial \sigma_\beta}$ (given by (A. 33)), and accordingly the sign of the overall derivative $\frac{\partial W_I}{\partial \sigma_\beta}$, can be positive, negative, or zero. Third, if the government solely manipulates media bias towards bad news, then

$$\frac{\partial W_I}{\partial \sigma_\phi} = \frac{\partial \alpha_p^H}{\partial \sigma_\phi} S_L(\alpha_p^H; \sigma) f(\alpha_p^H) - \frac{\partial \alpha_p^L}{\partial \sigma_\phi} S_H(\alpha_p^L; \sigma) f(\alpha_p^L). \tag{A. 54}$$

Since $\frac{\partial \alpha_p^H}{\partial \sigma_\phi} < 0$ (A. 12) and $\frac{\partial \alpha_p^L}{\partial \sigma_\phi} > 0$ (A. 18), the overall derivative $\frac{\partial W_I}{\partial \sigma_\phi} < 0$.

D.2 News cost manipulation and citizen welfare

When the government increases the citizens' cost of news access by $\theta \in (0, 1]$, the liberal outlet adjusts its level of bias towards bad news accordingly. The effect of the increased news cost on the aggregate welfare of citizens is as follows.

$$\frac{\partial W_I}{\partial \theta} = \frac{\partial \alpha_\theta^H(\theta, \sigma_\phi(\theta))}{\partial \theta} S_L(\alpha_\theta^H; \sigma) f(\alpha_\theta^H) - \frac{\partial \alpha_\theta^L(\theta, \sigma_\phi(\theta))}{\partial \theta} S_H(\alpha_\theta^L; \sigma) f(\alpha_\theta^L). \tag{A. 55}$$

When the liberal outlet's bias towards bad news increases in the citizen's cost of new access, i.e., $\frac{\partial \sigma_\phi}{\partial \theta} > 0$, it can be observed that $\frac{\partial \alpha_\theta^H(\theta, \sigma_\phi(\theta))}{\partial \theta} < 0$ (A. 44) and $\frac{\partial \alpha_\theta^L(\theta, \sigma_\phi(\theta))}{\partial \theta} > 0$ (A. 45). Consequently, the overall derivative $\frac{\partial W_I}{\partial \theta} < 0$. If the liberal outlet's bias towards bad news decreases in the citizen's cost of new access, i.e., $\frac{\partial \sigma_\phi}{\partial \theta} < 0$, the signs of $\frac{\partial \alpha_\theta^H(\theta, \sigma_\phi(\theta))}{\partial \theta}$ and $\frac{\partial \alpha_\theta^L(\theta, \sigma_\phi(\theta))}{\partial \theta}$, and accordingly the sign of $\frac{\partial W_I}{\partial \theta}$, can be positive, negative, or zero.