Theoretical Model

Supplement to *Social Media and Government Responsiveness: Evidence from Vaccine Procurement in China*

Yixin Mei, Ruoyu Qian, and Yanhui Wu

July 2026

This appendix presents the full model of the brief version in the paper, *Social Media and Government Responsiveness: Evidence from Vaccine Procurement in China*, with complete derivations and proofs.

1 Model

Players and states. A central government, the principal, delegates a procurement decision to a local official, the agent. The state of the world is risky or safe, $\theta \in \{R, S\}$, with prior $\pi \equiv \Pr(\theta = R) \in (0, 1)$. The agent chooses either rule-following open-bid procurement O or discretionary procurement D.

Payoffs. Open bidding is transparent but rigid and incurs an operational cost $c > 0$. Discretion is efficient when the state is safe but hazardous when risky: if the agent chooses D in state R and the choice is not corrected by oversight, safety failure occurs with probability $p \in (0, 1]$ and causes social harm $h > 0$. We assume $ph > c$, so that, in the risky state, open bidding is socially preferable to unmonitored discretion. The agent earns a private benefit $b > 0$ from discretion (e.g., rents from flexible contracting, relationship capital, or corruption) and zero from open bidding.

Information. The agent privately observes a symmetric binary signal $\hat{\theta} \in \{r, s\}$ with accuracy $q \in (\frac{1}{2}, 1)$. He acts on his posterior belief about the risky state, given by Bayes' rule:

$$\rho_r = \frac{\pi q}{\pi q + (1-\pi)(1-q)}, \qquad \rho_s = \frac{\pi(1-q)}{\pi(1-q) + (1-\pi)q}. \tag{1}$$

Given $q > \frac{1}{2}$, $\rho_r > \rho_s$: a risky signal raises the agent's assessment of danger more than a safe signal does. Note also $\rho_s < \pi < \rho_r$, and the identity $\pi(1-q) = \rho_s \pi_D$, where $\pi_D \equiv \pi(1-q) + (1-\pi)q$, which we use repeatedly below.

Monitoring. The principal commits to a safe-harbor scheme: open-bid O is never audited; a choice of D is audited with intensity $m \in [0,1]$. An audit reveals the state correctly. If it uncovers risky discretion, the principal overturns the procurement format from D to O and penalizes the agent by $C > 0$; the agent retains the private benefit b already accrued.[1] Auditing costs $\mu(V)\,k(m)$, where k is increasing and strictly convex in effort m with $k'(0) = 0$, and $\mu(V) > 0$ scales the cost with the information environment defined below. Auditing is thus allocatively valuable.

Social media: volume and composition. Let $V \in [0,1]$ denote total social media exposure in a jurisdiction and $\sigma \in [0,1]$ its sentiment composition, the share of content that is emotive rather than informational. Via the *information channel*, exposure lowers monitoring cost through $\mu(V)$:

$$\mu'(V) < 0, \qquad \mu''(V) > 0, \tag{2}$$

so μ is decreasing and convex: the first posts reveal genuine local conditions while later posts are largely retweets that add little new information, so the marginal cost saving $-\mu'(V)$ fades as volume grows. Via the *sentiment channel*, negative sentiment towards the government undermines public trust in the regime, so the sentiment-weighted exposure σV increases the principal's *perceived* harm:

$$\tilde{h}(\sigma V) \geq h, \qquad \tilde{h}(0) = h, \qquad \tilde{h}' > 0, \qquad \tilde{h}'' > 0, \tag{3}$$

where strict convexity captures the idea that outrage compounds when posts go viral. The contrast in curvature between (2) and (3) means that when volume rises, the informational return fades while the sentiment force accelerates, tilting the system toward the sentiment channel even when composition is fixed.

For expositional convenience, we collect the regularity conditions as below.

Assumption 1.

(a) ***Perceived benefit.*** *$p\,\tilde{h}(\sigma V) > c$ for all (σ, V), so auditing a discretionary case is worthwhile at the perceived harm.*

[1] If detection also strips the rent, all bounds below hold with C replaced by $b + C$; no result changes.

(b) ***Interior band.*** *The truth-telling band* (4) *is interior:* $0 < \underline{m} < \bar{m} < 1$.

(c) ***Interior pressure.*** *The principal's unconstrained optimum* $m^P(V, \sigma)$ *defined by* (6) *lies in* $(0, 1)$.

(d) ***Capitulation is reachable.*** *Sentiment can push desired pressure past the band:* $m^P(1, \sigma) > \bar{m}$ *for the relevant* σ*, so an interior capitulation threshold* $\bar{V}(\sigma) \in (0, 1)$ *exists.*

(e) ***Fading information return.*** $-\mu'(V) \to 0$ *as* $V \to 1$*, so the marginal informational cost saving is exhausted at high volume.*

(f) ***Risk is the exception.*** $\pi < \frac{1}{2}$*: the safe state is more likely than the risky state, so safety failures are rare events.*

(g) ***Heterogeneity range*** *(used only in Proposition 3). Rents are dispersed,* $b \sim U[0, \bar{b}]$*, and for all* (V, σ) *in the empirically relevant region* $C\rho_s\, m^P(V, \sigma) < \bar{b} \le C\rho_r\, m^P(V, \sigma)$*: a positive fraction of cases capitulate while in no case is rent-seeking discretion left undeterred after a risky signal. Since* ρ_r/ρ_s *is large when the risky state is rare and signals are accurate, this range is wide.*

2 Truth-Telling, Inspection, and Monitoring Failure

We focus on the truth-telling equilibrium in which the agent chooses open bidding after a risky signal and discretion after a safe one.

Lemma 1 (Truth-telling band)**.** *The agent follows his signal if and only if monitoring lies in an intermediate band,*

$$\underline{m} \equiv \frac{b}{C\rho_r} \;\le\; m \;\le\; \frac{b}{C\rho_s} \equiv \bar{m}, \tag{4}$$

which is nonempty. If $m < \underline{m}$ *he chooses discretion after either signal; if* $m > \bar{m}$ *he converts to universal open bidding (capitulation).*

Proof. After signal $x \in \{r, s\}$, discretion yields b minus the expected penalty, which is incurred only if the state is risky (posterior ρ_x) and the audit detects it (probability m). So discretion is worth $b - mC\rho_x$ against zero from open bidding. The agent chooses O after r iff $m \ge b/(C\rho_r)$ and D after s iff $m \le b/(C\rho_s)$; both hold iff (4) holds. Since $\rho_r > \rho_s$, the band is nonempty. If $m < \underline{m}$, discretion is strictly preferred after both signals; if $m > \bar{m}$, open bidding is strictly preferred after both. □

The lower bound $\underline{m}$ deters rent-seeking discretion when the signal is risky; the upper bound $\bar{m}$ keeps the agent from abandoning efficient discretion when the signal is safe. By Assumption 1(b) the band is interior.

Under signal following, audits fall on discretionary choices, which arise after the safe signal with probability $\pi_D = \pi(1-q) + (1-\pi)q$, of which the genuinely risky fraction is $\pi(1-q)$. Correcting one risky case saves the perceived harm $p\tilde{h}$ at cost c, so the principal chooses m to maximize

$$\Phi(m) = m\,\pi(1-q)\left[p\,\tilde{h}(\sigma V) - c\right] - \mu(V)\,k(m)\,\pi_D. \tag{5}$$

Lemma 2 (Inspection pressure)**.** *The principal's desired inspection pressure $m^P(V,\sigma)$ is the unique solution to*

$$\mu(V)\,k'(m^P)\,\pi_D = \pi(1-q)\left[p\,\tilde{h}(\sigma V) - c\right], \tag{6}$$

and it rises with volume through both channels and with the sentiment share: $\partial m^P/\partial V > 0$ and $\partial m^P/\partial\sigma > 0$ for $V > 0$. Moreover, m^P falls with signal quality, $\partial m^P/\partial q < 0$, and does not depend on the penalty C.

Proof. $\Phi''(m) = -\mu(V)k''(m)\pi_D < 0$, so Φ is strictly concave; with $k'(0) = 0$ and $p\tilde{h} > c$ (Assumption 1(a)), the maximizer is interior (Assumption 1(c)) and uniquely determined by (6). Implicitly differentiating (6),

$$\frac{\partial m^P}{\partial V} = \frac{\overbrace{-\mu'(V)\,k'(m^P)\,\pi_D}^{\text{information}} + \overbrace{\pi(1-q)\,p\,\sigma\,\tilde{h}'(\sigma V)}^{\text{sentiment}}}{\mu(V)\,k''(m^P)\,\pi_D} > 0, \tag{7}$$

and, symmetrically, $\partial m^P/\partial\sigma > 0$, since both are governed by $\mu' < 0$, $\tilde{h}' > 0$, and $k', k'' > 0$. For signal quality, use the identity $\pi(1-q) = \rho_s\pi_D$ to rewrite (6) as $\mu(V)\,k'(m^P) = \rho_s(q)\,[\,p\tilde{h}(\sigma V) - c\,]$, in which π_D has cancelled. From (1),

$$\rho_s'(q) = -\frac{\pi(1-\pi)}{[\pi(1-q)+(1-\pi)q]^2} < 0, \tag{8}$$

so the right side falls in q; since k' is strictly increasing, $\partial m^P/\partial q < 0$: better-informed agents are audited less. Finally, C does not appear in (6): the fine is a transfer that creates no allocative value, so the principal does not audit to collect it. □

We adopt the inspection-pressure interpretation: once public visibility triggers top-down scrutiny, the desired pressure $m^P(V,\sigma)$ is the effective audit intensity the official faces, and it is set by a superior authority responding to political harm $\tilde{h}$ rather than chosen by a benevolent planner who internalizes capitulation. Note that (6) characterizes the superior authority's desired inspection pressure conditional on the truth-telling benchmark. Once this pressure exceeds the band, the model interprets the resulting effective pressure as politically imposed scrutiny rather than a solution to a fully internalized principal optimization problem over all behavioral regimes.

A key discrepancy arises in the assessment of welfare. Citizens care about the actual harm h, while the regime, which has a political goal, cares about $\tilde{h}(\sigma V)$. Define m^I as the intensity that maximizes citizen welfare, i.e., the solution to (6) with h in place of $\tilde{h}$:

$$\mu(V)\, k'(m^I)\, \pi_D = \pi(1-q)\,(ph-c). \tag{9}$$

Lemma 3 (Two monitoring failures)**.** *The principal weakly over-monitors: $m^P \geq m^I$, with equality if and only if $\sigma V = 0$. Consequently, when $\sigma V > 0$ citizen welfare is strictly below its maximum: under-monitoring ($m < m^I$) leaves risky discretion undeterred and sustains safety failures, while over-monitoring ($m > m^I$) wastes audit resources and pushes officials toward inefficient rule-following.*

Proof. Since $\tilde{h}(\sigma V) \geq h$ with equality iff $\sigma V = 0$, the right side of (6) weakly exceeds that of (9); as k' is strictly increasing, $m^P \geq m^I$ with equality iff $\sigma V = 0$. Citizen welfare under signal following at intensity m (with the transfers b and C netting out) is

$$W(m;V) = -c\,\Pr(\hat{\theta} = r) + \pi(1-q)\,[-mc - (1-m)ph] - \mu(V)\,k(m)\,\pi_D, \tag{10}$$

which is strictly concave in m with first-order condition (9), so m^I is its unique maximizer and $W(m^I;V) - W(m^P;V) > 0$ whenever $m^P \neq m^I$. □

The key insight of Lemma 3 is that the information channel attacks under-monitoring by making oversight cheaper, while the sentiment channel creates over-monitoring by inflating political harm. Social media can therefore cure one monitoring failure while causing another.

Theorem 1 (When social media is welfare enhancing)**.** *Below the capitulation threshold, a marginal increase in social media visibility raises citizen welfare if and only if the direct cost-saving effect exceeds the belief-distortion effect:*

$$\underbrace{-\mu'(V)\,k(m^P)\,\pi_D}_{\textit{direct cost-saving}} > \underbrace{\pi(1-q)\,p\left[\tilde{h}(\sigma V) - h\right]\frac{\partial m^P}{\partial V}}_{\textit{belief-distortion}}. \tag{11}$$

Purely informational exposure ($\sigma = 0$) is always welfare enhancing, whereas for $\sigma V > 0$ the condition eventually fails as volume grows, because the direct cost-saving diminishes while *the distortion compounds.*

Proof. Below capitulation, realized welfare is $W^*(V) = W(m^P(V, \sigma); V)$. By the chain rule,

$$\frac{dW^*}{dV} = -\mu'(V)\, k(m^P)\, \pi_D + \left.\frac{\partial W}{\partial m}\right|_{m=m^P} \cdot \frac{\partial m^P}{\partial V}.$$

Differentiating (10) in m and substituting (6),

$$\left.\frac{\partial W}{\partial m}\right|_{m=m^P} = \pi(1-q)(ph-c) - \mu(V)k'(m^P)\pi_D = -\,\pi(1-q)\,p\left[\tilde{h}(\sigma V) - h\right] \le 0,$$

so $dW^*/dV > 0$ iff (11) holds. If $\sigma = 0$ then $\tilde{h} = h$, the distortion vanishes, and $dW^*/dV = -\mu'(V)k(m^P)\pi_D > 0$. If $\sigma V > 0$, the left side of (11) shrinks as V rises, because $-\mu'(V) \to 0$ (Assumption 1(e)) and $k(m^P) \le k(\bar{m})$ is bounded, while the right side grows: the wedge $\tilde{h}(\sigma V) - h$ is strictly increasing with increasing slope (by $\tilde{h}'' > 0$) and $\partial m^P/\partial V$ stays positive by (7). So, for any fixed positive sentiment share, the welfare effect of additional visibility eventually turns negative within the signal- following regime. □

3 Testable Propositions

Proposition 1 (Visibility and error rates)**.** *There exists a social media visibility threshold $\bar{V}$, below which the agent follows his signal, and when visibility increases, the principal monitors more, observed open bidding rises, type-I error falls, and type-II error is constant. When social media visibility exceeds $\bar{V}$, the agent discards his private information, capitulates to adopt universal open bidding, and type-II error jumps up.*

Proof. When the agent follows his signal, open bidding is observed in two disjoint events: he picks O directly after a risky signal, with probability $\Pr(\hat{\theta} = r) = \pi q + (1-\pi)(1-q)$; or he picks D after a safe signal, the state is in fact risky (probability $\pi(1-q)$), and the audit detects and converts the case (probability m^P). Summing, and computing the errors analogously,

$$B = \pi q + (1-\pi)(1-q) + m^P\,\pi(1-q), \quad T_1 = p\,(1-m^P)\,\pi(1-q), \quad T_2 = (1-\pi)(1-q), \tag{12}$$

where T_1 requires the risky state, a misleading safe signal, no correction, and a bad draw, and T_2 arises only through a mistaken risky signal in the safe state (independent of pressure). Under capitulation the agent chooses O after both signals, so $B = 1$, $T_1 = 0$, and $T_2 = 1 - \pi$.

By Lemma 2 and Assumption 1(d), m^P is continuous and strictly increasing in V with

$m^P(1,\sigma) > \bar{m}$, so $\bar{V}$ solving $m^P(\bar{V},\sigma) = \bar{m}$ exists in $(0,1)$ and is unique. For $V < \bar{V}$, pressure lies in the band and (12) applies: $\partial B/\partial V = \pi(1-q)\,\partial m^P/\partial V > 0$, $\partial T_1/\partial V = -p\,\pi(1-q)\,\partial m^P/\partial V < 0$, and T_2 does not depend on V. For $V > \bar{V}$, pressure exceeds $\bar{m}$, so by Lemma 1 the agent capitulates: B jumps to 1, T_1 drops to 0, and T_2 jumps from $(1-\pi)(1-q)$ up to $1-\pi$. □

Proposition 2 (Information and sentiment channels)**.**

(i) ***Separation case.*** *Both informational and sentiment posts increase open bidding. Moreover, informational marginal effects weaken as informational exposure accumulates, while sentiment marginal effects strengthen as emotive exposure accumulates.*

(ii) ***Entanglement.*** *Holding the sentiment composition constant, a larger volume of social media posts increases the share of open bidding.*

Proof. (i) Under the stated separation, informational posts $V_I = (1-\sigma)V$ enter only the monitoring cost $\mu(V_I)$ and sentiment posts $V_\Sigma = \sigma V$ enter only the perceived harm $\tilde{h}(V_\Sigma)$. Implicitly differentiating (6) with μ evaluated at V_I and $\tilde{h}$ at V_Σ,

$$\frac{\partial m^P}{\partial V_I} = \frac{-\mu'(V_I)\,k'(m^P)}{\mu(V_I)\,k''(m^P)} > 0, \qquad \frac{\partial m^P}{\partial V_\Sigma} = \frac{\pi(1-q)\,p\,\tilde{h}'(V_\Sigma)}{\mu(V_I)\,k''(m^P)\,\pi_D} > 0,$$

and by (12) both raise B through the coefficient $\pi(1-q)$. Along a path of rising volume, $-\mu'(V_I)$ shrinks toward zero (by $\mu'' > 0$ and Assumption 1(e)) while $\tilde{h}'(V_\Sigma)$ is strictly increasing (by $\tilde{h}'' > 0$); hence beyond a threshold volume $\partial m^P/\partial V_\Sigma > \partial m^P/\partial V_I$. (At low volume the ranking may reverse, since the first informational posts are highly revealing.) For capitulation, Lemma 3 shows pure information keeps pressure at the citizen optimum m^I; hence whenever $m^I \leq \bar{m}$, information alone never pushes pressure past the band, and $m^P > \bar{m}$ requires the sentiment wedge $\tilde{h} - h > 0$.

(ii) Holding σ constant, Lemma 2 gives $\partial m^P/\partial V > 0$, so $\partial B/\partial V = \pi(1-q)\,\partial m^P/\partial V > 0$ by (12). Within (7) the information term carries $-\mu'(V)$, which shrinks as V rises, while the sentiment term carries $\sigma\tilde{h}'(\sigma V)$, which grows; so at high volume the increase in open bidding is dominated by the sentiment channel. By Theorem 1 the accountability gain is mitigated. □

To derive the interaction effects estimated in the empirical analysis, we let the private benefit vary across procurement cases: $b \sim U[0,\bar{b}]$, drawn independently of the signal and the state (Assumption 1(g)). The band limit $\bar{m} = b/(C\rho_s)$ is then case-specific, so capitulation becomes an extensive margin with positive mass rather than a single threshold event, and the response of open bidding to visibility is smooth. For closed forms we adopt the quadratic audit cost $k(m) = \kappa m^2/2$, under which (6) yields

$$m^P = \rho_s(q)\, A(V,\sigma), \qquad A \equiv \frac{p\,\tilde{h}(\sigma V) - c}{\kappa\, \mu(V)}, \qquad A_V > 0. \tag{13}$$

Lemma 4 (Smoothed capitulation margin)**.** *Under Assumption 1(g), the fraction of cases in which the official capitulates is*

$$P(V; q, C) = \frac{C \rho_s\, m^P(V,\sigma)}{\bar{b}} \in (0,1), \tag{14}$$

and the expected open-bid share is

$$\mathbb{E}[B] = P \cdot 1 + (1-P)\, B_{\text{band}}, \qquad B_{\text{band}} \equiv \pi q + (1-\pi)(1-q) + m^P\, \pi(1-q). \tag{15}$$

Moreover, $P < \pi C/\bar{b}$.

Proof. A case with rent b capitulates iff $m^P > \bar{m}(b) = b/(C\rho_s)$, i.e., iff $b < C\rho_s m^P$; under $b \sim U[0,\bar{b}]$ this has probability (14), interior by Assumption 1(g). The same assumption, $\bar{b} \le C\rho_r m^P$, ensures every non-capitulating case follows the signal; (12) then gives (15). Finally, $m^P \le 1$ and $\rho_s < \pi$ imply $P < \pi C/\bar{b}$. □

Proposition 3 (Heterogeneous effects)**.** *Maintain Assumptions 1(f)–(g) and the quadratic audit cost (13).*

(i) ***Signal quality.*** *If $\bar{b} \ge 5\,\pi C$,[2] then the effect of social media visibility on the share of open bidding is dampened by the agent's signal quality: $\partial^2 \mathbb{E}[B]/\partial V\, \partial q < 0$. The interaction operates through both margins: within the band, higher q reduces the need to responses to auditing; at the capitulation margin, higher q reduces the flow of cases tipped into universal open bidding.*

(ii) ***Penalty.*** *The effect of social media visibility on the share of open bidding is amplified by the penalty: $\partial^2 \mathbb{E}[B]/\partial V\, \partial C > 0$. The interaction operates only through the capitulation margin: within the band, B_{band} and m^P are independent of C (Lemma 2).*

Proof. Substituting (13) into (15) and using $\pi(1-q) = \rho_s \pi_D$,

$$B_{\text{band}} = \Pr(\hat{\theta} = r) + \rho_s^2 \pi_D A, \qquad P = \frac{C\rho_s^2 A}{\bar{b}}, \qquad 1 - B_{\text{band}} = \pi_D(1 - \rho_s^2 A),$$

[2]This condition is sufficient but not necessary. It means that rents are sufficiently dispersed relative to the prior-weighted penalty. The condition holds for a wide range of parameter values. Violations arise only in extreme corners where capitulation is near-universal, the signal is nearly uninformative, and the prior is close to $\frac{1}{2}$.

so, writing $\mathbb{E}[B] = B_{\text{band}} + P(1 - B_{\text{band}})$ and differentiating in V,

$$\frac{\partial \mathbb{E}[B]}{\partial V} = \rho_s^2 \pi_D A_V \, \Psi, \qquad \Psi \equiv (1 - P) + \frac{C}{\bar{b}} \left(1 - \rho_s^2 A\right) > 1 - P, \tag{16}$$

where A and A_V depend on neither q nor C.

(ii) Penalty. Only Ψ depends on C. Since $\rho_s^2 A = \rho_s m^P$,

$$\frac{\partial^2 \mathbb{E}[B]}{\partial V \, \partial C} = \rho_s^2 \pi_D A_V \frac{\partial \Psi}{\partial C} = \frac{\rho_s^2 \pi_D A_V}{\bar{b}} \left(1 - 2\rho_s m^P\right) > 0.$$

Since $\rho_s m^P < \rho_s < \pi < \frac{1}{2}$, by Assumptions 1(c),(f) (using $m^P < 1$). As C enters only through P, the within-band component is identically zero.

(i) Signal quality. Let $r \equiv |\rho_s'|/\rho_s = \frac{1}{1-q} + \frac{1-2\pi}{\pi_D} > 0$. From (8) and $\pi_D' = 1 - 2\pi$,

$$\frac{\partial \ln(\rho_s^2 \pi_D)}{\partial q} = -2r + \frac{1 - 2\pi}{\pi_D} < -r < 0.$$

Both $\partial B_{\text{band}}/\partial V = \rho_s^2 \pi_D A_V$ and $\partial P/\partial V = C\rho_s^2 A_V/\bar{b}$ carry the factor ρ_s^2, strictly decreasing in q, so each margin of the interaction is negative. For the total effect, $\rho_s^2 A = P\bar{b}/C$ and $\partial_q(\rho_s^2 A) = -2r\rho_s^2 A$ give $\partial_q \Psi = 4rP$, hence

$$\frac{\partial}{\partial q} \ln \frac{\partial \mathbb{E}[B]}{\partial V} < -r + \frac{4rP}{\Psi} = r\left(\frac{4P}{\Psi} - 1\right) < 0,$$

because $\Psi > 1 - P > \frac{4}{5} > 4P$ whenever $P < \frac{1}{5}$, which holds under $\bar{b} \geq 5\pi C$ by Lemma 4. $\square$

Here, the dispersion condition $\bar{b} \geq 5\pi C$ guarantees that capitulation is the exception rather than the rule, disciplining the opposing weight effect (fewer capitulating cases shifts mass onto the band) that higher q would otherwise generate. In the limit of degenerate heterogeneity ($\bar{b} \to b$), the extensive margin collapses back to the single threshold $\bar{V}$ of Proposition 1, with $\partial \bar{V}/\partial q > 0$ and $\partial \bar{V}/\partial C < 0$.