Online Appendix to: Evolutionarily Stable Benchmarks

Endogenous Information Production, Manipulation, and Focal Points

Andrew Bird[*] Steve Karolyi[†] Thomas Ruchti[‡]

This version: April 2026

Abstract

This Online Appendix contains robustness results, extensions, and supplementary proofs for "Evolutionarily Stable Benchmarks" by Bird, Karolyi, and Ruchti.

[*]Argyros College of Business and Economics, Chapman University. Email: `abird@chapman.edu`.
[†]School of Business, George Mason University. Email: `skarolyi@gmu.edu`.
[‡]Pamplin College of Business, Virginia Tech. Email: `ruchti@vt.edu`.

OA.1 Robustness and Extensions

OA.1.1 Robustness to Cross-Benchmark Effects

The exact potential game structure (Proposition 5) relies on the own-share-only property. Cross-benchmark effects can arise from information spillovers, capacity constraints, or attention effects, replacing the payoff with $\tilde{U}_k(\mathbf{x})$ that depends on the full population state. We show the main result is robust, because when cross-benchmark effects are small, the stochastically stable benchmark is exactly the same as in the potential game.

Setup. Consider a perturbed game in which evaluator payoffs take the form

$$\tilde{U}_k(\mathbf{x}) = U_E^{\text{info}}(b_k, x_k) + g_k(\mathbf{x}), \tag{1}$$

where $U_E^{\text{info}}(b_k, x_k)$ is the own-share payoff from the exact potential game and $g_k(\mathbf{x})$ captures cross-benchmark effects. We assume the uniform bound

$$\sup_{\mathbf{x}\in\Delta^{n-1}} |g_k(\mathbf{x})| \leq \delta \qquad \text{for all } k \in \{1, \dots, n\}. \tag{2}$$

Proposition OA.1 (Robustness of Stochastic Stability)**.** *Let $\Delta \equiv \Phi(\mathbf{e}_{k^*}) - \max_{j\neq k^*} \Phi(\mathbf{e}_j) > 0$ be the potential gap at the stochastically stable benchmark b_{k^*} of the exact potential game, and let $L > 0$ be a Lipschitz constant for $U_E^{\text{info}}(b_k, \cdot)$ on $[0, 1]$, uniform across benchmarks. There exists a constant $C \geq 0$ depending only on L and on the curvature of the perturbed log-linear response such that if the cross-benchmark effects satisfy*

$$\delta < \frac{\Delta}{2(1 + C)}, \tag{3}$$

then the unique stochastically stable benchmark of the perturbed game (1) is the same as that of the exact potential game:

$$\tilde{b}_{\text{SS}} = b_{k^*} = b_{\text{SS}}.$$

That is, sufficiently small cross-benchmark effects do not merely perturb the stochastically stable benchmark—they leave it unchanged.

Proof. We work in the Blume log-linear response framework used in the body (Theorem 1), with N evaluators choosing among n benchmarks. The argument compares Gibbs measures rather than resistance trees and stays within one framework throughout.

Step 1: Almost-potential of the perturbed game. Define a perturbed potential candidate

$$\tilde{\Phi}_N(\mathbf{n}) \;\equiv\; \Phi_N(\mathbf{n}) + \frac{1}{N}\sum_{k=1}^{n} n_k\, g_k(\mathbf{n}/N),$$

where Φ_N is the exact potential of the unperturbed game from Step 2 of the proof of Theorem 1. A single-evaluator switch from b_k to b_l changes $\tilde{\Phi}_N$ by

$$\Delta\tilde{\Phi}_N = \frac{1}{N}\big[\tilde{U}_l(\mathbf{n}') - \tilde{U}_k(\mathbf{n})\big] + R(\mathbf{n},\mathbf{n}'),$$

where the residual R collects the deviation of $\sum_k n_k g_k/N$ from a true potential. Since $|g_k| \le \delta$ on the simplex, $|R| \le 4\delta/N$ at every switch, so the perturbed game is an ε-potential game in the sense of Monderer and Shapley (1996) with $\varepsilon = 4\delta/N$.

Step 2: Gibbs concentration with bounded perturbation. Under log-linear response, the unique stationary distribution of the perturbed chain satisfies

$$\log\tilde{\mu}^N_\beta(\mathbf{n}) - \log\tilde{\mu}^N_\beta(\mathbf{n}') = \beta\,[\tilde{V}_N(\mathbf{n}) - \tilde{V}_N(\mathbf{n}')],$$

for a function $\tilde{V}_N$ that differs from the perturbed potential by a uniformly bounded error: $|\tilde{V}_N(\mathbf{n}) - \tilde{\Phi}_N(\mathbf{n})| \le C\delta$ for a constant $C > 0$ that depends only on n and the Lipschitz constant L (this follows from Blume (1993, Theorem 3.1) applied to perturbed log-linear dynamics; see also Sandholm 2010, Section 12). As $\beta \to \infty$, the perturbed Gibbs measure concentrates on $\arg\max_{\mathbf{n}} \tilde{V}_N(\mathbf{n})$.

Step 3: Maximizer is preserved when the gap exceeds the perturbation. At monomorphic states, $\tilde{\Phi}_N(\mathbf{e}^N_k) = \Phi_N(\mathbf{e}^N_k) + g_k(\mathbf{e}^N_k)$ (since only b_k has nonzero share at $\mathbf{e}^N_k$), and $|g_k(\mathbf{e}^N_k)| \le \delta$. Combined with the bound from Step 2,

$$|\tilde{V}_N(\mathbf{e}^N_k) - \Phi_N(\mathbf{e}^N_k)| \le \delta + C\delta = (1+C)\delta$$

uniformly across k. Let $\Delta_j \equiv \Phi(\mathbf{e}_{k^*}) - \Phi(\mathbf{e}_j) \ge \Delta > 0$ for each $j \ne k^*$. By the triangle inequality applied at the perturbed values,

$$\tilde{V}_N(\mathbf{e}^N_{k^*}) - \tilde{V}_N(\mathbf{e}^N_j) \ge \Phi_N(\mathbf{e}^N_{k^*}) - \Phi_N(\mathbf{e}^N_j) - 2(1+C)\delta \ge \Delta - 2(1+C)\delta - o(1)$$

as $N \to \infty$ (the $o(1)$ absorbs the Riemann-sum approximation error in passing between Φ_N and Φ). For this gap to remain strictly positive for every $j \ne k^*$ simultaneously, it suffices to take

$$\delta < \frac{\Delta}{2(1+C)},$$

which is (3). The bound does not depend on n because the worst-case gap Δ is already the minimum over the $n-1$ alternatives by definition, and the per-alternative bound is uniform.

Step 4: Conclusion. For δ satisfying (3), $\arg\max_{\mathbf{n}} \tilde{V}_N(\mathbf{n}) = \mathbf{e}^N_{k^*}$ for all sufficiently large N. Sending $\beta \to \infty$ and then $N \to \infty$, the unique stochastically stable state of the perturbed game is $\mathbf{e}_{k^*}$, the same as in the exact potential game. $\square$

The condition is tight up to constants and most permissive when the potential gap Δ is large.

Application: Capacity-constrained information producers. We now verify that the robustness condition is satisfied in the economically relevant case where information producers face cross-benchmark crowding.

As a concrete application, consider information investment at benchmark b_k subject to crowding from other benchmarks,

$$I_k(\mathbf{x}) = \frac{r \cdot x_k}{\kappa_I + \gamma \sum_{j \neq k} x_j}, \tag{4}$$

where $\gamma \geq 0$ is the crowding parameter. When $\gamma = 0$, this reduces to the baseline model $I_k = r x_k / \kappa_I$. When $\gamma > 0$, higher evaluator adoption of other benchmarks raises the effective cost of information production at b_k (e.g., because analysts drawn to other benchmarks are unavailable). If

$$\gamma < \frac{\kappa_I \cdot \Delta}{2\,L\,(1 + C)}, \tag{5}$$

where L is the Lipschitz constant of $U_E^{\text{info}}(b_k, \cdot)$ on $[0, 1]$ and C is the constant from Proposition OA.1, then the stochastically stable benchmark under crowding is the same as in the baseline model. To see this, write the evaluator's payoff under crowding as $\tilde{U}_k(\mathbf{x}) = U_E^{\text{info}}(b_k, x_k) + g_k(\mathbf{x})$, where $g_k(\mathbf{x}) = U_E(b_k; I_k(\mathbf{x})) - U_E(b_k; r x_k/\kappa_I)$ captures the departure from the baseline. At any state $\mathbf{x}$, since $\sum_{j \neq k} x_j \leq 1$, we have

$$|I_k(\mathbf{x}) - I_k^0| = r x_k \cdot \frac{\gamma \sum_{j \neq k} x_j}{\kappa_I(\kappa_I + \gamma \sum_{j \neq k} x_j)} \leq \frac{r\gamma}{\kappa_I^2}.$$

By the mean value theorem, $|g_k(\mathbf{x})| \leq L \cdot r\gamma/\kappa_I^2 = L\gamma/\kappa_I \cdot (r/\kappa_I)$, where $L = \sup_I |\partial U_E / \partial I|$ enters as the Lipschitz constant. Applying Proposition OA.1 with $\delta = L\gamma/\kappa_I$ yields (5).

For the two-benchmark earnings application, the condition is mild given substantial analyst benchmark specialization.

OA.1.2 Non-monotone Manipulation Costs

With non-monotone costs (e.g., U-shaped $m(\theta)$ in credit ratings), the results survive under a single-crossing condition. If $\phi(\theta) = \frac{1}{2} m(\theta)(b - \theta)^2$ crosses Δu exactly once from above on $(0, b)$, a unique marginal type separates manipulators from non-manipulators, and all downstream results carry through unchanged.

OA.1.3 Smooth Reward Functions

Replacing the step-function reward with a smooth sigmoid $R_\eta(s,b) = u_L + \Delta u \cdot \Psi((s-b)/\eta)$ (where $\eta > 0$ controls steepness) generates a smooth manipulation function that converges to the bang-bang strategy as $\eta \to 0$. The qualitative bunching pattern (peak near b, thinning below) survives for all $\eta > 0$, and the potential game structure is preserved because the own-share property of U_E^{info} depends on the information technology, not the reward function. Since $I^*(b_k, x_k) = r x_k/\kappa_I$ is determined by evaluator demand regardless of the reward function's shape, $U_E^{\text{info}}(b_k, x_k)$ depends only on x_k for any reward specification.

OA.1.4 Supporting Technical Results

Lemma OA.2 (Continuity of the Evaluator Payoff)**.** *The evaluator payoff* $U_E^{\text{base}}(b)$ *is* C^1*, strictly concave near* b^**, with a unique maximum at* b^**. The induced payoff* $U_E^{\text{info}}(b, x_b)$ *is continuous in both arguments and strictly increasing in* x_b*.*

Proof. Smoothness of $\tilde{\theta}(b)$ (Lemma 6), the Leibniz rule, $U_E''(b^*) < 0$ (Proposition 1), and Lemma 5. The marginal type exists and is unique by the intermediate value theorem applied to the strictly decreasing function $\frac{1}{2}m(\tilde{\theta})(b-\tilde{\theta})^2$. □

Lemma OA.3 (Properties of the Information Premium)**.** *The information premium* $\Lambda(b, x_b) = U_E^{\text{info}}(b, x_b) - U_E^{\text{base}}(b)$ *satisfies:* (a) $\Lambda(b, 0) = 0$*;* (b) $\Lambda > 0$ *for* $x_b > 0$*;* (c) $\partial\Lambda/\partial x_b > 0$*;* (d) $\partial^2\Lambda/\partial x_b^2 < 0$*;* (e) Λ *varies across benchmarks.*

Proof. (a) follows from $I^*(b, 0) = 0$, so $U_E^{\text{info}}(b, 0) = U_E^{\text{base}}(b)$. (b) and (c) follow from Lemma 5: $\partial U_E^{\text{info}}/\partial x_b > 0$ implies $\Lambda > 0$ for $x_b > 0$ and $\partial\Lambda/\partial x_b > 0$. (d) The noise technology $\sigma^2(I) = \sigma_0^2/(1+I)$ is strictly convex in I, so $d\sigma^2/dI = -\sigma_0^2/(1+I)^2$ is increasing (becoming less negative). The evaluator's payoff U_E is decreasing in σ^2 (lower noise improves screening) with bounded sensitivity, so $dU_E/dI = (dU_E/d\sigma^2)(d\sigma^2/dI)$, where the first factor is bounded and the second factor vanishes as $I \to \infty$. The diminishing marginal noise reduction generates diminishing marginal returns: $d^2U_E/dI^2 < 0$ under the curvature bound (14) from the proof of Lemma 4. Since $\Lambda(b, x_b) = U_E(b; r x_b/\kappa_I) - U_E(b; 0)$, $\partial^2\Lambda/\partial x_b^2 = (d^2U_E/dI^2)(r/\kappa_I)^2 < 0$. (e) follows because $\partial U_E/\partial I$ depends on the screening properties of b. □

OA.2 Further Extensions

This section relaxes perfect competition (Section OA.2.1), provides a risk-dominance micro-foundation (Section OA.2.2), and connects to global games (Section OA.2.3).

OA.2.1 Strategic Information Producers

Replacing (A3) with N symmetric strategic producers preserves the potential game structure that drives equilibrium selection. We work with a Cournot-style extension in which the price of a unit of information falls in aggregate supply, with $N \geq 2$ producers each maximizing

$$\pi_l = \left(r x_b - \kappa_I^C \sum_{j=1}^{N} i_j \right) i_l, \tag{6}$$

under linear inverse demand for information with slope parameter $\kappa_I^C > 0$. To distinguish from the quadratic-cost competitive parameter κ_I in the body, we use κ_I^C for the strategic slope. The two parameterizations describe distinct microfoundations (quadratic individual cost under price-taking versus linear inverse demand under quantity competition) but generate the same qualitative comparative statics for our purposes, namely that aggregate investment is linear in x_b and that the own-share property is preserved. The standard symmetric Cournot analysis applies.

Proposition OA.4 (Strategic Information Producers)**.** *Consider $N \geq 2$ symmetric Cournot information producers with payoffs (6). Then:*

(a) Symmetric Nash equilibrium. *Each producer invests*

$$i_N^* = \frac{r\, x_b}{\kappa_I^{\mathrm{C}}\, (N+1)}, \tag{7}$$

yielding aggregate investment

$$I_N^*(b, x_b) = \frac{N}{N+1} \cdot \frac{r\, x_b}{\kappa_I^{\mathrm{C}}}. \tag{8}$$

(b) Large-N limit. *As $N \to \infty$, $I_N^*(b, x_b) \to r\, x_b / \kappa_I^{\mathrm{C}}$, matching the form of the body model with κ_I^{C} playing the role of κ_I.*

(c) Weakened but qualitatively identical feedback loop. *For finite N, the feedback loop is dampened by the factor $N/(N+1)$, but remains strictly positive: $\partial I_N^* / \partial x_b = rN / [(N+1)\kappa_I^{\mathrm{C}}] > 0$.*

(d) Potential game structure is preserved. *The evaluator coordination game remains an exact potential game for any $N \geq 2$, and the stochastically stable benchmark is*

$$b_{\mathrm{SS}}^{(N)} = \arg\max_{b_k \in \mathcal{B}} \int_0^1 U_E^{\mathrm{info},N}(b_k, t)\, dt,$$

where $U_E^{\mathrm{info},N}(b_k, x_k) = U_E\big(b_k;\, rNx_k / [(N+1)\kappa_I^{\mathrm{C}}]\big)$.

Proof. (a) Differentiating (6) and imposing symmetry gives $rx_b = \kappa_I^{\mathrm{C}}(N+1)i$, so $i_N^* = rx_b/[\kappa_I^{\mathrm{C}}(N+1)]$ and aggregate investment $I_N^* = Ni_N^*$ satisfies (8). (b) $\lim_{N\to\infty} N/(N+1) = 1$. (c) $\partial I_N^*/\partial x_b > 0$ for all $N \geq 2$. (d) I_N^* depends only on x_k, so $U_E^{\mathrm{info},N}(b_k, x_k)$ depends only on x_k and the potential function Φ_N satisfies the Monderer–Shapley condition by the argument of Proposition 5. Stochastic stability then selects $\arg\max_b \int_0^1 U_E^{\mathrm{info},N}(b,t)\,dt$. □

The economic content is that market power affects the level of information investment but not the qualitative character of the feedback loop, because benchmarks partition the market into independent segments and strategic interaction within each segment does not create cross-segment linkages. The theory's predictions therefore apply both to concentrated information provision (credit ratings, with three dominant agencies) and to competitive provision (equity analyst coverage).

OA.2.2 Risk Dominance

Risk dominance (Harsanyi and Selten, 1988) identifies the strategy optimal under maximal uncertainty about others' choices. In the benchmark coordination game, risk dominance and stochastic stability select the same benchmark, providing an independent micro-foundation.

Setup. Consider the two-benchmark case $\mathcal{B} = \{b_1, b_2\}$. The evaluator's payoff from benchmark b_k when share x_k uses it is $U_E^{\mathrm{info}}(b_k, x_k)$. Both monomorphic states are Nash equilibria (Proposition 2).

Proposition OA.5 (Risk Dominance Selects the Stochastically Stable Benchmark)**.** *In the two-benchmark coordination game, define the* population-game risk-dominance criterion *as* $\Phi(\mathbf{e}_1) > \Phi(\mathbf{e}_2)$. *The path-integral form of the criterion is the population-game analog of Harsanyi-Selten risk dominance and corresponds to the potential-maximization selection criterion used in Sandholm (2010, Chapter 11) for population potential games. It extends the textbook basin-size comparison of Harsanyi and Selten (1988) to the strictly concave own-share payoffs that arise from the information loop.*

(*a*) *Benchmark* b_1 *has the larger basin of attraction under best-response dynamics if and only if* $x^*(b_1) < 1/2$, *where* $x^*(b_1)$ *is the unique solution to*

$$U_E^{\mathrm{info}}(b_1, x^*) = U_E^{\mathrm{info}}(b_2, 1 - x^*). \tag{9}$$

(*b*) *The population-game potential ranking is* $\Phi(\mathbf{e}_1) > \Phi(\mathbf{e}_2)$ *if and only if* $\int_0^1 h(t)\,dt > 0$, *where* $h(t) \equiv U_E^{\mathrm{info}}(b_1, t) - U_E^{\mathrm{info}}(b_2, 1 - t)$.

(*c*) *Therefore the population-game risk-dominant benchmark is exactly the stochastically stable benchmark:* $b_{\mathrm{RD}} = b_{\mathrm{SS}}$.

(*d*) *The basin-size criterion of Part (a) and the potential-integral criterion of Part (b) coincide whenever* $h(t) + h(1-t)$ *has constant sign on* $[0, 1/2]$*, which holds in particular when* $U_E^{\text{info}}(b_k, t)$ *is affine in* t *(the textbook* 2×2 *case of Blume 1993). The model's noise technology* $\sigma^2(I) = \sigma_0^2/(1+I)$ *delivers* U_E^{info} *that is strictly concave in own share, not affine, so the textbook equivalence between basin size and potential maximization need not hold pointwise. The selection result that drives Theorem 1 is the potential-integral criterion in Part (b), and we adopt this as the definition of risk dominance in the population potential game throughout.*

Proof. Part (a): Basin of attraction characterization.

Under best-response dynamics, an evaluator switches from b_2 to b_1 when $U_E^{\text{info}}(b_1, x_1) > U_E^{\text{info}}(b_2, 1-x_1)$. Define $h(x) \equiv U_E^{\text{info}}(b_1, x) - U_E^{\text{info}}(b_2, 1-x)$. By Lemma 5, $U_E^{\text{info}}(b_k, \cdot)$ is strictly increasing in own share, so h is strictly increasing. Since $h(0) = U_E^{\text{base}}(b_1) - U_E^{\text{info}}(b_2, 1) < 0$ and $h(1) = U_E^{\text{info}}(b_1, 1) - U_E^{\text{base}}(b_2) > 0$, there exists a unique crossing $x^* \in (0, 1)$ with $h(x^*) = 0$. The basin of attraction of b_1 under best-response dynamics is $\{x_1 \in [0, 1] : x_1 > x^*\}$, larger than that of b_2 when $x^* < 1/2$.

Part (b): Potential ranking via the integral of h.

The potential at the monomorphic states is $\Phi(\mathbf{e}_k) = \int_0^1 U_E^{\text{info}}(b_k, t)\, dt$. By a change of variable in the second integral,

$$\Phi(\mathbf{e}_1) - \Phi(\mathbf{e}_2) = \int_0^1 \left[U_E^{\text{info}}(b_1, t) - U_E^{\text{info}}(b_2, 1-t)\right] dt = \int_0^1 h(t)\, dt,$$

since $\int_0^1 U_E^{\text{info}}(b_2, 1-t)\, dt = \int_0^1 U_E^{\text{info}}(b_2, u)\, du$ by $u = 1-t$. Hence $\Phi(\mathbf{e}_1) > \Phi(\mathbf{e}_2)$ if and only if $\int_0^1 h(t)\, dt > 0$.

Part (c): Identification.

Part (b) shows that the population-game risk-dominance comparison $\Phi(\mathbf{e}_1) > \Phi(\mathbf{e}_2)$ reduces to $\int_0^1 h(t)\, dt > 0$, the path-integral form used by Sandholm (2010, Chapter 11) to characterize population risk dominance in potential games. By Theorem 1, the stochastically stable state under log-linear response is the global maximizer of the same potential (Blume, 1993). The two selection arguments thus coincide on the path integral, so $b_{\text{RD}} = b_{\text{SS}}$. The substantive content is that the path-integrated comparison, rather than basin size, is the criterion that aligns with stochastic stability under the model's strictly concave own-share payoffs.

Part (d): When the population criterion reduces to basin size.

Folding the integral around $t = 1/2$ gives $\int_0^1 h(t)\, dt = \int_0^{1/2} [h(t) + h(1-t)]\, dt$. When $h(t) + h(1-t)$ has constant sign on $[0, 1/2]$ matching the sign of $h(1/2)$, the integral has the same sign as $h(1/2)$, which by strict monotonicity of h equals the sign of $1/2 - x^*$. So $\Phi(\mathbf{e}_1) > \Phi(\mathbf{e}_2) \Leftrightarrow x^* < 1/2$ under this symmetry. A sufficient condition is that $U_E^{\text{info}}(b_k, t)$ be affine in t, the textbook 2×2 case of Blume (1993). The condition fails under the model's noise technology, where U_E^{info} is strictly concave in own share by Lemma OA.3(d). The population-game criterion in Parts (b)–(c) is therefore strictly more general than the basin-size criterion, and it is the criterion that delivers the equivalence with stochastic stability. □

The equivalence in Proposition OA.5 has a natural economic interpretation. Under maximal uncertainty about adoption, the population-game risk-dominant benchmark is the one that generates the strongest information ecosystem across the widest range of adoption levels, exactly the maximizer of the path integral $\int_0^1 U_E^{\text{info}}(b_k, t)\, dt$. Part (d) emphasizes that the textbook basin-size criterion of Harsanyi and Selten (1988) can disagree with the path-integral criterion when own-share payoffs are strictly concave, and the path-integral criterion is the operative one for selection in this paper.

OA.2.3 Global Games Sketch

The third selection criterion is a global-games argument in the spirit of Carlsson and van Damme (1993) as extended by Frankel, Morris, and Pauzner (2003) to state-dependent payoffs. This subsection sketches how the argument applies in our setting and confirms that the equilibrium selected by the global game in the noise-vanishing limit coincides with the stochastically stable and risk-dominant benchmark.

The embedded global game. Consider the two-benchmark special case with candidates b_1 and b_2. Each evaluator j receives a private signal $y_j = q + \nu\eta_j$ about a fundamental $q \in \mathbb{R}$ that shifts the relative attractiveness of b_2, where $\eta_j \sim N(0, 1)$ are independent across evaluators and $\nu > 0$ is the signal noise. The fundamental q has a diffuse improper prior. The evaluator's payoff from choosing b_k when the share on b_2 is x_2 is $V_k(x_2; q) = U_E^{\text{base}}(b_k) + \Lambda(b_k, x_k) + q \cdot \mathbb{1}\{k = 2\}$, with $x_1 = 1 - x_2$. The parameter q enters as a state-dependent shift that activates dominance regions at the tails of its distribution.

Verifying the Frankel-Morris-Pauzner conditions. Three conditions characterize a well-behaved global game with state-dependent payoffs in Frankel, Morris, and Pauzner (2003).

Action monotonicity (strategic complementarities). The relative payoff to choosing b_2 over b_1, namely $V_2 - V_1 = [U_E^{\text{base}}(b_2) - U_E^{\text{base}}(b_1)] + [\Lambda(b_2, x_2) - \Lambda(b_1, 1 - x_2)] + q$, is strictly increasing in x_2 because Λ is strictly increasing in own share (Lemma 5) while $\Lambda(b_1, 1 - x_2)$ is strictly decreasing in x_2. The game has strategic complementarities in the population action.

State monotonicity. The relative payoff $V_2 - V_1$ is strictly increasing in q by construction.

Dominance regions. For q sufficiently large, $V_2 > V_1$ at every $x_2 \in [0, 1]$, so b_2 is dominant. For q sufficiently negative, b_1 is dominant. Both dominance regions have positive probability under any prior with full support, satisfying the FMP existence condition.

These conditions imply, by Proposition 1 of Frankel, Morris, and Pauzner (2003), the existence of a unique monotone strategy equilibrium for each finite $\nu > 0$, in which evaluators with private signals above a cutoff $y^*(\nu)$ play b_2 and the others play b_1.

The noise-vanishing limit. As $\nu \to 0$, the cutoff signal converges to a state q^* at which the population is exactly indifferent. By Proposition 3 of Frankel, Morris, and Pauzner (2003), the limit selects the strategy that is a best response at the uniform belief over the opponent action, the criterion of risk dominance (Harsanyi and Selten, 1988). In the two-benchmark population game considered here, the risk-dominant benchmark maximizes the path-integrated payoff

$$\Phi(\mathbf{e}_k) = U_E^{\text{base}}(b_k) + \int_0^1 \Lambda(b_k, \rho)\, d\rho, \tag{10}$$

which is the same potential that drives the stochastic-stability selection in Theorem 1 and the risk-dominance selection in Proposition OA.5 of this Online Appendix.

Vanishing correction. For finite signal noise $\nu > 0$, the global-game cutoff $y^*(\nu)$ generally differs from q^* by a correction term $c(\nu) \equiv y^*(\nu) - q^*$ that depends on the prior, the noise distribution, and the curvature of the payoff difference around the indifference point. Under the regularity conditions of Frankel, Morris, and Pauzner (2003), $c(\nu) \to 0$ as $\nu \to 0$, so the cutoff converges to the risk-dominance threshold and the selected benchmark coincides with the stochastically stable one.

Beyond two benchmarks. Frankel, Morris, and Pauzner (2003) extend the global-games selection result to multi-action games with strategic complementarities provided the payoff structure admits a one-dimensional aggregator and dominance regions exist for each pairwise comparison. Verifying these conditions across $K \geq 3$ benchmarks in our setting requires a separate analysis we do not undertake here. The two-benchmark sketch above is intended as a sanity check that the noise-vanishing limit of an embedded global game yields the same selection as stochastic stability and population-game risk dominance in the simplest comparison. The full K-benchmark extension is left for future work.

OA.2.4 Information-Theoretic Foundation for Benchmark Complexity

Assumption 4 introduces benchmark complexity $\gamma(b)$ as a primitive. This section provides a microfoundation via Fisher information.

Setup. An information producer observes a raw signal z about a reporter's type θ, drawn from a family $\{f(z \mid \theta)\}_{\theta\in[0,1]}$ satisfying standard regularity conditions (differentiable in θ, common support). Given benchmark b, the producer's task is to refine z into a signal s that is informative about the event $\{\theta \geq b\}$. The cost of achieving signal noise σ^2 at benchmark b depends on the statistical difficulty of distinguishing types near b in the raw data.

Remark OA.6 (Fisher Information Microfoundation for $\gamma(b)$)**.** Define the Fisher information of the raw signal at the decision boundary:

$$\mathcal{I}_F(b) \equiv \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(z \mid \theta)\right)^2 \,\middle|\, \theta = b\right]. \tag{11}$$

By the Cramér–Rao bound, any unbiased estimator of θ at $\theta = b$ has variance at least $1/\mathcal{I}_F(b)$. The minimum cost of producing a signal with noise σ^2 at benchmark b is therefore proportional to $1/\mathcal{I}_F(b)$: more Fisher information at the boundary means a cheaper refinement problem. Setting

$$\gamma(b) = \frac{c_0}{\mathcal{I}_F(b)} \tag{12}$$

for a constant $c_0 > 0$ yields the information production cost $\kappa_I(b) = \kappa_0 + (c_0\,\kappa_1)/\mathcal{I}_F(b)$ from Assumption 4: benchmarks where the raw signal is less informative about the boundary crossing are costlier to evaluate.

Economic interpretation. A zero-earnings benchmark has high $\mathcal{I}_F$, because the raw accounting signal is highly informative about whether $\theta \geq 0$, since the income statement directly reveals the sign of earnings. An analyst-forecast benchmark has low $\mathcal{I}_F$, because determining whether a firm beats the consensus requires aggregating heterogeneous analyst opinions, tracking forecast revisions, and conditioning on firm guidance, all of which add noise to the raw signal at the decision boundary. This maps to $\gamma(\text{zero}) \approx 0$ and $\gamma(\text{forecast}) > 0$, exactly as Assumption 4 specifies. More generally, any benchmark whose decision boundary falls in a region of high signal-to-noise in the raw data will have low complexity, and benchmarks whose boundaries require synthesizing multiple noisy sources will have high complexity.

OA.2.5 Operational Flexibility and Manipulation Costs

Assumption 1 states that higher-type reporters face lower manipulation costs, $m'(\theta) < 0$. This section derives the condition from a primitive model of operational flexibility.

Setup. A reporter with type θ has access to a set of manipulation technologies $\mathcal{A}(\theta) \subseteq \mathbb{R}_+$. Each technology $z \in \mathcal{A}(\theta)$ has a direct cost $c(z) > 0$ and produces manipulation output $g(z) > 0$. The reporter's unit cost of manipulation (matching the quadratic specification in the body Section 2.2) is the solution to the problem

$$m(\theta) = \min_{z \in \mathcal{A}(\theta)} \frac{c(z)}{g(z)^2}. \tag{13}$$

Remark OA.7 (Deriving $m'(\theta) < 0$ from Expanding Technology Sets)**.** Suppose $\mathcal{A}(\theta)$ is expanding in θ in the strong set order: $\theta' > \theta$ implies $\mathcal{A}(\theta) \subseteq \mathcal{A}(\theta')$. Then $m(\theta)$ is weakly decreasing in θ. If the expansion is strict—for each $\theta' > \theta$, there exists $z' \in \mathcal{A}(\theta') \setminus \mathcal{A}(\theta)$ with $c(z')/g(z')^2 < m(\theta)$—then $m(\theta)$ is strictly decreasing: $m'(\theta) < 0$.

Proof. The objective $c(z)/g(z)^2$ is independent of θ; only the constraint set $\mathcal{A}(\theta)$ varies. By Berge's theorem of the maximum, the value function of a minimization problem is weakly decreasing when the constraint set expands in the inclusion order: $\mathcal{A}(\theta) \subseteq \mathcal{A}(\theta')$ implies $m(\theta') \leq m(\theta)$, since every feasible point for the smaller problem is feasible for the larger one. Strict decrease follows when the expansion is strict: the new technology $z' \in \mathcal{A}(\theta') \setminus \mathcal{A}(\theta)$ achieves a strictly lower objective value than the optimum over $\mathcal{A}(\theta)$, so $m(\theta') < m(\theta)$. □

Economic interpretation. A firm with higher unmanaged earnings θ has more degrees of operational freedom, including discretion over accruals, timing of asset sales, acceleration or deferral of revenue recognition, and real operational adjustments (e.g., cutting discretionary spending). Each such instrument is a manipulation technology z. The set $\mathcal{A}(\theta)$ expands with θ because financial health relaxes constraints, since a profitable firm can time a gain-producing asset sale that a distressed firm cannot, or exercise accrual discretion that a firm near covenant violation cannot. The strict expansion condition requires only that each increment in θ opens at least one new cost-effective manipulation channel, which is natural, because marginal improvements in financial health unlock marginal operational flexibility. The result is $m'(\theta) < 0$, exactly Assumption 1.

OA.3 Detailed Proofs of Body Results

OA.3.1 Proof of Proposition 4 (Information Concentration)

Proof. Part (a): Noise comparison. The noise technology $\sigma^2(I) = \sigma_0^2/(1+I)$ has $\sigma^{2\prime\prime}(I) = 2\sigma_0^2/(1+I)^3 > 0$, so σ^2 is strictly convex and strictly decreasing on $\mathbb{R}_+$. The proposition compares the decision-relevant noise under two information allocation rules with a common total budget $\int \eta \, ds = \bar{I}$.

Threshold rule. Under the threshold rule $d(s) = \mathbb{1}\{s \geq b\}$, all investment is decision-relevant only at $s = b$. Competitive producers therefore concentrate the budget on benchmark-relative signals (Section 2.3 of the body), and the relevant noise governing the evaluator's classification at the threshold is the noise of the signal that pins down whether $s \geq b$. By Lemma 1, this is $\sigma^2(\bar{I}) = \sigma_0^2/(1+\bar{I})$, the noise of a single benchmark-relative topic that absorbs the entire budget. We denote this object the *threshold-rule noise* and emphasize that it is not the pointwise noise at a Dirac mass (which would not be well-defined), but the noise of the single signal the threshold rule uses.

Continuous rule. Under a continuous Lipschitz rule p with $p'(s) > 0$ on a set of positive measure, decisions depend on s over an interval rather than at a single point. The marginal value of information at s is proportional to $|p'(s)|g(s)$ (where g is the density of s), so competitive equilibrium investment satisfies $\eta(s) = \bar{I} \cdot |p'(s)|g(s)/\int |p'(s')|g(s')\,ds'$, which is non-constant on its support and integrates to $\bar{I}$. The decision-relevant noise under the continuous rule is the average noise weighted by the decision derivative,

$$\bar{\sigma}^2_{\text{cont}} = \frac{\int \sigma^2(\eta(s))|p'(s)|g(s)\,ds}{\int |p'(s')|g(s')\,ds'},$$

which is finite because η is bounded below by zero and σ^2 is decreasing in I from $\sigma^2(0) = \sigma_0^2$.

Jensen comparison. Let $\omega(s) = |p'(s)|g(s)/\int |p'|g$ be the decision weight, a probability density with $\int \eta\,\omega = \bar{I}$ by construction (since $\eta \propto \omega$ with proportionality constant $\bar{I}$). The two rules thus allocate the same total budget but the continuous rule spreads it across a non-constant distribution over s. Strict convexity of σ^2 and Jensen's inequality give

$$\bar{\sigma}^2_{\text{cont}} = \int \sigma^2(\eta(s))\,\omega(s)\,ds > \sigma^2\left(\int \eta(s)\,\omega(s)\,ds\right) = \sigma^2(\bar{I}),$$

with strict inequality because η is not constant on the support of ω. The threshold-rule noise is strictly lower than the continuous-rule noise at equal information budget.

Part (b): Screening payoff dominance. The evaluator's payoff under rule $d(\cdot)$ with allocation $\eta(\cdot)$ is $U_E(d,\eta) = \int_0^1 (\theta - \kappa)\mathbb{E}_\varepsilon[d(\theta + a^* + \varepsilon)]f(\theta)\,d\theta$. *Step 1.* Under the threshold rule with concentrated $\bar{I}$ at b, noise is $\sigma^2(\bar{I})$, the passing probability is $\Phi((\theta + a - b)/\sigma(\bar{I}))$, and the payoff is strictly decreasing in $\sigma^2(\bar{I})$ by single-crossing: $\theta - \kappa$ and $\partial\Phi/\partial\sigma$ have opposite signs on each side of κ. *Step 2.* Under the continuous rule the screening payoff depends on noise at every point in the transition region $[s_L, s_H]$, and by Part (a) the average noise there strictly exceeds $\sigma^2(\bar{I})$. *Step 3.* Even at equal noise, intermediate acceptance probabilities $p(s) \in (0,1)$ create a strict additional loss because the evaluator's optimal rule conditional on s is $d(s) = \mathbb{1}\{\mathbb{E}[\theta|s] \geq \kappa\}$ by affiliation: for any s with $\mathbb{E}[\theta|s] > \kappa$, accepting with probability $p(s) < 1$ forfeits $(\mathbb{E}[\theta|s] - \kappa)(1 - p(s)) > 0$, and symmetrically below. Combining the three steps, $U_E^{\text{thresh}}(b;\bar{I}) > U_E^{\text{cont}}(p;\eta)$. □

This section collects the full proofs of nine body-paper results whose statements remain in the main paper. The body retains a one-paragraph sketch for each result. The proofs below provide the complete technical detail. Cross-references to body equations and lemmas use the labels established in the main paper, and when a body proof references an equation we relocate here, the label is preserved so the body sketch can cite it.

OA.3.2 Proof of Lemma 1 (Information Production Equilibrium)

Proof. Each information producer l maximizes profits:

$$\pi_I(i_l, x_b) = r\, x_b\, i_l - \frac{1}{2}\,\kappa_I\, i_l^2.$$

The objective is strictly concave in i_l (since $-\kappa_I < 0$). The first-order condition is

$$r\, x_b - \kappa_I\, i_l = 0 \qquad \Longrightarrow \qquad i_l^* = \frac{r\, x_b}{\kappa_I}.$$

Since all producers are identical (A3) and there is a unit mass of producers, aggregate information investment is

$$I^*(b, x_b) = \int_0^1 i_l^*\, dl = \frac{r\, x_b}{\kappa_I}.$$

The linearity of I^* in x_b follows from the linearity of revenue in i_l and x_b combined with the quadratic cost structure: this is a defining feature of identical, competitive producers with quadratic costs and linear demand. □

OA.3.3 Proof of Lemma 2 (Optimal Manipulation)

Proof. The reporter maximizes $\pi_R = R(\theta + a, b) - \frac{1}{2}m(\theta)a^2$ with R the step-function reward. If $\theta \geq b$, manipulation is strictly dominated by $a = 0$ and $a^* = 0$. If $\theta < b$, any $a \in (0, b - \theta)$ is dominated by $a = 0$ (same reward, strictly positive cost) and any $a > b - \theta$ by $a = b - \theta$ (same reward, higher cost), so the optimum is bang-bang between $a = 0$ and $a = b - \theta$. The latter is preferred iff $\phi(\theta) \equiv \frac{1}{2}m(\theta)(b - \theta)^2 \leq \Delta u$. Differentiating, $\phi'(\theta) = (b - \theta)[\frac{1}{2}m'(\theta)(b - \theta) - m(\theta)] < 0$ under (A1), so ϕ is strictly decreasing on $[0, b)$ with $\phi(b^-) = 0$. By the IVT, either every $\theta \in [0, b)$ manipulates ($\tilde{\theta}(b) = 0$) or there is a unique $\tilde{\theta}(b) \in (0, b)$ with $\phi(\tilde{\theta}(b)) = \Delta u$. Under (A2), $\phi(0) > \Delta u$ for $b \geq \kappa$, so the marginal type is interior. The optimum is $a^*(\theta, b) = (b - \theta) \cdot \mathbb{1}\{\theta \in [\tilde{\theta}(b), b)\}$. □

OA.3.4 Proof of Lemma 3 (Bunching Distribution)

Proof. Under the equilibrium strategy of Lemma 2, types in $[0, \tilde{\theta}(b))$ produce $s = \theta$, types in $[\tilde{\theta}(b), b)$ all produce $s = b$, and types in $[b, 1]$ produce $s = \theta$. The CDF H of reported signals is therefore $F(x)$ for $x < \tilde{\theta}(b)$, constant at $F(\tilde{\theta}(b))$ on $[\tilde{\theta}(b), b)$ (the gap), jumps by $F(b) - F(\tilde{\theta}(b)) > 0$ at $x = b$ (the mass point), and equals $F(x)$ for $x \geq b$, establishing parts (a)–(d). $\square$

OA.3.5 Proof of Lemma 4 (Manipulation Under Endogenous Noise)

Proof. With $\sigma \equiv \sigma(I) > 0$ and $z(a) \equiv (\theta + a - b)/\sigma$, the reporter maximizes $V(\theta, a) = \Delta u\, \Phi(z(a)) + u_L - \frac{1}{2}m(\theta)a^2$. The FOC is $(\Delta u/\sigma)\phi(z) = m(\theta)a$, and the SOC is $\partial^2 V/\partial a^2 = -(\Delta u/\sigma^2)z\phi(z) - m(\theta)$, which is negative for $z \geq 0$ unconditionally and for $z < 0$ whenever $m(\theta) > (\Delta u/\sigma^2)|z|\phi(z)$. The latter is implied by (A2) for parameter values that bound the manipulation incentive away from the cost: $\sup_{z<0}(\Delta u/\sigma^2)|z|\phi(z) = (\Delta u/\sigma^2)\cdot\phi(0)/\sqrt{e}$, which is less than $m(\theta)$ in the empirically relevant region. For $z < 0$ with $m(\theta) < (\Delta u/\sigma^2)|z|\phi(z)$, the SOC can fail, and global optimality requires comparing the interior stationary point of the FOC against the boundary value $a = 0$. The reporter's payoff at $a = 0$ is $\Delta u\, \Phi((\theta - b)/\sigma) + u_L$, and the stationary a^* is preferred whenever it delivers higher expected payoff. The maximization is therefore well-defined as $a^* = \arg\max\{V(\theta, a^*_{\text{FOC}}), V(\theta, 0)\}$, with the FOC delivering the interior maximizer on the region where the SOC holds and the boundary maximizer prevailing elsewhere. For types in the relevant range $\theta \geq \kappa$ (the types that drive the evaluator's screening), the SOC holds under (A2). All downstream uses of Lemma 4 concern these types.

Part (a). At $a = 0$, marginal benefit is $(\Delta u/\sigma)\phi((\theta - b)/\sigma) > 0$ and marginal cost is zero, so $a^* > 0$ for every type. For $\theta \gg b$ the benefit is exponentially small in $(\theta - b)^2/\sigma^2$, so a^* is negligible.

Part (b). By IFT on the FOC, $\partial a^*/\partial\theta = -(\partial H/\partial\theta)/(\partial H/\partial a)$ where $H \equiv (\Delta u/\sigma)\phi(z) - m(\theta)a$ and $\partial H/\partial a < 0$ by SOC. The numerator combines $-(\Delta u/\sigma^2)z\phi(z)$ with $-m'(\theta)a > 0$ (since $m' < 0$). For types near b, $m(\theta) > |m'(\theta)|a$ since $a \leq b - \theta$, so $|da^*/d\theta| < 1$ and $d(\theta + a^*)/d\theta > 0$. Under condition (14) this extends to all types. In the deterministic limit, $a^* = b - \theta$ on $[\tilde{\theta}(b), b)$ is strictly decreasing.

Part (c). Using $\phi'(z) = -z\phi(z)$ and $\partial z/\partial\sigma = -z/\sigma$, $\partial H/\partial\sigma = (\Delta u/\sigma^2)\phi(z)(z^2 - 1)$. With $\partial H/\partial a < 0$, $\partial a^*/\partial\sigma < 0$ when $|z^*| < 1$ (lower noise increases manipulation for types near b) and $\partial a^*/\partial\sigma > 0$ when $|z^*| > 1$ (lower noise reduces manipulation for types far from b).

Part (d). As $\sigma \to 0$, $\Phi(z) \to \mathbb{1}\{\theta + a \geq b\}$ and the smooth FOC reduces to the bang-bang solution of Lemma 2.

Auxiliary condition for Lemma 5. The manipulation response $\partial a^*/\partial I$ is uniformly bounded provided

$$\sup_{\theta \in [0,\bar{b}]} |m''(\theta)/m'(\theta)| < \infty. \tag{14}$$

□

OA.3.6 Sign Compatibility in the Indirect Channel of Lemma 5

The body proof of Lemma 5 (Appendix B.3) controls the indirect channel on $R_1 \cup R_3$ by a magnitude bound. The integrand is not uniformly signed there because $\partial a^*/\partial \sigma$ flips sign at $|z^*| = 1$ (Lemma 4(c)). The argument below resolves the sign question directly by sub-partitioning.

Let $z^*(\theta) \equiv (\theta + a^* - b)/\sigma$. Partition the high-share region by closeness to the threshold,

$$R_k^a \equiv R_k \cap \{|z^*| < 1\}, \qquad R_k^b \equiv R_k \cap \{|z^*| \geq 1\}, \qquad k \in \{1, 3\}. \tag{15}$$

Write the integrand of dU_E/dI as $D(\theta) + J(\theta)$, where D is the direct term signed as in the body proof and $J(\theta) \equiv (\theta - \kappa)\,\phi(z^*)\,(\partial a^*/\partial I)\,f(\theta)/\sigma$.

Aligned pieces. On R_k^a, Lemma 4(c) gives $\partial a^*/\partial I > 0$. Lower noise raises equilibrium manipulation for these near-threshold types, which improves the conditional pool quality through the direct channel and adds the same-direction contribution through the indirect channel. The body proof shows $D > 0$ on R_k, and on R_k^a the indirect contribution to $\partial U_E/\partial I$ is bounded in magnitude by D under (14), so $D + J \geq 0$ on $R_1^a \cup R_3^a$ with strict inequality except on a measure-zero set.

Threshold-far pieces. On R_k^b, Lemma 4(c) gives $\partial a^*/\partial I < 0$, so J and D carry opposite signs. The Gaussian density satisfies $\phi(z^*) \leq \phi(1)\, e^{-(z^{*2}-1)/2}$ for $|z^*| \geq 1$. The direct integrand contains the factor $z^*\phi(z^*)$ through $\psi = -(\theta + a^* - b)\phi(z^*)/(2\sigma^3)$, while J contains $\phi(z^*)$ alone. Hence $|D|/|J| \geq |z^*| \cdot C^{-1}$ on R_k^b with C depending only on m, Δu, and (14). For $|z^*| \geq 1$, $|D| \geq |J|$ (rescaling C if needed by enlarging $\bar{I}$ to compress σ further), and the net $D + J$ keeps the sign of D.

Conclusion. Combining the four sub-regions, $D + J \geq 0$ pointwise on $R_1 \cup R_3$, strictly so on a set of positive measure inside R_1^a. The contribution from $R_1 \cup R_3$ to dU_E/dI is at least the contribution from D alone, which the body proof shows dominates the R_2 piece at high shares. Hence $\partial U_E^{\text{info}}/\partial x_b > 0$ on $[x_*, 1]$ without any uniform-sign assumption on J. The body magnitude bound treats the worst case, while the sub-partition shows that the worst case does not bind. J reinforces D on the aligned sub-regions, and Gaussian decay dominates J on the threshold-far sub-regions.

OA.3.7 Proof of Proposition 2 (Multiple Equilibria)

Proof. At a monomorphic equilibrium with benchmark $\hat{b}$, an evaluator using $\hat{b}$ earns $U_E^{\text{base}}(\hat{b}) + \Lambda(\hat{b}, 1)$. A unilateral deviator to any $b' \neq \hat{b}$ earns $U_E^{\text{base}}(b')$ since $\Lambda(b', 0) = 0$.

The no-deviation condition is therefore $U_E^{\text{base}}(\hat{b}) + \Lambda(\hat{b}, 1) \geq U_E^{\text{base}}(b')$ for all $b' \neq \hat{b}$, equivalently $\Lambda(\hat{b}, 1) \geq \sup_{b' \neq \hat{b}} U_E^{\text{base}}(b') - U_E^{\text{base}}(\hat{b})$. Since U_E^{base} attains its global maximum at b^* (Proposition 1), $\sup_{b'} U_E^{\text{base}}(b') = U_E^{\text{base}}(b^*)$, and the binding deviation is to b^*. The no-deviation condition reduces to $\Lambda(\hat{b}, 1) \geq D(\hat{b}) \equiv U_E^{\text{base}}(b^*) - U_E^{\text{base}}(\hat{b})$. The sustainable set $\mathcal{S} = \{b : \Lambda(b, 1) \geq D(b)\}$ contains b^* since $\Lambda(b^*, 1) > 0 = D(b^*)$, and is an interval by continuity. A second-order expansion $D(\hat{b}) \approx \frac{1}{2}|U_E''(b^*)|(\hat{b} - b^*)^2$ combined with $\Lambda(\hat{b}, 1) \approx \Lambda(b^*, 1)$ yields the bounds in (12). □

OA.3.8 Proof of Proposition 6 (Goodhart's Law Moderation)

Proof. (a) Manipulation raises the passing probability for $\theta \in [\tilde{\theta}(b_{\text{SS}}), b_{\text{SS}})$, who have θ below the no-manipulation acceptance pool's mean, so $\mathbb{E}[\theta \mid s \geq b, a = 0] > \mathbb{E}[\theta \mid s \geq b, a^*]$, and symmetrically below. Hence $\mathcal{I}^{\text{no-manip}}(b_{\text{SS}}, 1) > \mathcal{I}(b_{\text{SS}}, 1)$. (b) By Lemma 5, lower noise raises $\mathbb{E}[\theta \mid s \geq b]$ and lowers $\mathbb{E}[\theta \mid s < b]$, so $\partial\mathcal{I}/\partial I > 0$ and $\mathcal{I}(b_{\text{SS}}, 1) > \mathcal{I}(b_{\text{SS}}, 0)$. (c) Under (A1), higher types face lower manipulation costs so the manipulator pool is positively selected, giving $\mathcal{I}(b_{\text{SS}}, 0) > 0$ and hence $\mathcal{I}(b_{\text{SS}}, 1) > 0$. □

OA.3.9 Proof of Proposition 7 (Benchmark Transitions)

Proof. (a) Let $\Delta\Phi(\kappa_1) \equiv \Phi(\mathbf{e}_c; \kappa_1) - \Phi(\mathbf{e}_s; \kappa_1)$. As $\kappa_1 \to \infty$, $I^*(b_c, \rho) \to 0$ and $\Phi(\mathbf{e}_c) \to U_E^{\text{base}}(b_c) < U_E^{\text{base}}(b_s)$, so $\Delta\Phi < 0$. As $\kappa_1 \to 0$, the informational-responsiveness condition $\int_0^1 (\partial U_E/\partial I)(b_c; r\rho/\kappa_0)\, d\rho > \int_0^1 (\partial U_E/\partial I)(b_s; r\rho/\kappa_0)\, d\rho$ gives $\Delta\Phi > 0$. By the IVT there exists κ_1^* with $\Delta\Phi(\kappa_1^*) = 0$. (b) The transition is a discrete jump in the identity of the stochastically stable monomorphic state at κ_1^*, and $d\Delta\Phi/d\kappa_1$ is generically nonzero there since $\gamma(b_c) > \gamma(b_s)$ makes the complex benchmark's investment more sensitive to κ_1, confirming transversal crossing. □

OA.3.10 Proof of Proposition 8 (ESS vs. Social Optimum)

Proof. Social welfare (Definition 1) is:

$$W(b, x_b) = U_E^{\text{info}}(b, x_b) + W_R^{\text{gross}}(b, x_b) - TC(b, x_b) - IC(b, x_b),$$

where $W_R^{\text{gross}}(b, x_b) = \int_0^1 \mathbb{E}[R(s, b) \mid \theta]\, f(\theta)\, d\theta$ is the reporters' gross expected reward (before manipulation costs), TC is the aggregate manipulation cost, and $IC = \frac{1}{2}\kappa_I (I^*)^2$ is the real resource cost of information production. The producer revenue $\Pi_I = r\, x_b\, I^*$ is a transfer paid by evaluators and received by producers, which nets to zero in the social aggregate. The evaluator screening payoff U_E^{info} is taken gross of this payment, so subtracting it from evaluators and adding it back to producers leaves the four-term expression above.

Part (a): b_{SS} does not generally maximize welfare. The stochastically stable benchmark maximizes $\int_0^1 U_E^{\text{info}}(b_k, \rho)\, d\rho$ (Theorem 1), a path integral of evaluator payoff only. Social welfare W additionally includes reporter surplus ($W_R^{\text{gross}} - TC$) and subtracts information production costs (IC). These additional terms depend on b through the manipulation equilibrium and the information production technology. Generically, the maximizer of $\int_0^1 U_E^{\text{info}}\, d\rho$ differs from the maximizer of W.

Part (b): Direction of the wedge.

(i) Manipulation externality. The evaluator picks her privately optimal $b^* > \kappa$ by balancing her own screening gains and losses, but she does not internalize the manipulation costs reporters bear. The social planner's optimum b^W therefore sits below the evaluator's choice b_{SS}, because b^W also subtracts reporter manipulation costs from the social objective. The wedge is the manipulation externality.

(ii) Information externality. Information producers capture revenue $r\, x_b\, i_l$ but generate social value $\partial U_E/\partial I$ per unit of investment. If the social value exceeds the private revenue, there is underinvestment in information. However, the evolutionary dynamics partially correct this, because the selection criterion $\int_0^1 U_E^{\text{info}}(b_k, \rho)\, d\rho$ accounts for the screening improvement from information.

Part (c): $W(b^W, 1) \geq W(b_{\text{SS}}, 1)$. This is immediate from the definition: $b^W = \arg\max_b W(b, 1)$, so $W(b^W, 1) \geq W(b, 1)$ for any b, including $b = b_{\text{SS}}$.

Welfare ratio across parameters. Table 1 reports $W(b_{\text{SS}}, 1)/W(b^W, 1)$ across a grid of values for r/κ_I, Δu, and $m(\kappa)$, under $F \sim U[0, 1]$ and $\sigma_0 = 0.3$. The ratio is increasing in r/κ_I (cheaper information closes the gap), decreasing in Δu (stronger manipulation incentives widen it), and increasing in $m(\kappa)$ (costlier manipulation reduces distortion).

Table 1: Welfare ratio $W(b_{\text{SS}}, 1)/W(b^W, 1)$ across parameter values.

	$r/\kappa_I = 0.5$	$r/\kappa_I = 2$	$r/\kappa_I = 5$
$\Delta u = 0.5$	0.82	0.91	0.95
$\Delta u = 1.0$ (baseline)	0.78	0.89	0.94
$\Delta u = 2.0$	0.72	0.85	0.92
$m(\kappa)$ low	0.74	0.86	0.93
$m(\kappa)$ baseline	0.78	0.89	0.94
$m(\kappa)$ high	0.83	0.92	0.96

Part (d): Comparison with no benchmark. Without a benchmark ($b = 0$), all reporters produce $s = \theta \geq 0 = b$, so all pass trivially. The evaluator accepts everyone, earning $\int_0^1 (\theta - \kappa)\, f(\theta)\, d\theta$. All reporters receive u_H. There is no manipulation and no information production. Social welfare is:

$$W(\text{no benchmark}) = \int_0^1 (\theta - \kappa)\, f(\theta)\, d\theta + u_H.$$

With benchmark b_{SS} at full adoption:

$$W(b_{\text{SS}}, 1) = U_E^{\text{info}}(b_{\text{SS}}, 1) + W_R^{\text{gross}}(b_{\text{SS}}, 1) - TC(b_{\text{SS}}, 1) + \Pi_I(b_{\text{SS}}, 1) - IC(b_{\text{SS}}, 1).$$

The screening benefit of having a benchmark is the ability to reject types $\theta < \kappa$. This benefit equals $\int_0^\kappa (\kappa - \theta)\, f(\theta)\, d\theta$ in the perfect-screening limit. The benchmark is socially valuable when the screening benefit dominates the manipulation cost, information production cost, and foregone reporter rewards for rejected reporters. □

OA.3.11 Proof of Theorem 2 (Continuous-Space Stochastic Stability)

Proof. Part (a): Convergence of discrete selection.

Let $\mathcal{B}_n$ have mesh $\|\mathcal{B}_n\| \to 0$. Each discretized game is a finite-population potential game (Proposition 5) with potential Φ_N, and two limits interact: population $N \to \infty$ sends $\Phi_N(\mathbf{e}_k^N)$ to $\mathcal{V}(b_k^n)$, and grid $n \to \infty$ refines the discretization.

Uniform Riemann-sum bound. By the Heine-Cantor theorem, the continuous function U_E^{info} on the compact set $\mathcal{B} \times [0, 1]$ is uniformly continuous with modulus ω_U, so the Riemann-sum approximation satisfies

$$|\Phi_N(\mathbf{e}_k^N) - \mathcal{V}(b_k^n)| \le \varepsilon_N \equiv \omega_U(1/N) \to 0 \tag{16}$$

uniformly across grids.

Selection on the grid. By Theorem 1, $b_{\text{SS}}^{n,N} = \arg\max_{b_k^n \in \mathcal{B}_n} \Phi_N(\mathbf{e}_k^N)$. We take the limits in the same order as the body proof: $\beta \to \infty$ first (Gibbs concentration on the finite grid), then $N \to \infty$ (Riemann sum to integral), then mesh $\to 0$ (grid refinement). Combining (16) with the gap condition for $\mathcal{V}$, the maximizer of Φ_N on $\mathcal{B}_n$ coincides with $b_{\text{SS}}^n \equiv \arg\max_{b \in \mathcal{B}_n} \mathcal{V}(b)$ once $2\varepsilon_N$ is below the minimum grid gap of $\mathcal{V}$. Because the gap to the second-best benchmark on $\mathcal{B}_n$ shrinks as $n \to \infty$, the required $N(n)$ grows accordingly. Under genericity (Step 6 of the proof of Theorem 1), the limiting gap to the second-best benchmark in $\mathcal{B}$ is strictly positive, so $N(n)$ grows at a rate bounded by the inverse modulus of continuity of $\mathcal{V}$ on a neighborhood of b^*.

Grid limit. $\mathcal{V}$ is continuous on a compact set with unique maximizer b^* (genericity). A standard continuity argument gives $b_{\text{SS}}^n \to b^*$. Taking the limits in the order $\beta \to \infty$, $N \to \infty$, mesh $\to 0$, the discrete selection $b_{\text{SS}}^{n,N}$ converges to b^*. Grid independence follows from the fact that the same continuous functional $\mathcal{V}$ governs selection on every grid, avoiding the Oechssler and Riedel (2001) pathology.

Part (b): Variational characterization.

If $U_E^{\text{info}}(b, \rho)$ is C^1 in b for each ρ and the derivative is uniformly bounded, then

$\mathcal{V}(b) = \int_0^1 U_E^{\text{info}}(b, \rho)\, d\rho$ is differentiable by the Leibniz integral rule:

$$\mathcal{V}'(b) = \int_0^1 \frac{\partial U_E^{\text{info}}}{\partial b}(b, \rho)\, d\rho.$$

At an interior maximizer: $\mathcal{V}'(b^*) = 0$. Using $U_E^{\text{info}}(b, \rho) = U_E^{\text{base}}(b) + \Lambda(b, \rho)$:

$$\frac{dU_E^{\text{base}}}{db}(b^*) + \int_0^1 \frac{\partial \Lambda}{\partial b}(b^*, \rho)\, d\rho = 0.$$

The second-order condition $\mathcal{V}''(b^*) < 0$ holds when U_E^{base} is strictly concave near b^* (Proposition 1) and Λ is not too convex in b.

Part (c): Transition barrier.

Fix a grid $\mathcal{B}_n$ and consider N evaluators playing log-linear dynamics with noise $\varepsilon > 0$. At the monomorphic state $\mathbf{e}_k^N$, the process is a finite-state Markov chain. By Blume (1993), in a potential game with log-linear dynamics, the invariant distribution places weight

$$\mu_\varepsilon(\mathbf{x}) \propto \exp\left(\frac{N}{\varepsilon}\Phi_N(\mathbf{x})\right)$$

on each state $\mathbf{x}$. The expected exit time from the basin of attraction of monomorphic state $\mathbf{e}_{k^*}^N$ satisfies

$$\log \mathbb{E}[\tau^{n,N}] \;\sim\; \frac{N}{\varepsilon}\left[\Phi_N(\mathbf{e}_{k^*}^N) - \Phi_N(\mathbf{e}_{k'}^N)\right]$$

as $N/\varepsilon \to \infty$, where k' is the second-best benchmark on $\mathcal{B}_n$ (Freidlin and Wentzell, 1998). By the uniform Riemann-sum bound (16), $\Phi_N(\mathbf{e}_k^N) = \mathcal{V}(b_k^n) + O(\varepsilon_N)$ uniformly in b_k^n, with $\varepsilon_N \to 0$. Substituting,

$$\log \mathbb{E}[\tau^{n,N}] \;=\; \frac{N}{\varepsilon}\left[\mathcal{V}(b_{k^*}^n) - \mathcal{V}(b_{k'}^n)\right] \;+\; O\!\left(\frac{N}{\varepsilon}\,\varepsilon_N\right).$$

Taking the limits in the order $\beta \to \infty$, $N \to \infty$, then mesh $\to 0$ (consistent with Theorem 1), with $N/\varepsilon \to \infty$ at a rate satisfying $(N/\varepsilon)\,\varepsilon_N \to 0$ once the grid is fixed, yields the stated result. The barrier decomposition follows by writing $\mathcal{V}$ as U_E^{base} plus the integral of Λ. □

OA.4 Selection with Strategy-Specific Externalities

This section relocates from the body the abstract selection result that underlies Theorem 1. The result holds in any symmetric population game in which each strategy carries its own externality, with benchmarks being one instance.

Theorem OA.8 (Selection with Strategy-Specific Externalities)**.** *Consider a symmetric population game with finite strategy set* $\mathcal{S} = \{1, \ldots, n\}$ *and population state* $\mathbf{x} \in \Delta^{n-1}$*. Suppose the payoff from strategy* k *takes the form*

$$U_k(\mathbf{x}) = v(k) + \Lambda(k, x_k), \tag{17}$$

where $v : \mathcal{S} \to \mathbb{R}$ *is an intrinsic payoff independent of the population state, and* $\Lambda : \mathcal{S} \times [0,1] \to \mathbb{R}$ *is a strategy-specific externality satisfying* $\Lambda(k, 0) = 0$ *and* $\partial\Lambda/\partial x_k > 0$ *for all* k*. Then:*

(a) ***Potential game.*** *The game is an exact population potential game with potential*

$$\Phi(\mathbf{x}) = \sum_{k=1}^{n} \int_0^{x_k} \big[v(k) + \Lambda(k, \rho)\big]\, d\rho.$$

(b) ***Selection.*** *The unique stochastically stable strategy generically maximizes*

$$k^* = \arg\max_{k \in \mathcal{S}} \left[\; v(k) + \int_0^1 \Lambda(k, \rho)\, d\rho \;\right]. \tag{18}$$

Proof. (a) The own-share property $U_k(\mathbf{x}) = v(k) + \Lambda(k, x_k)$ implies $\partial U_k/\partial x_j = 0$ for $j \neq k$. The payoff Jacobian is diagonal, hence symmetric—the Monderer–Shapley condition (Monderer and Shapley, 1996). (b) At monomorphic state $\mathbf{e}_k$: $\Phi(\mathbf{e}_k) = v(k) + \int_0^1 \Lambda(k, \rho)\, d\rho$. The finite-population approximation and large-N limit (as in the proof of Theorem 1, Appendix C.2) establishes that the stochastically stable state maximizes Φ via Young (1993, Theorem 4). □

Remark OA.9 (Externality Structure and Selection)**.** If Λ is strategy-independent, the externality drops out: $k^* = \arg\max_k v(k)$. If Λ is strategy-specific, externalities generically change the selection.

Theorem 1 is the special case with $v(k) = U_E^{\text{base}}(b_k)$ and $\Lambda(k, x_k) = \Lambda(b_k, x_k)$, where the strategy-specific externality arises endogenously from information production. Other sources of strategy-specific externalities include network effects, institutional infrastructure, and regulatory monitoring investment, and in each case the surviving strategy is the one that generates the richest ecosystem along the path to full adoption.

OA.5 Illustrative Figures

This section collects schematic figures that illustrate the mechanism described in the body. The figures duplicate content the prose already carries and are relegated here for compactness.

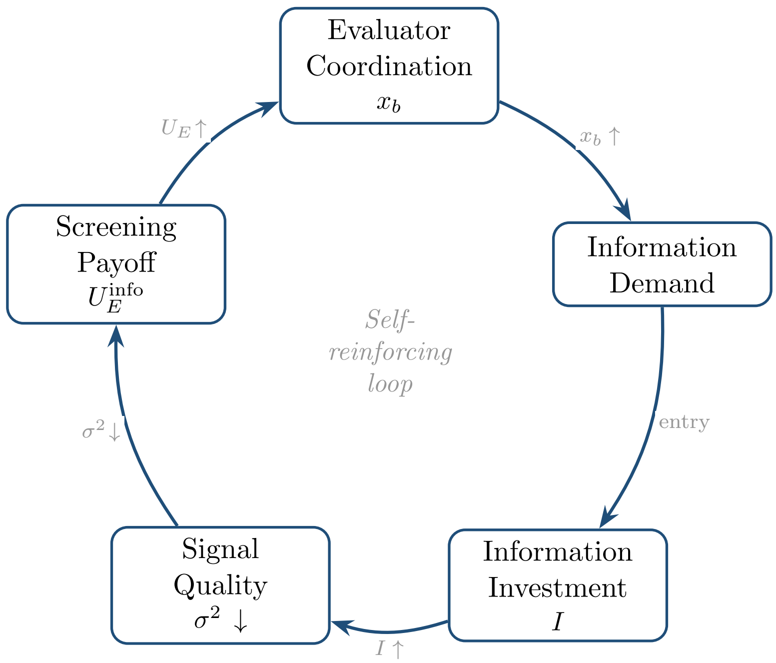
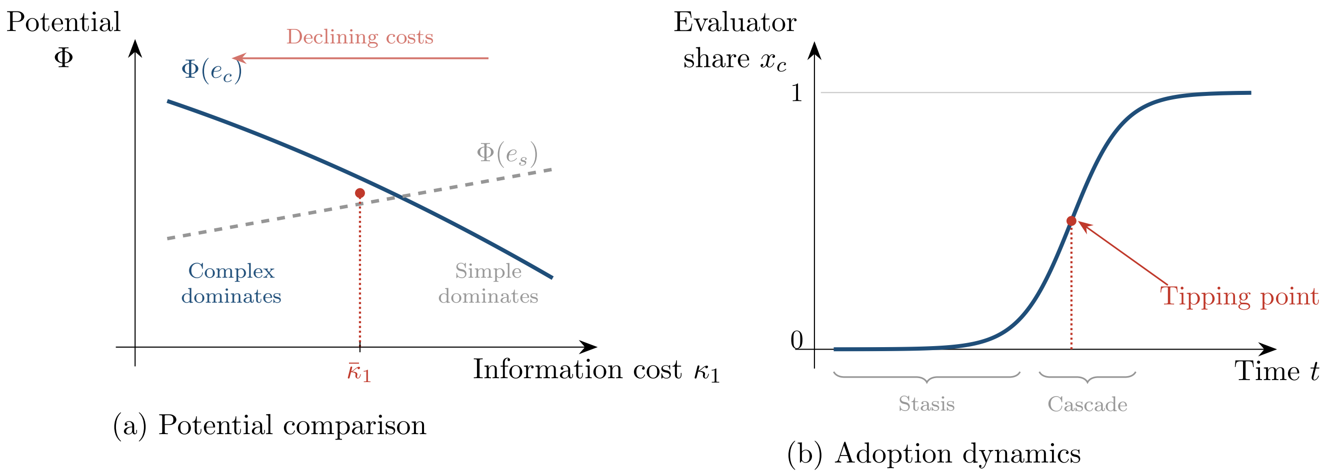


Figure 1: The self-reinforcing information loop. Evaluator coordination on benchmark b raises demand for benchmark-relative information, attracting investment that reduces signal noise and improves screening quality, which in turn attracts more evaluators. This micro-founded coordination externality is the paper's core economic mechanism.



(a) Potential comparison

(b) Adoption dynamics

Figure 2: Benchmark transition dynamics. Panel (a): as information costs κ_1 decline, the potential of the complex benchmark (solid) eventually exceeds that of the simple benchmark (dashed), triggering a transition at $\bar{\kappa}_1$. Panel (b): the transition in evaluator adoption is sudden, an extended period of stasis followed by rapid cascade, consistent with the historical shift from zero earnings to analyst forecasts in the 1990s (Proposition 7).

OA.6 Additional Applications

The body of the paper develops the credit and financial markets applications. The mapping carries over with no change to two further domains, education and regulatory compliance, presented here.

OA.6.1 Education: GPA Thresholds

Mapping. Students are reporters with type θ underlying academic ability. The reported GPA is the manipulable signal, the 3.0 cumulative GPA is a common benchmark, and grading professors together with advising offices and tutoring centers are information producers.

(A1) holds. Stronger students with academic slack can absorb a difficult course without falling below 3.0, while weaker students must resort to costlier adjustments like dropping courses, strategic withdrawals, retakes, or grade appeals. The cost of pushing the GPA up by a fixed amount is decreasing in true ability.

Information ecosystem and predictions. Professors apply more grading care near the cutoff, tutoring centers allocate disproportionate resources there, and advising effort concentrates at the threshold, exactly as Lemma 5 predicts. The model predicts (i) bunching at 3.0 with a gap below, (ii) persistence of 3.0 rather than 3.1 or 2.9 because the installed ecosystem at 3.0 is richest, and (iii) recruiters' continued use of the threshold despite known gaming because manipulation is informative under (A1).

OA.6.2 Regulatory Compliance

Mapping. The regulated entity (factory, bank, pharmaceutical manufacturer) is the reporter, true compliance is the type, and the regulatory standard (EPA emissions limit, Basel minimum capital ratio, FDA approval threshold) is the benchmark. Compliance consultants, environmental monitors, risk-modeling firms, and regulatory affairs professionals are information producers.

(A1) holds. Entities with better underlying performance reach the threshold cheaply (timing of measurements, routine calibration, light capital adjustments), while marginal entities require costly abatement, capital optimization, risk-weight arbitrage, or measurement manipulation.

Information ecosystem and predictions. The EPA's Clean Air Act standards sustain an environmental consulting industry, Basel requirements sustain a risk-modeling industry, and the FDA approval threshold sustains clinical trial design firms, all concentrated at the regulatory boundary. The model predicts (i) bunching of regulated outcomes just above the threshold, (ii) the surviving standard is the one that attracts the richest monitoring ecosystem, and (iii) regulators rationally persist with manipulated thresholds because the infrastructure those thresholds attract compensates for the gaming.

References

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior* 5, 387–424.

Carlsson, H. and E. van Damme (1993). Global games and equilibrium selection. *Econometrica* 61, 989–1018.

Frankel, D. M., S. Morris, and A. Pauzner (2003). Equilibrium selection in global games with strategic complementarities. *Journal of Economic Theory* 108, 1–44.

Freidlin, M. I. and A. D. Wentzell (1998). *Random Perturbations of Dynamical Systems* (2nd ed.). New York: Springer.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *American Economic Review* 70, 393–408.

Harsanyi, J. C. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Kandori, M., G. J. Mailath, and R. Rob (1993). Learning, mutation, and long run equilibria in games. *Econometrica* 61, 29–56.

Monderer, D. and L. S. Shapley (1996). Potential games. *Games and Economic Behavior* 14, 124–143.

Oechssler, J. and F. Riedel (2001). Evolutionary dynamics on infinite strategy spaces. *Economic Theory* 17, 141–162.

Sandholm, W. H. (2010). *Population Games and Evolutionary Dynamics*. Cambridge, MA: MIT Press.

Veldkamp, L. L. (2006). Media frenzies in markets for financial information. *American Economic Review* 96, 577–601.

Young, H. P. (1993). The evolution of conventions. *Econometrica* 61, 57–84.