

Toy Models based on Published Papers

Industrial Organization, EconS 594

Felix Munoz-Garcia

Washington State University

Contents

Price Discrimination

1. Third-Degree Price Discrimination and Demand Curvature.....3
Based on Aguirre, Cowan, and Vickers (2010)
2. Third-Degree Price Discrimination in Oligopoly When Markets Are Covered 15
Based on Dertwinkel-Kalt and Wey (2023)
3. Two-Part Tariffs with Two Firms 24
Based on Armstrong and Vickers (2001)

Entry Barriers

4. Discounts as a Barrier to Entry.....34
Based on Ide, Montero, and Figueroa (2016)

Collusion and Cartels

5. Forward Trading and Collusion 42
Based on Liski and Montero (2006)
6. Collusion, Leniency Programs, and Whistleblower Rewards 50
Based on Aubert, Rey, and Kovacic (2006)

Signaling, Disclosure, and Certification

7. Communicating Quality: Disclosure or Price Signaling..... 59
Based on Daughety and Reinganum (2008)
8. Signaling a Clean Product with Its Price.....68
Based on Mahenc (2008)
9. Penetration or Skimming Pricing for Credence Products? 78
Based on Mahenc (2025)
10. Label Confusion and the Groucho Effect 86
Based on Harbaugh, Maxwell, and Roussillon (2011)
11. Eco-labeling with Noisy Certification Tests 103
Based on Mason (2011)

Information Design and Persuasion

12. Bayesian Persuasion: Four Papers 114
Based on Kamenica and Gentzkow (2011) and related papers
13. Funding a Review Agency under Obfuscation.....145
Based on Figueroa and Lemus (2024)

Platforms and the Attention Economy

14. The Attention Economy 156
Based on Chen (2026)

Environmental Regulation and Innovation

15. Greener or Cheaper: Economies of Scope in R&D 165
Based on Strandholm, Espinola-Arredondo, and Munoz-Garcia (2025)

Third-Degree Price Discrimination and Demand Curvature

A step-by-step exercise in the spirit of Aguirre, Cowan, and Vickers (2010)

Companion note to Aguirre, I., S. Cowan, and J. Vickers (2010), “Monopoly Price Discrimination and Demand Curvature,”

American Economic Review 100(4): 1601–1615. Full references at the end.

Overview. A monopolist that can tell its customers apart by an exogenous characteristic, such as age, location, or the time of purchase, charges each group its own monopoly price. Moving from a single price to discriminatory prices raises the firm’s profit, harms the consumers in the markets where the price rises, and benefits those in the markets where it falls, so the effect on welfare, the sum of consumer surplus and profit, can have either sign. Aguirre, Cowan, and Vickers (2010) show that the sign depends on simple comparisons of the curvature of demand in the markets. The effect of discrimination on welfare is the sum of a misallocation effect, because consumers face different prices, which always lowers welfare, and an output effect, which raises welfare when total output increases, because the price exceeds marginal cost. This note builds the mechanism with the pair of demands in the paper’s Example 1, a linear demand and an exponential demand, and moves one parameter of the exponential demand to produce every case. Step (a) solves pricing with and without discrimination; Step (b) relaxes a constraint on the price difference and decomposes the welfare effect; Step (c) states the increasing ratio condition and shows when discrimination lowers welfare; Step (d) shows when it raises welfare, and when welfare first rises and then falls; Step (e) turns to total output; Step (f) treats demands with constant curvature of inverse demand; Step (g) treats constant-elasticity demands; and Step (h) lists what the toy model keeps and drops.

1. The Toy Model

Firm and markets. A monopolist with constant marginal cost $c = 1$ sells in two markets whose demands are independent of each other. Utility is quasi-linear, so social welfare W is consumer surplus plus profit. Demand in a generic market is $q(p)$, which is twice differentiable and decreasing, and the profit in a market is $\pi(p) = (p - c) q(p)$. With third-degree price discrimination the firm sets a price in each market; without it, the firm must charge the same *uniform* price in both. Throughout, both markets are served at the uniform price, so discrimination does not open a new market.

Curvature. Four objects describe a demand function. The price elasticity is $\eta \equiv -\frac{pq'}{q}$ and the Lerner index is $L \equiv \frac{p-c}{p}$. The curvature of direct demand is $\alpha \equiv -\frac{pq''}{q'}$, the elasticity of the slope of demand, and the curvature of inverse demand $p(q)$ is $\sigma \equiv -\frac{qp''}{p'} = \frac{q q''}{(q')^2}$. The two curvatures are linked by

$$\sigma = \frac{\alpha}{\eta}. \quad (1)$$

Linear demand has $\alpha = \sigma = 0$, convex demand has $\alpha > 0$ and $\sigma > 0$ (inverse demand is convex too), and concave demand has $\alpha < 0$ and $\sigma < 0$. The second derivative of profit is

$$\pi''(p) = 2q' + (p - c) q'' = q' (2 - L\alpha), \quad (2)$$

so profit is strictly concave if and only if $2 - L\alpha > 0$, which is assumed throughout.

The two demands. Market L has linear demand and market E has exponential demand,

$$q_L(p) = 9 - p, \quad q_E(p) = 8 e^{-(p-1)/b}, \quad b > 0. \quad (3)$$

The two demands are normalized to be equal at the cost price, $q_L(c) = q_E(c) = 8$, so the scale b of market E is the only free parameter and curvature is the only difference between the markets: demand E is convex, with $\alpha > 0$, while demand L is the borderline linear case. In market L , $\alpha_L = \sigma_L = 0$ and $\eta_L = \frac{p}{9-p}$. In market E , $\sigma_E = 1$ and $\alpha_E = \eta_E = \frac{p}{b}$, so $2 - L\alpha_E = 2 - \frac{p-1}{b}$ and profit is strictly concave at every price below $1 + 2b$; the examples below satisfy this condition. Writing the margin as $p - c$, demand E falls by the factor e^{-1} when the margin rises by b .

2. Step (a): Pricing with and without Discrimination

Discriminatory prices. With discrimination the firm maximizes π in each market, so

$$\pi'(p^*) = q(p^*) + (p^* - c) q'(p^*) = 0 \iff L^* \eta^* = 1, \quad (4)$$

the Lerner condition. Since $\alpha = \sigma\eta$, it follows that $L^* \alpha^* = \sigma^*$ and $2 - L^* \alpha^* = 2 - \sigma^* > 0$ at the discriminatory price. In the toy,

$$p_L^* = \frac{9+c}{2} = 5, \quad p_E^* = b + c = b + 1. \quad (5)$$

The uniform price. Without discrimination the firm chooses the price $\bar{p}$ that solves $\pi'_L(\bar{p}) + \pi'_E(\bar{p}) = 0$, that is,

$$10 - 2\bar{p} + 8e^{-(\bar{p}-1)/b} \left(1 - \frac{\bar{p}-1}{b}\right) = 0, \quad (6)$$

which has no closed form (in terms of the margin $m = \bar{p} - c$ it reads $8 - 2m + 8e^{-m/b}(1 - \frac{m}{b}) = 0$). Since the two marginal profits cancel at $\bar{p}$, one market has $\pi' < 0$ and the other has $\pi' > 0$ there. Write w for the market with $\pi'_w(\bar{p}) = q_w [1 - L\eta_w] < 0$ and s for the market with $\pi'_s(\bar{p}) = q_s [1 - L\eta_s] > 0$.

Result 1 (weak and strong markets). At the uniform price:

- the weak market w has the higher elasticity, $\eta_w(\bar{p}) > \eta_s(\bar{p})$;
- the uniform price lies between the discriminatory prices, $p_w^* < \bar{p} < p_s^*$, because profit is strictly concave in each market;
- with discrimination, the price falls in the weak market and rises in the strong market.

In the toy, market E is the weak market if and only if $b + 1 < 5$, that is, $b < 4$. Table 1 reports the three values of b used below, $b = 5, 3, 1.8$.

b	Weak market	p_w^*	p_s^*	$\bar{p}$	$\eta_w(\bar{p})$	$\eta_s(\bar{p})$
5	L	5	6	5.2546	1.4030	1.0509
3	E	4	5	4.7217	1.5739	1.1036
1.8	E	2.8	5	4.4604	2.4780	0.9825

Table 1: Discriminatory and uniform prices in the toy model, with $c = 1$ and $q_E(p) = 8e^{-(p-1)/b}$.

Figure 1 depicts how the uniform price is found, with the price on the horizontal axis. The solid curve is the marginal profit of the strong market, which is positive at $\bar{p}$, and the dashed curve is minus the marginal profit of the weak market. The uniform price is where the two curves cross, so that $\pi'_s(\bar{p}) = -\pi'_w(\bar{p})$. Panel (a) has $b = 5$, where market L is weak, and panel (b) has $b = 3$, where the roles are reversed.

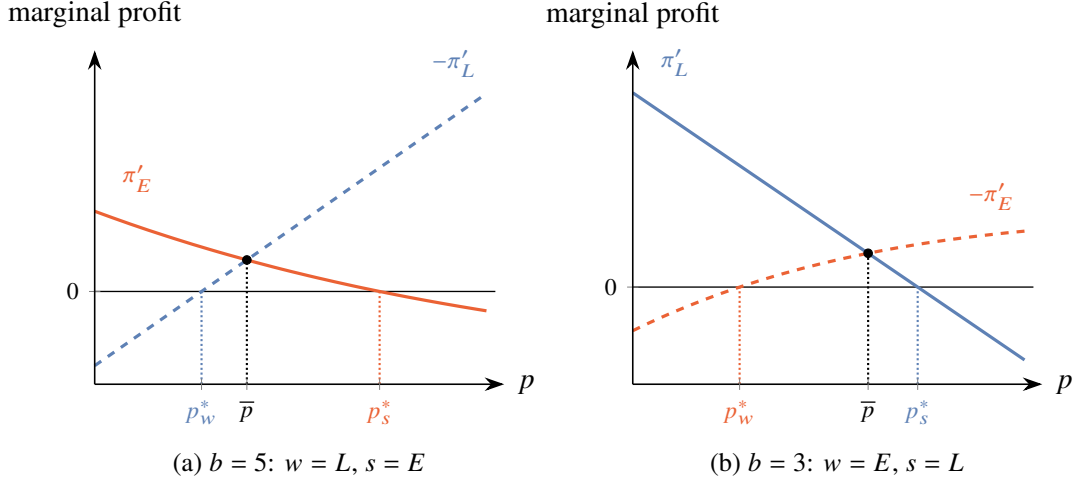


Figure 1: Marginal profit of the strong market (solid) and minus the marginal profit of the weak market (dashed), with $c = 1$ and $q_E(p) = 8 e^{-(p-1)/b}$.

3. Step (b): Relaxing the Price-Difference Constraint

The paper moves from the uniform price to discrimination in small steps. The firm chooses its prices subject to the constraint $p_s - p_w \leq r$, where $r \geq 0$ is the degree of discrimination allowed. When the constraint binds, the firm maximizes $\pi_w(p_w) + \pi_s(p_w + r)$, so

$$\pi'_w(p_w) + \pi'_s(p_w + r) = 0. \quad (7)$$

At $r = 0$ the firm charges the uniform price $\bar{p}$, and the constraint stops binding at $r^* = p_s^* - p_w^*$, which equals 1 for $b = 5$ and $b = 3$ and equals 2.2 for $b = 1.8$. Differentiating (7) with respect to r ,

$$p'_w(r) = \frac{-\pi''_s}{\pi''_w + \pi''_s} < 0, \quad p'_s(r) = \frac{\pi''_w}{\pi''_w + \pi''_s} > 0, \quad (8)$$

so the price in the weak market falls and the price in the strong market rises as r grows, both in the direction of the discriminatory prices.

The welfare effect. The marginal effect of a price on welfare in a market is $\frac{dW}{dp} = (p - c) q'(p)$: a price change moves quantity, and each unit of quantity has social value equal to the price-cost margin. Hence

$$W'(r) = (p_w - c) q'_w p'_w(r) + (p_s - c) q'_s p'_s(r). \quad (9)$$

Adding and subtracting $\bar{p}$ and writing $Q'(r) \equiv q'_w p'_w + q'_s p'_s$ for the marginal effect on total output,

$$W'(r) = \underbrace{(p_w - \bar{p}) q'_w p'_w + (p_s - \bar{p}) q'_s p'_s}_{\text{misallocation effect}} + \underbrace{(\bar{p} - c) Q'(r)}_{\text{value of the output effect}}. \quad (10)$$

Each term of the misallocation effect is the product of a price difference, a negative slope of demand, and a price change of the same sign as the difference, so the misallocation effect is zero at $r = 0$ and negative for $r > 0$. Integrating (10) over $[0, r^*]$, the total welfare effect equals a negative misallocation term plus $(\bar{p} - c)$ times the total change in output, so an increase in output is necessary for welfare to rise.

Figure 2 depicts the case $b = 1.8$, with r on the horizontal axis. Panel (a) shows the two prices leaving $\bar{p} = 4.4604$ in opposite directions until they reach $p_w^* = 2.8$ and $p_s^* = 5$ at $r^* = 2.2$. Panel (b) shows the

decomposition (10). At $r = 0$ the misallocation effect is zero, so $W'(0) = 2.109$ is the value of the output effect. As r grows the misallocation effect becomes more negative, falling to -1.74 at r^* , while the output term falls to 1.56 , and $W'(r)$ turns negative at $\hat{r} = 2.0566$.

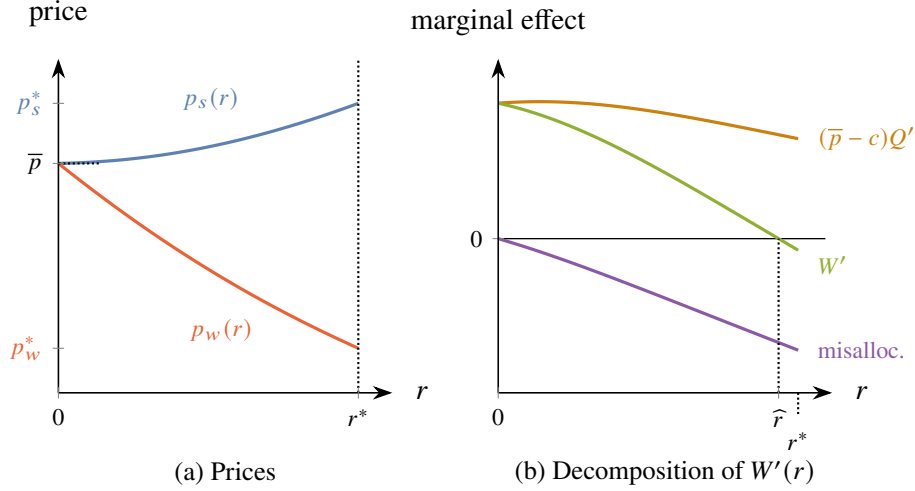


Figure 2: Prices and the decomposition of the marginal welfare effect against the allowed price difference r , with $b = 1.8$; the labels in panel (b) name the curves that end at r^* : the output term $(\bar{p} - c)Q'(r)$, the marginal welfare effect $W'(r)$, and the misallocation effect.

4. Step (c): The Increasing Ratio Condition and When Welfare Falls

The paper rewrites (9) using a ratio that depends on curvature. Define

$$z(p) \equiv \frac{(p - c) q'(p)}{2q'(p) + (p - c) q''(p)} = \frac{p - c}{2 - L\alpha}, \quad (11)$$

the ratio of the marginal effect of the price on welfare to the second derivative of profit. Writing $(p - c)q' = z\pi''$ in (9) and using (8),

$$W'(r) = \underbrace{\frac{-\pi_w'' \pi_s''}{\pi_w'' + \pi_s''}}_{> 0} [z_w(p_w(r)) - z_s(p_s(r))], \quad (12)$$

so the marginal welfare effect has the sign of $z_w - z_s$. In the toy,

$$z_L(p) = \frac{p - 1}{2}, \quad z_E(p) = \frac{b(p - 1)}{2b - (p - 1)}. \quad (13)$$

The increasing ratio condition (IRC). The ratio $z(p)$ is increasing in p in each market. It holds for linear demands, for inverse demands with constant positive curvature (which include exponential and constant-elasticity demands), for direct demands with constant curvature, positive or negative, for logit and probit demands, and for demands derived from the lognormal distribution; it always holds near marginal cost, since $z'(c) = \frac{1}{2}$. It is not universal: with constant negative curvature of inverse demand it fails at high enough prices. In the toy, both z_L and z_E are increasing on the range where profit is strictly concave.

Result 2 (single peak). Under the IRC, welfare $W(r)$ is either decreasing in r , or increasing in r , or first increasing and then decreasing.

The intuition is direct. As r grows, p_w falls and p_s rises, so under the IRC $z_w(p_w(r))$ falls and $z_s(p_s(r))$ rises. The gap $z_w - z_s$ in (12) therefore decreases in r and changes sign at most once, from positive to negative. The paper states this as a Lemma: at any $\hat{r}$ with $W'(\hat{r}) = 0$, $W''(\hat{r}) < 0$. Which of the three shapes arises depends on the sign of $W'(0)$ and of $W'(r^*)$.

Figure 3 depicts the two ratios along the constraint path, with r on the horizontal axis, for $b = 5$, $b = 3$, and $b = 1.8$. In panel (a) the strong market has the higher ratio for every r , so $W' < 0$ throughout; in panel (b) the weak market does, so $W' > 0$ throughout; and in panel (c) the curves cross at $\hat{r} = 2.0566$, so W' changes sign.

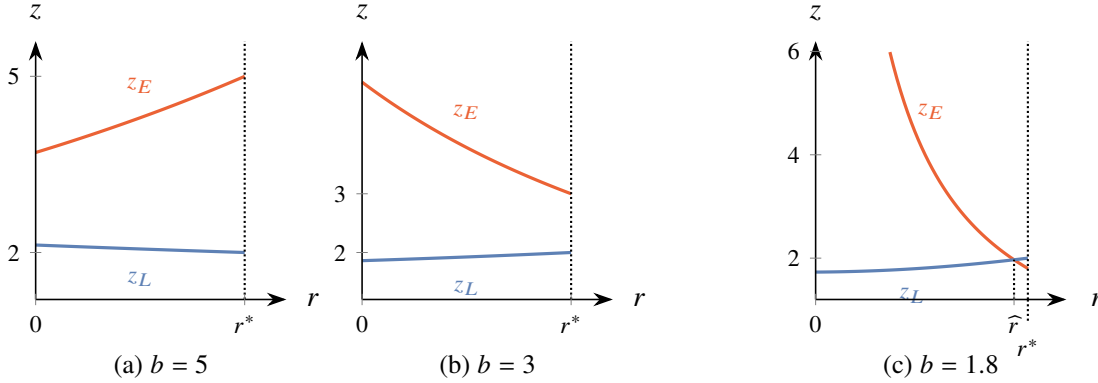


Figure 3: The ratios $z_w(p_w(r))$ and $z_s(p_s(r))$ against the allowed price difference r , for L (blue) and E (red); in panel (c) the curve for E is not drawn where it exceeds 6, which is the case for $r < 0.77$.

When discrimination lowers welfare. At $r = 0$ the two markets share the price $\bar{p}$ and the Lerner index, so $z_w(\bar{p}) - z_s(\bar{p})$ has the sign of $\alpha_w(\bar{p}) - \alpha_s(\bar{p})$, the difference in the curvature of direct demand. By Result 2, if $W'(0) \leq 0$ then welfare is decreasing for every $r > 0$.

Result 3 (Proposition 1 of the paper). Under the IRC, if the direct demand in the strong market is at least as convex as in the weak market at the uniform price, $\alpha_s(\bar{p}) \geq \alpha_w(\bar{p})$, then discrimination reduces welfare.

At the uniform price the misallocation effect is zero, so the welfare effect has the sign of the output effect, and the condition on α is the condition under which output does not rise locally. The IRC extends this local result to every larger r . The condition covers the case in which demand in the strong market is an affine transformation of demand in the weak market, $q_s(p) = M + N q_w(p)$ with $M, N > 0$ (Cowan, 2007), since the curvature of direct demand at a given price is then the same in both markets. In the toy, $\alpha_E(\bar{p}) = \frac{\bar{p}}{b} > 0 = \alpha_L$, so Result 3 applies whenever E is the strong market, that is, $b > 4$. For $b = 5$, $\alpha_E(5.2546) = 1.0509 > 0$ and discrimination lowers welfare by 1.13. Conversely, for welfare to rise, direct demand in the weak market must be strictly more convex than in the strong market at the uniform price.

Figure 4 depicts why a more convex weak market gains more. The inverse demand $p_1(q) = 10 - q$ is linear, and $p_2(q) = 10 - q + 0.06(q - \bar{q})^2$ is more convex, with the same price and slope at the uniform quantity $\bar{q} = 3$, so the uniform price $\bar{p} = 7$ is the same for both. Their marginal revenue curves also cross at $\bar{q}$, but for $q > \bar{q}$ the price p_2 is higher, so a given increase in output lowers the price by less and $MR_2(q) > MR_1(q)$. The discriminatory quantity is therefore higher with the more convex demand, $q_2 = 5.53 > q_1 = 4.5$, and the welfare gain in the market is larger for two reasons: the output increase is bigger, and the marginal social value of output, the price minus c , is higher at every quantity above $\bar{q}$.

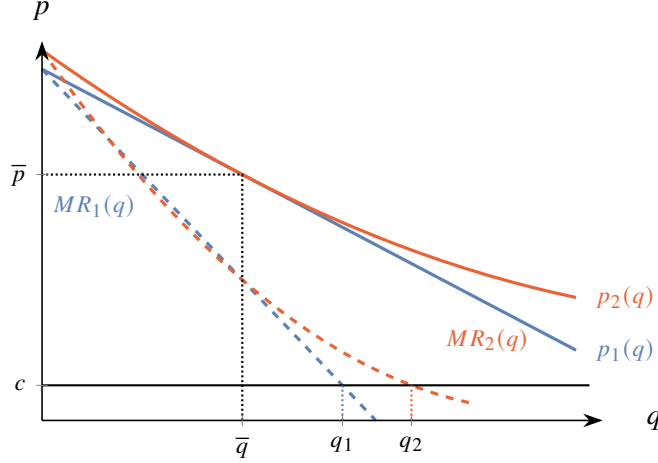


Figure 4: Linear inverse demand $p_1(q)$ and a more convex inverse demand $p_2(q)$ with the same price and slope at $\bar{q} = 3$, with their marginal revenue curves (dashed) and $c = 1$.

5. Step (d): When Welfare Rises

The sufficient condition for welfare to rise is found by looking at what happens at the discriminatory prices. Since $L^* \alpha^* = \sigma^*$, the ratio (11) at a discriminatory price is

$$z(p^*) = \frac{p^* - c}{2 - \sigma^*}. \quad (14)$$

If $z_w(p_w^*) \geq z_s(p_s^*)$, then $W'(r^*) \geq 0$ by (12), and by Result 2 welfare is increasing for every $r < r^*$.

Result 4 (Proposition 2 of the paper). Under the IRC, if $\frac{p_w^* - c}{2 - \sigma_w^*} \geq \frac{p_s^* - c}{2 - \sigma_s^*}$, then welfare is higher with discrimination.

Since $p_w^* < p_s^*$, the condition requires $\sigma_w^* > \sigma_s^*$: inverse demand in the weak market is more convex than in the strong market at the discriminatory prices, and the prices are close enough together that the misallocation effect is small. The condition depends only on the discriminatory prices and curvatures, so it does not depend on the multiplicative size of a market. In the toy, with E the weak market, $z_E(p_E^*) = \frac{b}{2-1} = b$ and $z_L(p_L^*) = \frac{4}{2} = 2$, so the condition is $b \geq 2$. For $b = 3$, discrimination raises welfare by 1.031, and by 1.527 if market E is twice as large at the cost price, $q_E(c) = 16$.

Equation (14) also has an empirical use. The pass-through of a cost change to the discriminatory price is $\frac{dp^*}{dc} = \frac{1}{2 - \sigma^*}$, which is 1 in market E and $\frac{1}{2}$ in market L , so a regression of the monopoly price on marginal cost estimates σ^* without estimating the demand function.

The intermediate case. The opposite of the condition of Result 4 gives $W'(r^*) < 0$. Combined with $\alpha_w(\bar{p}) > \alpha_s(\bar{p})$, which gives $W'(0) > 0$, welfare first rises and then falls.

Result 5 (Proposition 3 of the paper). Under the IRC, if

- direct demand in the weak market is more convex than in the strong market at the uniform price, $\alpha_w(\bar{p}) > \alpha_s(\bar{p})$, and
- inverse demand in the strong market is at least as convex as in the weak market at the discriminatory prices, $\sigma_s^* \geq \sigma_w^*$,

then welfare rises for $r < \hat{r}$ and falls for $r > \hat{r}$, with a unique peak $\hat{r} \in (0, r^*)$.

The proof uses the second condition only through $W'(r^*) < 0$, which holds in the toy whenever E is weak and $b < 2$, even though $\sigma_E^* = 1 > \sigma_L^* = 0$. The result does not say whether full discrimination raises or lowers welfare. In the toy the intermediate case is mild: profit in market E is strictly concave at the uniform price only for $b \geq 1.73$, so the case is confined to $1.73 \leq b < 2$. For $b = 1.8$ the peak is +2.440 at $\hat{r} = 2.0566$ and full discrimination leaves welfare +2.427 above the uniform-price level. Figure 6(a) in Step (f) shows a hump that is easy to see, in a pair of demands to which Result 5 also applies, and in which full discrimination ends below the uniform-price level.

Figure 5 depicts $W(r) - W(0)$ for the cases of Results 3 and 4, with r on the horizontal axis: welfare falls with $b = 5$ in panel (a) and rises with $b = 3$ in panel (b).

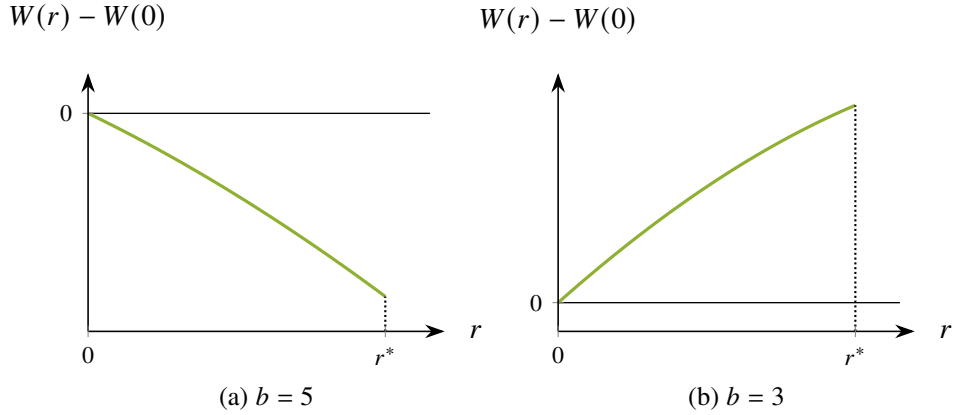


Figure 5: Welfare relative to the uniform price against the allowed price difference r , with $q_E(p) = 8e^{-(p-1)/b}$.

Table 2 collects the cases. The two thresholds, $b = 4$ (market E becomes the weak market) and $b = 2$ (the condition of Result 4 fails), divide the values of b into the three regimes. Both thresholds follow from the discriminatory prices alone, so they do not depend on the size of market E , and they are the same for any normalization of $q_E(c)$.

b	Result	Shape of $W(r)$	$W(r^*) - W(0)$
$b > 4$ (5)	Result 3	decreasing	-1.134
$2 \leq b < 4$ (3)	Result 4	increasing	+1.031
$1.73 \leq b < 2$ (1.8)	Result 5	rises, then falls	+2.427

Table 2: Welfare regimes in the toy model, with $q_E(p) = 8e^{-(p-1)/b}$.

6. Step (e): The Effect on Total Output

Since a rise in output is necessary for welfare to rise, the paper applies the same method to output. Using (8), the marginal output effect is

$$Q'(r) = \underbrace{\frac{-q'_w q'_s}{\pi''_w + \pi''_s}}_{>0} [L_w \alpha_w - L_s \alpha_s], \quad (15)$$

and, since $\alpha = \eta\sigma$,

$$L_w \alpha_w - L_s \alpha_s = L_w \eta_w (\sigma_w - \sigma_s) + (L_w \eta_w - L_s \eta_s) \sigma_s. \quad (16)$$

At the uniform price the Lerner indices are equal, so $Q'(0)$ has the sign of $\alpha_w(\bar{p}) - \alpha_s(\bar{p})$; at the discriminatory prices $L\eta = 1$ in each market, so $Q'(r^*)$ has the sign of $\sigma_w - \sigma_s$. In between, $L_w \leq L_s$, and the mild assumption that $L\eta$ is increasing in the price implies $L_w \eta_w > L_s \eta_s$.

Result 6 (Proposition 4 of the paper). Output changes with discrimination as follows:

- if demand is concave in the strong market and less concave, or convex, in the weak market ($\alpha_w \geq 0 \geq \alpha_s$, or $0 \geq \alpha_w \geq \alpha_s$), output increases;
- if demand is convex in the strong market and concave, or less convex, in the weak market ($\alpha_s \geq 0 \geq \alpha_w$, or $\alpha_s \geq \alpha_w \geq 0$), output decreases;
- if inverse demands are convex, and more so in the weak market ($\sigma_w \geq \sigma_s \geq 0$), output increases;
- if inverse demands are concave, and more so in the weak market ($\sigma_w \leq \sigma_s \leq 0$), output decreases.

Linear demands, where $\alpha = \sigma = 0$, sit on the boundary of all four cases, which is Pigou's (1920) result that output does not change. The paper's summary is that output rises if both direct and inverse demand are more convex in the weak market, and does not rise if both are at least as convex in the strong market. In the toy, $\alpha_E > 0 = \alpha_L$ and $\sigma_E = 1 > 0 = \sigma_L$. If E is the weak market ($b < 4$), the first and third cases apply and output rises; if E is the strong market ($b > 4$), the second case applies and output falls. This holds whatever the size of market E .

Table 3 reports the decomposition (10) integrated over $[0, r^*]$. In the case $b = 5$ output falls, so both terms are negative and welfare falls. In the cases $b = 3$ and $b = 1.8$ output rises, and the output term exceeds the misallocation term in both. The pair of demands of Step (f) with $\sigma = \frac{1}{2}$ is a case in which the misallocation term is larger: output rises by 0.211 and welfare falls by 0.491, which is the point of Varian's (1985) necessary condition: more output is necessary but not sufficient.

b	ΔQ	$(\bar{p} - c) \Delta Q$	Misallocation	ΔW
5	-0.219	-0.930	-0.204	-1.134
3	+0.351	+1.306	-0.275	+1.031
1.8	+1.233	+4.268	-1.841	+2.427

Table 3: Effect of full discrimination on output, $\Delta Q = Q(r^*) - Q(0)$, and the decomposition of the welfare effect, $\Delta W = W(r^*) - W(0)$, with $q_E(p) = 8 e^{-(p-1)/b}$.

7. Step (f): Constant Curvature of Inverse Demand

When the two inverse demands have the same constant curvature $\sigma > 0$, Result 5 applies, because $\alpha = \sigma\eta$ is higher in the weak market, and Result 6 says that output rises. These results do not say whether welfare rises or falls with full discrimination. The paper answers with a second technique, which restricts quantities instead of prices. Consider inverse demands with constant curvature,

$$p_i(q) = a_i - \frac{b_i q^{1-\sigma}}{1-\sigma} \quad (\sigma \neq 1), \quad p_i(q) = a_i - b_i \ln q \quad (\sigma = 1). \quad (17)$$

The family contains linear demand ($\sigma = 0$), exponential demand ($\sigma = 1$), and quadratic demand $q = \left(\frac{a-p}{2b}\right)^2$ ($\sigma = \frac{1}{2}$), and its discriminatory price is $p^* = \frac{(1-\sigma)a+c}{2-\sigma}$ for $\sigma \neq 1$ and $p^* = b + c$ for $\sigma = 1$. Let $\bar{q}_i = q_i(\bar{p})$ and let $\Pi_i(q_i) = \pi_i(p_i(q_i))$ be profit as a function of quantity. The firm maximizes $\sum_i \Pi_i(q_i)$ subject to a limit $t \geq 0$ on how far quantities move from the uniform-price quantities,

$$-p'_w(\bar{q}_w)(q_w - \bar{q}_w) \leq -p'_s(\bar{q}_s)(q_s - \bar{q}_s) + t, \quad (18)$$

and the constrained-optimal quantities satisfy

$$\frac{\Pi'_w(q_w(t))}{p'_w(\bar{q}_w)} + \frac{\Pi'_s(q_s(t))}{p'_s(\bar{q}_s)} = 0. \quad (19)$$

At $t = 0$ the firm sells the uniform-price quantities. As t grows the firm raises quantity in the weak market and cuts it in the strong market, so prices move toward their discriminatory levels. The marginal welfare effect has the sign of

$$\delta(t) \equiv \frac{p_w - c}{2 - \sigma_w} \frac{p'_w(\bar{q}_w)}{p'_w(q_w)} - \frac{p_s - c}{2 - \sigma_s} \frac{p'_s(\bar{q}_s)}{p'_s(q_s)}, \quad (20)$$

and $\delta(0)$ has the sign of $\sigma_w - \sigma_s$. With constant σ_i , the derivative $\delta'(t)$ has the sign opposite to $1 - L\eta\sigma$ when this term has the same sign in both markets: δ keeps falling if $L\eta\sigma \leq 1$ in both markets and keeps rising if $L\eta\sigma \geq 1$ in both.

Result 7 (Proposition 5 of the paper). With constant curvature of inverse demand:

- if $1 > \sigma_s \geq \sigma_w$ and $L(\bar{p})\eta_w(\bar{p})\sigma_w \leq 1$, discrimination reduces welfare;
- if $\sigma_w \geq \sigma_s > 1$ and $L(\bar{p})\eta_s(\bar{p})\sigma_s \geq 1$, discrimination raises welfare.

If the discriminatory prices are close, $L\eta \approx 1$ and the second condition in each case follows from the first. For a common curvature σ the result is then that welfare falls if $\sigma < 1$, because the misallocation effect outweighs the positive output effect, and rises if $\sigma > 1$, because the output effect is strong enough to exceed the misallocation effect.

Figure 6 has two panels. Both markets have the discriminatory prices $p_w^* = 3$ and $p_s^* = 4$ and the discriminatory quantity 10, with $c = 1$, and a_i and b_i are set to match. Panel (a) depicts welfare against the allowed price difference r for $\sigma = \frac{1}{2}$: welfare rises to +0.179 at $\hat{r} = 0.337$ and then falls to 0.491 below its uniform-price level at $r^* = 1$, the hump of Result 5 with the net loss of Result 7. Panel (b) depicts the effect of full discrimination on welfare and output against a common curvature σ , with σ on the horizontal axis. Output is unchanged at $\sigma = 0$, the linear case, falls for $\sigma < 0$, and rises for $\sigma > 0$. Welfare falls for every σ below 0.965 and rises above it; the threshold is slightly below 1 because the two discriminatory prices, 3 and 4, are not equal. At $\sigma = \frac{1}{2}$ discrimination changes welfare by -0.491 and output by +0.211, and the condition of the first case of Result 7 holds, since $L(\bar{p})\eta_w(\bar{p})\sigma_w = 0.69 \leq 1$; at $\sigma = \frac{3}{2}$ it changes welfare by +0.530 and output by +0.612, and the condition of the second case holds, since $L(\bar{p})\eta_s(\bar{p})\sigma_s = 1.38 \geq 1$.

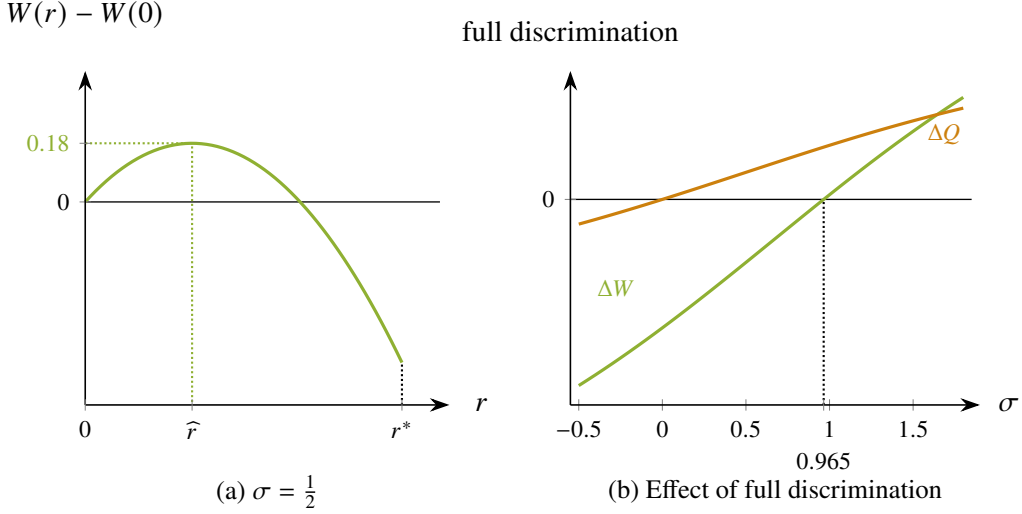


Figure 6: (a) Welfare against the allowed price difference r with a common curvature $\sigma = \frac{1}{2}$; (b) change in welfare (ΔW) and in output (ΔQ) from full discrimination against the common curvature σ of inverse demand; $p_w^* = 3$, $p_s^* = 4$, $q_w^* = q_s^* = 10$, and $c = 1$.

8. Step (g): Constant Elasticities

Constant-elasticity demands $q_i(p) = A_i p^{-\eta_i}$ with $\eta_i > 1$ are a case of constant inverse curvature, with $\sigma_i = 1 + \frac{1}{\eta_i}$, and also of constant direct curvature, with $\alpha_i = 1 + \eta_i$. The weak market has the higher elasticity $\eta_w > \eta_s$, so it has the higher α and the lower σ . Result 5 applies, so welfare first rises and then falls with r , but Results 3, 4, 6, and 7 are silent. The discriminatory prices are

$$p_i^* = \frac{c \eta_i}{\eta_i - 1}. \quad (21)$$

Take $c = 1$, $\eta_s = 2$, $\eta_w = 3$, and A_i such that both markets sell 10 at their discriminatory prices. Then $p_w^* = 1.5$, $p_s^* = 2$, and $\bar{p} = 1.7549$.

The paper's technique is to move along a path of “harmonic mean” prices,

$$\frac{1}{\rho_i(m)} = \frac{m}{p_i^*} + \frac{1-m}{\bar{p}}, \quad 0 \leq m \leq 1, \quad (22)$$

where $m = 0$ is the uniform price and $m = 1$ is full discrimination. Output at these prices is $\tilde{Q}(m) = \sum_i A_i \rho_i(m)^{-\eta_i} = \sum_i A_i \left(\frac{m}{p_i^*} + \frac{1-m}{\bar{p}} \right)^{\eta_i}$, a sum of strictly convex functions of m because $\eta_i > 1$. Its slope at $m = 0$ is zero by the first-order condition of the uniform price, so $\tilde{Q}(1) > \tilde{Q}(0)$ and output rises. Welfare $\tilde{W}(m)$ also has zero slope at $m = 0$, and its second derivative is

$$\tilde{W}''(m) = \sum_i -q'_i(\rho_i) [\rho'_i(m)]^2 [L_i(\rho_i) (\eta_i - 1) - 1], \quad (23)$$

which is negative if $L_i(\rho_i)(\eta_i - 1) \leq 1$ in both markets. This holds in the strong market because $L_s \leq \frac{1}{\eta_s}$ gives $L_s(\eta_s - 1) \leq \frac{\eta_s - 1}{\eta_s} < 1$, and in the weak market if $\eta_w - 1 \leq \frac{1}{L(\bar{p})}$. The first-order condition of the uniform price gives $\frac{1}{L(\bar{p})} = s_w \eta_w + (1 - s_w) \eta_s$, where s_w is the weak market's share of output at $\bar{p}$, so the weak-market condition is $(\eta_w - \eta_s)(1 - s_w) \leq 1$.

Result 8 (Proposition 6 of the paper). With constant-elasticity demands:

- total output is higher with discrimination;
- social welfare is lower with discrimination if $(\eta_w - \eta_s)(1 - s_w) \leq 1$, where $1 - s_w$ is the strong market's share of output at the uniform price, and in particular if $\eta_w - \eta_s \leq 1$.

In the toy, $s_w = 0.325$ and $\eta_w - \eta_s = 1$, so the condition holds, output rises by 0.767, and welfare falls by 0.292. Figure 7 depicts the argument and its limit. Panel (a) shows $\tilde{Q}(m) - \tilde{Q}(0)$ and $\tilde{W}(m) - \tilde{W}(0)$ along the harmonic path for $\eta_w = 3$: output is convex and welfare concave in m , both with zero slope at $m = 0$. Panel (b) shows the effect of full discrimination on welfare against the elasticity gap $\eta_w - \eta_s$, with $\eta_s = 2$ and both markets selling 10 at their discriminatory prices. Welfare falls for gaps up to about 2.57, well beyond the sufficient condition, and rises for larger gaps: with $\eta_w = 5$ output rises by 4.415 and welfare rises by 0.276. The condition of Result 8 is sufficient only, and it can fail when elasticities differ substantially, as Ippolito's (1980) numerical examples illustrate.

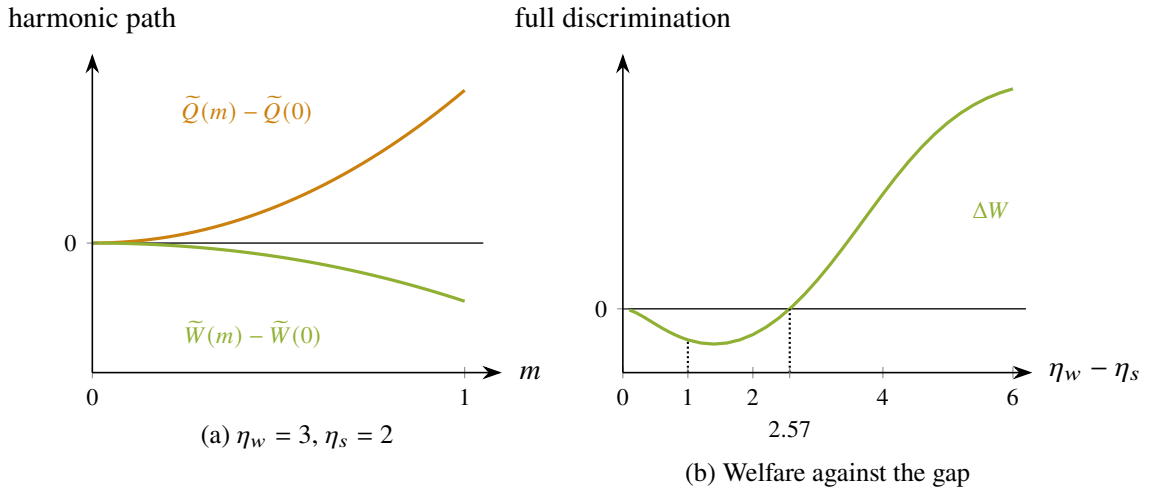


Figure 7: Output and welfare along the harmonic-mean path of prices, and the change in welfare from full discrimination against the elasticity gap, with $c = 1$ and $q_w^* = q_s^* = 10$.

9. Step (h): What the Toy Model Keeps and Drops

The pair of linear and exponential demands, together with the constant-curvature families of Steps (f) and (g), keeps every main conclusion of the paper:

1. Under the IRC, welfare is monotone in the degree of discrimination or has a single interior peak (Step c, Lemma).
2. Discrimination reduces welfare if direct demand in the strong market is at least as convex as in the weak market at the uniform price (Step c, Proposition 1).
3. Welfare is higher with discrimination if inverse demand in the weak market is sufficiently more convex and the discriminatory prices are close (Step d, Proposition 2), and welfare rises and then falls when direct demand is more convex in the weak market but inverse demand is not (Step d, Proposition 3).
4. Output rises if direct and inverse demand are both more convex in the weak market, and does not rise if both are at least as convex in the strong market (Step e, Proposition 4).

5. With constant curvature, the threshold between falling and rising welfare is $\sigma = 1$ when the discriminatory prices are close (Step f, Proposition 5), and with constant elasticities discrimination raises output and lowers welfare when the elasticities are not far apart (Step g, Proposition 6).

The toy also shows what the conditions of the paper leave open. The regime of Results 3 to 5 depends on the scale b of market E and not on its size, and in the intermediate case the sign of the net effect of full discrimination is not fixed by the conditions: it is positive in the linear-exponential pair and negative in the pair of quadratic demands of Step (f).

Relative to Aguirre, Cowan, and Vickers (2010), the toy model uses two markets, and the paper notes that all the results generalize to more than two markets by constraining a weighted sum of the differences between the prices and the uniform price (Schmalensee, 1981), with weights equal to marginal profitability at the uniform price. The toy assumes that both markets are served at the uniform price, so it leaves aside discrimination that opens new markets, and it assumes a monopolist with constant marginal cost and strictly concave profit in each market. Each of these features can be relaxed one at a time.

References

- Aguirre, I., S. Cowan, and J. Vickers (2010). “Monopoly Price Discrimination and Demand Curvature.” *American Economic Review*, 100(4), pp. 1601–1615. <https://doi.org/10.1257/aer.100.4.1601>
- Cowan, S. (2007). “The Welfare Effects of Third-Degree Price Discrimination with Nonlinear Demand Functions.” *RAND Journal of Economics*, 38(2), pp. 419–428.
- Ippolito, R. A. (1980). “Welfare Effects of Price Discrimination When Demand Curves Are Constant Elasticity.” *Atlantic Economic Journal*, 8(2), pp. 89–93.
- Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.
- Schmalensee, R. (1981). “Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination.” *American Economic Review*, 71(1), pp. 242–247.
- Varian, H. R. (1985). “Price Discrimination and Social Welfare.” *American Economic Review*, 75(4), pp. 870–875.

Third-Degree Price Discrimination in Oligopoly When Markets Are Covered

A step-by-step exercise in the spirit of Dertwinkel-Kalt and Wey (2023)

Companion note to Dertwinkel-Kalt, M., and C. Wey (2023), “Third-Degree Price Discrimination in Oligopoly When Markets Are Covered,” *Journal of Industrial Economics* 71(2): 464–490. This note follows the working-paper version, DICE Discussion Paper 336 (2020). Full references at the end.

Overview. A monopolist that discriminates among markets changes welfare through two channels: some consumers stay home when their price rises, and the consumers who buy pay different prices. An oligopoly adds a third: when one firm raises its price, some consumers switch to a rival. Dertwinkel-Kalt and Wey (2023) switch off the first channel by assuming that every consumer buys from some firm in every market, so that market demand is completely inelastic and only the switching channel remains. In their linear covered-demand model, third-degree price discrimination does not change any firm’s output in any market, so it leaves total welfare unchanged, but it raises prices on average. Each firm’s Lerner index under uniform pricing is the weighted harmonic mean of its Lerner indices under discrimination, which is lower than the weighted arithmetic mean, so uniform pricing reduces market power, raises consumer surplus, and lowers every firm’s profit. The consumer surplus gain can be computed from the prices and quantities of the discriminatory equilibrium alone. This note builds the model with two firms and two markets, so that every number is simple. Section 1 sets up the demand system; Step (a) solves the discriminatory equilibrium; Step (b) solves the uniform-pricing equilibrium and shows that outputs do not change; Step (c) derives the arithmetic- and harmonic-mean formulas; Step (d) measures the consumer surplus gain; Step (e) adds arbitrage costs; Step (f) draws three policy implications; and Step (g) lists what the toy model keeps and drops.

1. The Toy Model

Firms and markets. Two firms, 1 and 2, produce differentiated products at constant marginal costs $c_1 = 1$ and $c_2 = 4$, so firm 1 is the low-cost firm. They sell in two independent markets, A and B . In each market the consumers are spread uniformly on a Hotelling line of length one, with firm 1 at the left end and firm 2 at the right end. Every consumer buys exactly one unit from one of the two firms, so the market is *covered*. Each market has a mass of 12 consumers, so the two markets differ only in the transport cost per unit of distance, which is 6 in market A and 3 in market B .

Demand. In market j with mass M_j and transport cost t_j , the consumer who is indifferent between the two firms sits at $x_0 = \frac{1}{2} \left(1 + \frac{p_2^j - p_1^j}{t_j}\right)$, so firm 1 sells $M_j x_0$ and firm 2 sells the rest. Writing $a_j \equiv \frac{M_j}{2}$ and $b_j \equiv \frac{M_j}{2t_j}$,

$$D_1^j = a_j + b_j(p_2^j - p_1^j), \quad D_2^j = a_j + b_j(p_1^j - p_2^j), \quad (1)$$

with $a_A = a_B = 6$, $b_A = 1$, and $b_B = 2$. This is the two-firm case of the paper’s *linear covered-demand* model with m firms and n markets, in which $D_i^j = a_j + b_j \sum_{i' \neq i} (p_{i'}^j - p_i^j)$. The ratio $\frac{b_j}{a_j}$ measures how intensely firms compete in market j : it is high when the transport cost is low, so that consumers switch easily. In this micro-foundation $\frac{a_j}{b_j} = t_j$, so the only difference between the markets is the ratio $\frac{a_A}{b_A} = 6$ against $\frac{a_B}{b_B} = 3$. Four properties of the demand system are used throughout, and each can be checked in (1):

(A1) a firm’s own-price slope is $\frac{\partial D_i^j}{\partial p_i^j} = -(m-1)b_j$, which equals $-b_j$ for $m = 2$;

- (A2) demand depends only on price *differences*: adding the same amount to every price in market j changes no one's demand;
- (A3) market demand is inelastic: $\sum_i D_i^j = m a_j$, which is 12 in both markets;
- (A4) the demand difference between two firms is pinned down by their price difference: $D_i^j - D_{i'}^j = m b_j (p_{i'}^j - p_i^j)$.

Figure 1 depicts the demand of firm 1 against the price difference $p_2^j - p_1^j$ in the two markets, as in (1). At a zero price difference firm 1 sells half of each market, 6 consumers, and each unit of price difference moves b_j consumers to it, so a given price difference moves more consumers in the market with the lower transport cost.

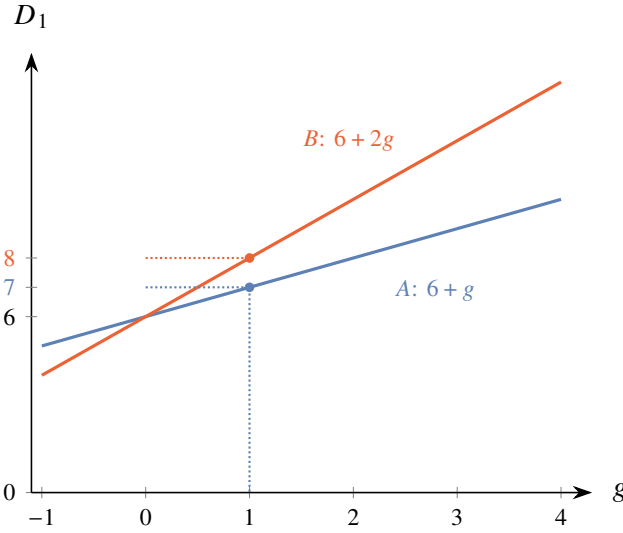


Figure 1: Demand of firm 1 in markets A and B against the price difference $g = p_2^j - p_1^j$; both lines start at 6 when $g = 0$, and the dotted lines mark $g = 1$.

Pricing regimes. Under *discriminatory pricing* each firm sets a price in each market, p_i^j . Under *uniform pricing* each firm must charge the same price in both markets, $\hat{p}_i$, for example because arbitrage or a regulation forces it. Firm i 's profit is $\sum_j D_i^j (p_i^j - c_i)$, firms choose prices simultaneously, and we look at the unique interior Nash equilibrium of each regime.

2. Step (a): The Discriminatory Equilibrium

Under discrimination each firm solves a separate problem in each market, with first-order condition $-(m - 1)b_j (p_i^j - c_i) + D_i^j = 0$, that is,

$$p_i^j - c_i = \frac{D_i^j}{(m - 1)b_j}. \quad (2)$$

The margin of a firm equals its output divided by the slope of its demand, as for any price-setting firm. Subtract (2) for firm i' and for firm i in the same market and use (A4): $-(m - 1)b_j [(p_{i'}^j - p_i^j) - (c_{i'} - c_i)] - m b_j (p_{i'}^j - p_i^j) = 0$, so

$$p_{i'}^j - p_i^j = \frac{m - 1}{2m - 1} (c_{i'} - c_i) \quad \text{in every market } j. \quad (3)$$

With two firms the price difference is one third of the cost difference, $\frac{3}{3} = 1$, in both markets. Using (3) in (A4) and (2), the equilibrium prices and outputs are

$$p_i^j = c_i + \frac{a_j}{(m-1)b_j} + \frac{1}{2m-1} \sum_{i' \neq i} (c_{i'} - c_i), \quad D_i^j = a_j + b_j \frac{m-1}{2m-1} \sum_{i' \neq i} (c_{i'} - c_i). \quad (4)$$

A firm's price is its cost, plus a markup $\frac{a_j}{(m-1)b_j}$ that depends only on the market and falls with the intensity of competition $\frac{b_j}{a_j}$, plus a term that depends on its cost relative to its rivals. In the toy,

$$p_1^A = 1 + 6 + 1 = 8, \quad p_2^A = 4 + 6 - 1 = 9, \quad p_1^B = 1 + 3 + 1 = 5, \quad p_2^B = 4 + 3 - 1 = 6,$$

and the outputs are $D_1^A = 7$, $D_2^A = 5$, $D_1^B = 8$, $D_2^B = 4$. Market A, where consumers switch less readily, is the high-price market for both firms.

Figure 2 depicts the two markets at these prices. The indifferent consumer sits at $x_0 = \frac{7}{12}$ in market A and at $\frac{8}{12} = \frac{2}{3}$ in market B: the low-cost firm serves more than half of each market, and more in the market where consumers switch easily, because there the same price difference moves more consumers.

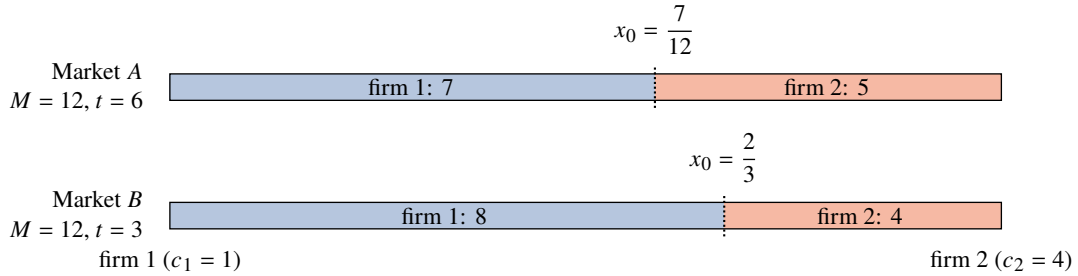


Figure 2: The two markets at the discriminatory prices: each firm sells to the consumers on its side of the indifferent consumer x_0 .

3. Step (b): The Uniform-Pricing Equilibrium

Under uniform pricing firm i chooses one price $\hat{p}_i$ for both markets. Summing its first-order conditions, $\sum_j [- (m-1)b_j (\hat{p}_i - c_i) + D_i^j] = 0$, so

$$\hat{p}_i - c_i = \frac{\sum_j D_i^j}{(m-1) \sum_j b_j}. \quad (5)$$

The same subtraction as in Step (a), applied to the summed conditions, gives $\hat{p}_{i'} - \hat{p}_i = \frac{m-1}{2m-1} (c_{i'} - c_i)$: the price difference between two firms is the same under uniform pricing as under discrimination. By (A2), outputs depend only on price differences, so they are the same under both regimes.

Result 1 (Proposition 1 of the paper). In the linear covered-demand model with constant marginal costs:

- all bilateral price differences are the same under discriminatory and uniform pricing, $p_{i'}^j - p_i^j = \hat{p}_{i'} - \hat{p}_i = \frac{m-1}{2m-1} (c_{i'} - c_i)$, in every market;
- every firm's output in every market is the same under the two regimes, $D_i^j = a_j + b_j \frac{m-1}{2m-1} \sum_{i' \neq i} (c_{i'} - c_i)$.

Intuitively, the regime moves the *level* of each firm's prices, through its markup, but not the gaps between firms, and in a covered market only the gaps matter. The price difference $\frac{m-1}{2m-1}$ of the cost difference rises from $\frac{1}{3}$ with two firms toward $\frac{1}{2}$ as the number of firms grows. Since outputs and who serves whom do not change, total welfare does not change either: the market is covered, so no consumer is lost, and each consumer travels the same distance to the same firm.

Table 1 reports the two equilibria. The uniform prices are $\hat{p}_1 = 1 + \frac{15}{3} = 6$ and $\hat{p}_2 = 4 + \frac{9}{3} = 7$, whose difference is 1 as in (3), and each firm earns less than under discrimination.

Firm	Market	p_i^j	$\widehat{p}_i$	D_i^j	profit, discrimination	profit, uniform
1	A	8	6	7	$7 \cdot 7 = 49$	$15 \cdot 5 = 75$
	B	5		8	$8 \cdot 4 = 32$	
total for firm 1					81	75
2	A	9	7	5	$5 \cdot 5 = 25$	$9 \cdot 3 = 27$
	B	6		4	$4 \cdot 2 = 8$	
total for firm 2					33	27

Table 1: Discriminatory and uniform equilibria, with $c_1 = 1$ and $c_2 = 4$; outputs are the same under both regimes.

4. Step (c): Elasticities and Lerner Indices

Define firm i 's elasticity in market j under discrimination and its aggregate elasticity under uniform pricing,

$$E_i^j \equiv -\frac{\partial D_i^j}{\partial p_i^j} \frac{p_i^j}{D_i^j}, \quad \hat{E}_i \equiv -\frac{\sum_j \frac{\partial D_i^j}{\partial \hat{p}_i} \hat{p}_i}{\sum_j D_i^j}, \quad (6)$$

and the Lerner indices $L_i^j \equiv \frac{p_i^j - c_i}{p_i^j} = \frac{1}{E_i^j}$ and $\hat{L}_i \equiv \frac{\hat{p}_i - c_i}{\hat{p}_i} = \frac{1}{\hat{E}_i}$, which the first-order conditions equate to the inverse elasticities. Let $w_i^j \equiv \frac{D_i^j}{\sum_j D_i^j}$ be the share of firm i 's output sold in market j . By Result 1 the weights are the same under both regimes. Summing (2) over markets and comparing with the summed uniform condition, $\sum_j (m-1)b_j(\hat{p}_i - c_i) = \sum_j (m-1)b_j(p_i^j - c_i)$, which says that $\hat{p}_i = \frac{\sum_j b_j p_i^j}{\sum_j b_j}$: the uniform price is the weighted arithmetic mean of the discriminatory prices with the weights b_j . Since the two sides of that equality are $\hat{E}_i \sum_j D_i^j$ and $\sum_j E_i^j D_i^j$ (multiply (6) through), dividing by $\sum_j D_i^j$ gives the arithmetic-mean formula for elasticities, and inverting, with $L = \frac{1}{E}$, gives the harmonic-mean formula for Lerner indices.

Result 2 (Proposition 2 of the paper). For every firm i :

- (arithmetic mean formula) the aggregate elasticity under uniform pricing is the weighted arithmetic mean of the market elasticities, $\hat{E}_i = \sum_j w_i^j E_i^j$;
- (harmonic mean formula) the Lerner index under uniform pricing is the weighted harmonic mean of the market Lerner indices,

$$\hat{L}_i = \frac{1}{\sum_j w_i^j \frac{1}{L_i^j}}; \quad (7)$$

- since a harmonic mean is below the arithmetic mean unless all the numbers are equal, $\widehat{L}_i < \sum_j w_i^j L_i^j$ whenever the Lerner indices differ across markets, so uniform pricing reduces market power;
- under the same condition, the uniform price is below the quantity-weighted mean of the discriminatory prices, $\widehat{p}_i < \sum_j w_i^j p_i^j$.

The Lerner indices of firm i differ across markets if and only if its discriminatory prices differ, that is, if and only if the ratio $\frac{a_i}{b_j}$ differs across markets (given $c_i > 0$, since a firm with zero cost has a Lerner index of one in every market), by (4).

In the toy, the elasticities of firm 1 are $E_1^A = \frac{1 \cdot 8}{7} = 1.143$ and $E_1^B = \frac{2 \cdot 5}{8} = 1.250$, and its aggregate elasticity under uniform pricing is $\widehat{E}_1 = \frac{3 \cdot 6}{15} = 1.2 = \frac{7}{15} \cdot 1.143 + \frac{8}{15} \cdot 1.250$. Its Lerner indices are $L_1^A = \frac{7}{8}$ and $L_1^B = \frac{4}{5}$, and

$$\widehat{L}_1 = \frac{1}{\frac{7}{15} \cdot \frac{8}{7} + \frac{8}{15} \cdot \frac{5}{4}} = \frac{5}{6} = 0.8333 = \frac{6-1}{6},$$

below the arithmetic mean $\frac{7}{15} \cdot \frac{7}{8} + \frac{8}{15} \cdot \frac{4}{5} = 0.8350$. For firm 2, $E_2^A = \frac{9}{5} = 1.8$ and $E_2^B = \frac{2 \cdot 6}{4} = 3$, so $\widehat{E}_2 = \frac{3 \cdot 7}{9} = \frac{7}{3} = 2.333$, and $L_2^A = \frac{5}{9}$ and $L_2^B = \frac{1}{3}$ give $\widehat{L}_2 = \frac{3}{7} = 0.429 = \frac{7-4}{7}$, below the arithmetic mean $\frac{37}{81} = 0.457$.

Figure 3 depicts the harmonic mean formula for firm 2, with the share w of its output sold in market A on the horizontal axis. The arithmetic mean $wL^A + (1-w)L^B$ is the straight line between the two Lerner indices, and the harmonic mean is the curve below it. At the firm's actual share $w = \frac{5}{9}$ the two means are 0.457 and 0.429, and the harmonic mean is the Lerner index $\widehat{L}_2$ that the firm earns under uniform pricing. The gap is larger the more the market Lerner indices differ.

Lerner index of firm 2

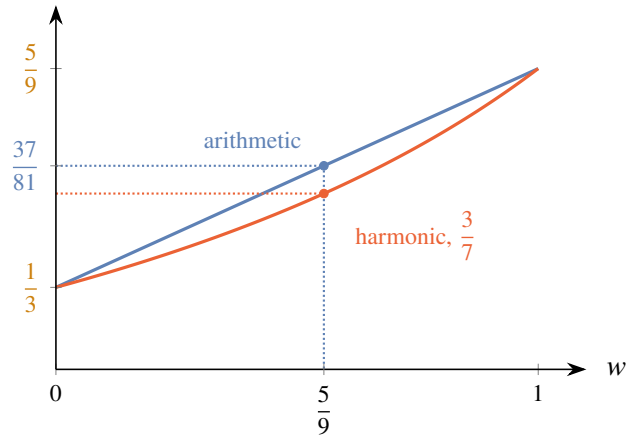


Figure 3: Weighted arithmetic and harmonic means of firm 2's market Lerner indices $L^B = \frac{1}{3}$ and $L^A = \frac{5}{9}$ against the share w of its output sold in market A.

Figure 4 depicts the prices of the two firms on a price axis. Each uniform price lies between the firm's two discriminatory prices, closer to the low-price market, and below the quantity-weighted mean of the discriminatory prices (6.4 for firm 1 and 7.667 for firm 2). The uniform price is the arithmetic mean with the weights b_j , $\frac{1 \cdot 8 + 2 \cdot 5}{3} = 6$, and the quantity-weighted mean gives more weight to the high-price market because the firm sells more there for each unit of b_j .

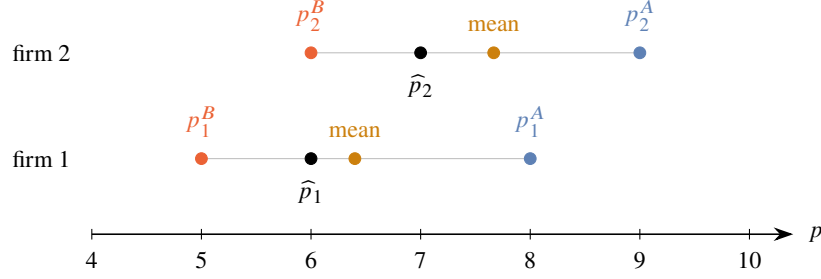


Figure 4: Discriminatory prices in markets B and A , uniform price $\hat{p}_i$, and the quantity-weighted mean of the discriminatory prices (mean), for each firm.

5. Step (d): The Consumer Surplus Gain from Uniform Pricing

Since outputs and the allocation of consumers to firms do not change, total welfare is the same under the two regimes, so the gain in consumer surplus equals the loss in total profit,

$$\widehat{CS} - CS = \sum_i \pi_i - \sum_i \hat{\pi}_i = \sum_i \sum_j (p_i^j - \hat{p}_i) D_i^j, \quad (8)$$

where the last equality uses the equal outputs. In the toy, firm 1 loses $81 - 75 = 6$, which equals $(8 - 6)7 + (5 - 6)8 = 14 - 8$, and firm 2 loses $33 - 27 = 6 = (9 - 7)5 + (6 - 7)4$, so the consumer surplus gain is 12. The firms' profit loss is the same for both firms, an equality that turns out to be general (see below).

The paper's second corollary writes the gain with the discriminatory data alone. By (2), a firm's profit under discrimination is $\pi_i = \sum_j \frac{(D_i^j)^2}{(m-1)b_j}$, and by (5) its profit under uniform pricing is $\hat{\pi}_i = \frac{(\sum_j D_i^j)^2}{(m-1)\sum_j b_j}$, so

$$\widehat{CS} - CS = \frac{1}{m-1} \sum_i \left[\sum_j \frac{(D_i^j)^2}{b_j} - \frac{(\sum_j D_i^j)^2}{\sum_j b_j} \right], \quad (9)$$

where b_j can be recovered from observables by (A4): in the toy, $b_A = \frac{7-5}{2(9-8)} = 1$ and $b_B = \frac{8-4}{2(6-5)} = 2$. The bracket is nonnegative by the Cauchy-Schwarz inequality, and it is zero only if $\frac{D_i^j}{b_j}$ is the same in all markets. For firm 1 it equals $\frac{7^2}{1} + \frac{8^2}{2} - \frac{15^2}{3} = 49 + 32 - 75 = 6$.

Result 3 (Corollaries 1 and 2 of the paper). Relative to uniform pricing, discrimination:

- lowers consumer surplus by $\sum_i \sum_j (p_i^j - \hat{p}_i) D_i^j > 0$, which equals the rise in total profit, and leaves total welfare unchanged;
- raises every firm's profit, so the discriminatory equilibrium is a Pareto improvement for the firms;
- produces a consumer surplus loss that can be computed from the prices and quantities of the discriminatory equilibrium alone, with (9).

Reading the formula. Since $D_i^j = b_j \left(\frac{a_j}{b_j} + \Delta_i \right)$, where $\Delta_i \equiv \frac{m-1}{2m-1} \sum_{i' \neq i} (c_{i'} - c_i)$ is a firm-specific constant, the term Δ_i cancels from the bracket in (9), and the bracket becomes a variance across markets of the ratio $\frac{a_j}{b_j}$ with the weights b_j :

$$\widehat{CS} - CS = \frac{m}{m-1} \sum_j b_j \left(\frac{a_j}{b_j} - \bar{r} \right)^2, \quad \bar{r} \equiv \frac{\sum_j a_j}{\sum_j b_j}. \quad (10)$$

Every firm loses the same amount, $\frac{1}{m-1} \sum_j b_j \left(\frac{a_j}{b_j} - \bar{r}\right)^2$, whatever its cost, and the gain depends on the demand parameters only through the dispersion of $\frac{a_j}{b_j}$ across markets, that is, through how much the discriminatory prices differ across markets. With two markets, $\widehat{CS} - CS = \frac{m}{m-1} \frac{b_A b_B}{b_A + b_B} \left(\frac{a_A}{b_A} - \frac{a_B}{b_B}\right)^2 = 2 \cdot \frac{2}{3} \cdot 3^2 = 12$. Figure 5 depicts this relation against the difference $s = \frac{a_A}{b_A} - \frac{a_B}{b_B}$ between the two markets, which is also the difference $p^A - p^B$ between the discriminatory prices of each firm with two firms, and the difference $t_A - t_B$ between the transport costs. The gain is zero when the markets have the same ratio and grows with the square of the difference.

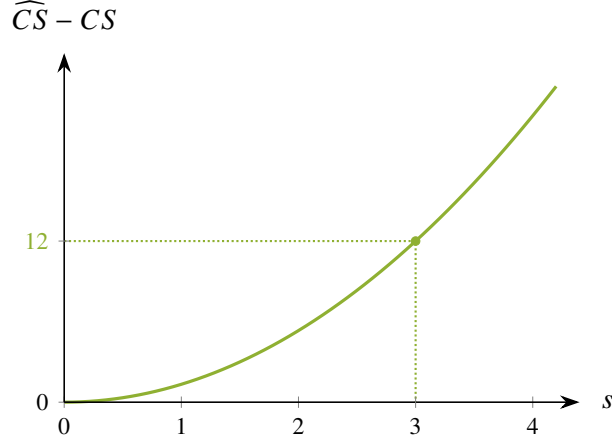


Figure 5: Consumer surplus gain from uniform pricing against the difference $s = \frac{a_A}{b_A} - \frac{a_B}{b_B}$, with $b_A = 1$ and $b_B = 2$.

Recovering the uniform price from market data. The discriminatory data also give the uniform price. From $\pi_i - \widehat{\pi}_i = \sum_j (p_i^j - \widehat{p}_i) D_i^j$ and (9),

$$\widehat{p}_i = \frac{\sum_j p_i^j D_i^j - (\pi_i - \widehat{\pi}_i)}{\sum_j D_i^j},$$

which for firm 1 gives $\frac{96-6}{15} = 6$ and for firm 2 gives $\frac{69-6}{9} = 7$. No marginal cost and no demand estimation is needed, only prices and quantities.

6. Step (e): Arbitrage Costs

Discrimination is feasible only if consumers cannot arbitrage between markets. Suppose they can, at a cost $r \geq 0$ per unit, so that each firm's price difference between the two markets cannot exceed r : $p_i^A - p_i^B \leq r$. Without the constraint, each firm's price difference is $r^* = p_i^A - p_i^B = 3$, the same for both firms (it equals $\frac{1}{m-1} \left(\frac{a_A}{b_A} - \frac{a_B}{b_B}\right)$), and for $r \geq r^*$ the constraint does not bind. For $r < r^*$ the constraint binds for both firms.

Result 4 (Proposition 3 of the paper). With two firms and two markets and a binding arbitrage constraint $r < r^*$, the equilibrium prices are

$$p_i^A(r) = \widehat{p}_i + \vartheta r, \quad p_i^B(r) = \widehat{p}_i - (1 - \vartheta) r, \quad \vartheta \equiv \frac{b_B}{b_A + b_B}, \quad (11)$$

and all price differences between firms and all outputs remain the same as under unconstrained discrimination and uniform pricing.

The prices (11) satisfy the constraint with equality, and a firm that moves both prices by the same amount leaves the price difference unchanged. The firm's first-order condition along the constraint is the uniform-pricing condition plus a term proportional to $\vartheta b_A - (1 - \vartheta) b_B$, which is zero at $\vartheta = \frac{b_B}{b_A + b_B}$. In the toy, $\vartheta = \frac{2}{3}$, so firm 1 charges $6 + \frac{2}{3}r$ in market A and $6 - \frac{1}{3}r$ in market B , and firm 2 charges $7 + \frac{2}{3}r$ and $7 - \frac{1}{3}r$; at $r = 3$ the prices are the discriminatory prices 8 and 5 (and 9 and 6). The arbitrage cost splits the price movement between the markets in proportion to b_B and b_A : the market where consumers switch more readily, B , absorbs the smaller price cut. The quantity-weighted average price of each firm rises with r , so the consumer surplus gain from a tighter constraint is

$$\widehat{CS} - CS(r) = 2 [\vartheta D_i^A - (1 - \vartheta) D_i^B] r = 4r \quad (r \leq 3),$$

where each firm's bracket equals $\frac{2 \cdot 7 - 1 \cdot 8}{3} = 2$ for firm 1 and $\frac{2 \cdot 5 - 1 \cdot 4}{3} = 2$ for firm 2.

Figure 6 depicts firm 1's prices in panel (a) and the consumer surplus gain in panel (b), against the arbitrage cost r . The prices fan out from the uniform price 6 at $r = 0$ to the discriminatory prices at $r^* = 3$ and stay there. The gain from a given arbitrage cost is linear in r , so any policy that makes arbitrage easier, such as integrating markets, lowers r and raises consumer surplus.

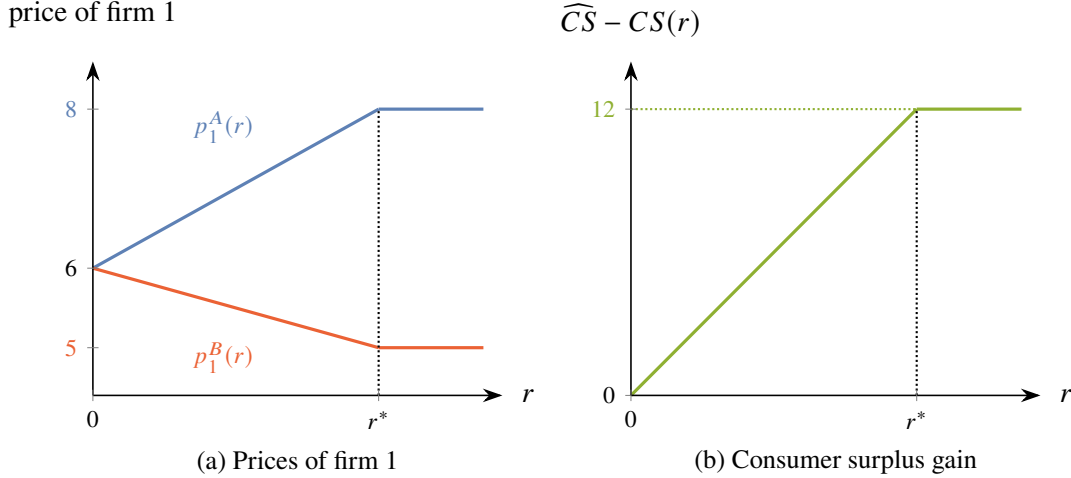


Figure 6: Prices of firm 1 and consumer surplus gain relative to uniform pricing against the arbitrage cost r .

7. Step (f): Three Policy Implications

Uniform pricing raises consumer surplus at no welfare cost. Consumers lose 12 from discrimination in the toy, and the firms gain exactly that, so a rule that forces uniform prices moves surplus from firms to consumers. The authors stress that this matters for antitrust authorities, whose objective is typically consumer surplus. Because the market is covered, uniform pricing costs nothing in welfare, in contrast with a monopolist, where the welfare effect of discrimination depends on the curvature of demand (Aguirre, Cowan, and Vickers, 2010).

Firms have a collective interest in segmenting markets. Each firm earns 6 more under discrimination, so discrimination is a Pareto improvement for the firms. Firms may therefore want to segment markets and prevent arbitrage, for example through geo-blocking. The result relies on all firms agreeing on which market is the high-price market, as they do here: the high-price market is A for both firms. The paper's survey

of the literature notes that when firms disagree on where the high price belongs, discrimination intensifies competition and firms are in a prisoners' dilemma; the toy model rules this out.

Market integration is a consumer-friendly policy. By Result 4, a lower arbitrage cost raises consumer surplus linearly. Policies that make arbitrage easy, such as the Single Market in the European Union, or that ban geo-blocking, work through this channel. If markets are fully integrated, discrimination is impossible and each firm must charge one price, which is the uniform-pricing equilibrium of Step (b).

8. Step (g): What the Toy Model Keeps and Drops

The toy model keeps every main conclusion of the paper:

1. Price differences between firms are the same under discrimination and uniform pricing, and outputs do not change with the regime (Step (b), Proposition 1).
2. The uniform-price elasticity is the weighted arithmetic mean, and the uniform-price Lerner index the weighted harmonic mean, of the market-specific values, so uniform pricing lowers every Lerner index and every average price (Step (c), Proposition 2).
3. Consumers gain, firms lose, and the gain can be computed from observables (Step (d), Corollaries 1 and 2).
4. With arbitrage costs, prices move linearly with r and outputs do not change (Step (e), Proposition 3).

The toy also shows two things that the general formulas leave implicit. First, all firms lose the same amount from uniform pricing, and the loss depends on the demand parameters only through the dispersion of $\frac{a_j}{b_j}$ across markets. Second, the strictness of the harmonic-mean inequality does not depend on whether the firms are symmetric: it requires only that the discriminatory prices differ across markets, that is, that $\frac{a_j}{b_j}$ differ.

Relative to Dertwinkel-Kalt and Wey (2023), the toy model has two firms and two markets instead of m firms and n markets, a Hotelling micro-foundation with one line per market, and constant marginal costs. As in the paper, market demand is completely inelastic, so the model isolates the switching channel and cannot say anything about the stay-home channel that drives the welfare effects in monopoly. Products are symmetrically substitutable, demands are linear, firms agree on which market has the higher price, and the equilibria are unique and interior. The paper notes that its results extend to market-specific delivery costs c^j that affect all firms equally, which leave the price differences unchanged, and to loyal consumers who always buy from one firm, which raise a_j but not b_j . Each of these features can be relaxed one at a time.

References

- Aguirre, I., S. Cowan, and J. Vickers (2010). "Monopoly Price Discrimination and Demand Curvature." *American Economic Review*, 100(4), pp. 1601–1615.
- Dertwinkel-Kalt, M., and C. Wey (2023). "Third-Degree Price Discrimination in Oligopoly When Markets Are Covered." *Journal of Industrial Economics*, 71(2), pp. 464–490. Working-paper version: DICE Discussion Paper No. 336 (2020), Heinrich Heine University Düsseldorf, <https://hdl.handle.net/10419/216730>.

Two-Part Tariffs with Two Firms

Competition in nonlinear pricing on a Hotelling line

Based on Armstrong and Vickers (2001); textbook treatment in Belleflamme and Peitz (2015, §9.3.2). Full references at the end.

1. The Model

We study two firms that compete in *two-part tariffs*. A two-part tariff asks a consumer who buys q units of the good to pay a fixed fee m_i regardless of the quantity, plus a constant per-unit price p_i for each unit consumed, so that the total payment to firm i is

$$T_i(q) = m_i + p_i q.$$

This is a simple form of nonlinear pricing: the average price paid per unit, $\frac{T_i(q)}{q} = p_i + \frac{m_i}{q}$, falls as the consumer buys more. Our question is what such a tariff looks like when firms compete, and how the outcome compares with ordinary (linear) pricing in which firms may set only the per-unit price.

The two firms sell horizontally differentiated products and sit at the two ends of a Hotelling line of unit length. Firm 1 is located at 0 and firm 2 at 1. A unit mass of consumers is spread uniformly along the line; a consumer at location $x \in [0, 1]$ incurs a “transportation” (mismatch) cost τ per unit of distance to the firm she buys from, where $\tau > 0$ measures the degree of product differentiation. Both firms produce at the same constant marginal cost c . Figure 1 depicts the market.

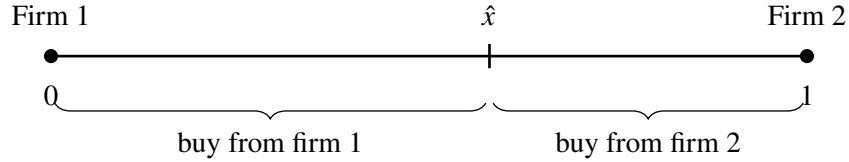


Figure 1: The Hotelling line. The consumer at $\hat{x}$ is indifferent between the two firms; consumers to her left buy from firm 1 and those to her right from firm 2.

The game has two stages, which we solve by backward induction:

Stage 1. Firms simultaneously choose their two-part tariffs (p_i, m_i) , $i = 1, 2$.

Stage 2. Each consumer observes both tariffs, chooses the firm to patronize, and then decides how many units to buy.

We start with Stage 2 (the consumers’ problem) and then move back to Stage 1 (the firms’ problem).

2. Stage 2: The Consumers

Consider a consumer who has decided to buy from a firm charging per-unit price p_i . She chooses her quantity q to maximize the utility she gets from consuming the good, $u(q)$, net of the per-unit payment $p_i q$. Her *indirect utility* is the value of this problem,

$$v(p_i) \equiv \max_q \{ u(q) - p_i q \}, \quad (1)$$

where $u(q)$ is increasing and concave. The first-order condition of (1) is

$$u'(q) = p_i,$$

which defines the consumer's *Walrasian demand* $q(p_i)$: she keeps buying units as long as the marginal utility of the good exceeds its per-unit price, and stops where the two are equal.

Roy's identity. A fact we use repeatedly is that the derivative of the indirect utility with respect to the per-unit price equals minus the quantity demanded,

$$v'(p_i) = -q(p_i). \quad (2)$$

This follows from the envelope theorem applied to (1): because q is chosen optimally, a marginal change in p_i affects v only through the direct term $-p_i q$, whose derivative is $-q(p_i)$. Intuitively, a one-dollar rise in the per-unit price costs the consumer exactly one dollar on each of the $q(p_i)$ units she was already buying.

Bringing back the fixed fee m_i , the net surplus a consumer obtains from firm i 's tariff is her indirect utility from the variable good minus the fixed fee,

$$w(p_i, m_i) = \underbrace{v(p_i)}_{\text{surplus from } q(p_i) \text{ units}} - \underbrace{m_i}_{\text{fixed fee}}.$$

To keep the notation compact, write $w_1 \equiv w(p_1, m_1)$ and $w_2 \equiv w(p_2, m_2)$ for the two “deals” on offer.

A consumer located at x therefore enjoys total utility

$$\begin{aligned} r - \tau x + w_1 & \quad \text{if she buys from firm 1,} \\ r - \tau(1 - x) + w_2 & \quad \text{if she buys from firm 2,} \end{aligned}$$

where r is the (large) gross value of consuming the product category at all. We assume the market is *covered*, so every consumer buys from one of the two firms; the constant r is then common to both options and drops out of every comparison below.

The indifferent consumer. The consumer $\hat{x}$ who is just indifferent between the two firms solves

$$r - \tau \hat{x} + w_1 = r - \tau(1 - \hat{x}) + w_2.$$

Cancelling r and solving for $\hat{x}$,

$$\begin{aligned} w_1 - w_2 + \tau &= 2\tau \hat{x} \\ \hat{x} &= \frac{\tau + w_1 - w_2}{2\tau}. \end{aligned} \quad (3)$$

Everyone to the left of $\hat{x}$ prefers firm 1 and everyone to the right prefers firm 2. Because consumers are uniformly spread on a line of unit mass, firm 1's demand is exactly $\hat{x}$ and firm 2's demand is $1 - \hat{x}$:

$$D_1 = \hat{x} = \frac{\tau + w_1 - w_2}{2\tau}, \quad D_2 = 1 - \hat{x} = \frac{\tau + w_2 - w_1}{2\tau}.$$

A firm attracts more consumers by offering a better deal (a larger w_i), whether it improves that deal through a lower per-unit price or a lower fixed fee.

3. Stage 1: The Firms

Take firm 1. Each of its $\hat{x}$ customers buys $q(p_1)$ units, on which the firm earns the per-unit margin $p_1 - c$, and each pays the fixed fee m_1 . Firm 1 therefore solves

$$\max_{p_1, m_1} \pi_1 = \underbrace{\hat{x}}_{\text{mass of buyers}} \cdot \left[\underbrace{(p_1 - c) q(p_1)}_{\text{variable margin}} + \underbrace{m_1}_{\text{fixed fee}} \right]. \quad (4)$$

It is cleaner to let the firm choose the *deal* w_1 it offers rather than the fee m_1 directly. Since $w_1 = v(p_1) - m_1$, we have $m_1 = v(p_1) - w_1$; substituting this and (3) into (4) turns the problem into a choice of the per-unit price p_1 and the deal w_1 ,

$$\max_{p_1, w_1} \pi_1 = \underbrace{\frac{\tau + w_1 - w_2}{2\tau}}_{\hat{x}} \cdot \left[(p_1 - c) q(p_1) + \underbrace{v(p_1) - w_1}_{m_1} \right]. \quad (5)$$

This change of variables is what makes the model tractable, because the market share $\hat{x}$ now depends on w_1 but *not* on the per-unit price p_1 . The two instruments split their jobs: p_1 shapes the size of the surplus firm 1 creates with each of its customers, while w_1 determines how many customers it wins.

First-order condition in p_1 . Because $\hat{x}$ does not depend on p_1 , differentiating (5) with respect to p_1 and dividing by $\hat{x} > 0$ leaves only the bracketed term,

$$q(p_1) + q'(p_1)(p_1 - c) + v'(p_1) = 0.$$

Now use Roy's identity (2), $v'(p_1) = -q(p_1)$. The first and last terms cancel,

$$\underbrace{q(p_1) + q'(p_1)(p_1 - c)}_{v'(p_1)} - \underbrace{q(p_1)}_{<0 \text{ (law of demand)}} = 0 \implies \underbrace{q'(p_1)}_{<0} (p_1 - c) = 0.$$

Since $q'(p_1) < 0$, the only way this can hold is $p_1 - c = 0$, that is,

$$p_1^* = c \quad (\text{and, by the same argument, } p_2^* = c). \quad (6)$$

Interpretation. Each firm sets its per-unit price equal to marginal cost, so consumption is efficient. The reason is that the per-unit price does not help firm 1 win customers (that is the job of w_1); it only sets how much surplus the firm generates with each customer it already has. That surplus, variable profit plus the consumer's own gross surplus, is maximized when the good is priced at marginal cost, and the firm then collects its share of it through the fixed fee. Marginal-cost pricing makes the pie as large as possible; the fee decides how it is split.

First-order condition in w_1 . Differentiating (5) with respect to w_1 gives two terms, one from the market share $\hat{x}$ (which rises at rate $\frac{1}{2\tau}$) and one from the fee inside the bracket (which falls one-for-one with w_1):

$$\frac{1}{2\tau} \left[(p_1 - c) q(p_1) + v(p_1) - w_1 \right] - \frac{\tau + w_1 - w_2}{2\tau} = 0.$$

Multiplying through by 2τ (and recognizing $2\tau \hat{x} = \tau + w_1 - w_2$),

$$\left[(p_1 - c) q(p_1) + v(p_1) - w_1 \right] - (\tau + w_1 - w_2) = 0. \quad (7)$$

Impose the optimal per-unit price $p_1 = c$ from (6), so the variable margin $(c - c) q(p_1) = 0$ vanishes, and invoke symmetry, $w_1 = w_2$, so that $w_1 - w_2 = 0$:

$$\left[\underbrace{(c - c) q(p_1)}_0 + v(p_1) - w_1 \right] - (\tau + \underbrace{w_1 - w_2}_0) = 0 \implies v(p_1) - w_1 - \tau = 0.$$

Finally recall that $w_1 = v(p_1) - m_1$, i.e. $v(p_1) - w_1 = m_1$, so the last equality reads $m_1 - \tau = 0$, i.e.

$$m_1^* = \tau \quad (\text{and, by the same argument, } m_2^* = \tau). \quad (8)$$

The equilibrium tariff. Putting (6) and (8) together, each firm charges the two-part tariff

$$T(q) = \underbrace{\tau}_{\text{fixed fee}} + \underbrace{c}_{\text{price per unit}} q.$$

Competition drives the per-unit price down to marginal cost, and the firms make their money entirely on the fixed fee, which they can raise up to the value of the differentiation parameter τ . The fee is the reward for being differentiated: the more differentiated the two products (the larger τ), the softer the competition and the higher the fee each firm can sustain.

Equilibrium profits. At the symmetric equilibrium each firm serves half the market, $\hat{x} = \frac{1}{2}$, earns a zero variable margin, and collects the fee $m^* = \tau$ from each customer:

$$\pi_1 = \underbrace{\frac{\tau + w_1 - w_2}{2\tau}}_{=\frac{1}{2}} \left[\underbrace{(p_1 - c) q_1}_0 + \underbrace{m_1}_\tau \right] = \frac{\tau}{2}. \quad (9)$$

Adding the two firms, industry profit under the two-part (nonlinear) tariff is

$$\pi_{NL} = \tau.$$

We keep this number in mind: it is the benchmark against which linear pricing is measured.

4. A Benchmark: Linear Pricing

Consider instead a setting of ordinary *linear pricing*, in which firms may charge only a per-unit price, with no fixed fee ($m_i = 0$). The consumer's deal is simply $w_i = v(p_i)$, and firm i 's profit per customer is the variable margin $\pi_i(p_i) \equiv (p_i - c) q(p_i)$. We want to compare the two regimes for a general demand curve $q(\cdot)$, so a small change of variables helps.

Let firm j charge the candidate symmetric price p , and write the rival's price p_i as a point between marginal cost and p ,

$$p_i = c + k(p - c), \quad k \in [0, 1].$$

The number k measures how far the price sits above marginal cost: $k = 0$ means $p_i = c$ and $k = 1$ means $p_i = p$. Define the consumer's surplus as a function of this index,

$$s(k) \equiv v(c + k(p - c)), \quad \text{so that} \quad \begin{cases} s(0) = v(c) & \text{"low price" (all surplus to the consumer),} \\ s(1) = v(p) & \text{"high price."} \end{cases} \quad (10)$$

Differentiating s and using Roy's identity (2),

$$s_k(k) \equiv \frac{\partial s(k)}{\partial k} = (p - c) v'(c + k(p - c)) = -(p - c) q(c + k(p - c)) < 0. \quad (11)$$

The consumer is worse off as the price rises, so $s_k < 0$.

The point of this parametrization is that the firm's per-customer profit can be written in terms of s_k . At price $p_i = c + k(p - c)$,

$$\pi_i(c + k(p - c)) = \underbrace{[c + k(p - c) - c]}_{=k(p-c)} q(c + k(p - c)) = k(p - c) q(c + k(p - c)) = -k s_k(k), \quad (12)$$

where the last equality just reads off (11). In particular, at $k = 1$ the per-customer (equivalently, unit-mass industry) profit under price p is

$$\pi_L \equiv \pi(p) = -s_k(1). \quad (13)$$

The firm's problem. Facing the rival's price p , firm i chooses its own price, i.e. chooses k , to solve

$$\max_k \pi_i = \frac{\tau + s(k) - s(1)}{2\tau} \cdot (-k s_k(k)), \quad (14)$$

where we used $v(p_i) = s(k)$ and $v(p) = s(1)$ in the market share (3), and the per-customer profit (12). At a symmetric equilibrium the rival's price is a best response to itself, so $p_i = p$, i.e. $k = 1$.

First-order condition in k . Differentiating the product in (14) and setting it to zero,

$$-k (s_k(k))^2 - [\tau + s(k) - s(1)] [s_k(k) + k s_{kk}(k)] = 0. \quad (15)$$

Evaluate this at the symmetric equilibrium $k = 1$, where $s(1) - s(1) = 0$ and the first bracket collapses to τ ,

$$\begin{aligned} -[s_k(1)]^2 - \tau [s_k(1) + s_{kk}(1)] &= 0 \\ \implies -s_k(1) &= s_{kk}(1) + \frac{1}{\tau} [s_k(1)]^2. \end{aligned} \quad (16)$$

Using $\pi_L = -s_k(1)$ from (13), and hence $[s_k(1)]^2 = \pi_L^2$, equation (16) says that industry profit under linear pricing satisfies

$$\pi_L = s_{kk}(1) + \frac{1}{\tau} \pi_L^2. \quad (17)$$

The sign of s_{kk} . Differentiate (11) once more,

$$s_{kk}(k) = -(p - c) q'(c + k(p - c)) \cdot (p - c) = \underbrace{-(p - c)^2}_{>0} \underbrace{q'(c + k(p - c))}_{<0 \text{ (law of demand)}} > 0. \quad (18)$$

So $s_{kk}(1) > 0$: the consumer's surplus falls at an increasing rate as the price climbs. Feeding this into (17),

$$\pi_L = \underbrace{s_{kk}(1)}_{>0} + \frac{1}{\tau} \pi_L^2 > \frac{1}{\tau} \pi_L^2 \implies \pi_L > \frac{1}{\tau} \pi_L^2 \implies \tau > \pi_L,$$

where the last step divides by $\pi_L > 0$. Since industry profit under the two-part tariff is $\pi_{NL} = \tau$, we have established the first comparison,

$$\pi_{NL} = \tau > \pi_L. \quad (19)$$

Result 1. Industry profits are higher under two-part tariffs than under linear pricing. Even though competition forces the per-unit price down to marginal cost, the fixed fee lets firms recover more than they could through a single linear price. Nonlinear pricing softens competition.

5. Welfare

Turn to total welfare, the sum of producer surplus and consumer surplus. At price $p_i = c + k(p - c)$, welfare per consumer is the firm's variable profit plus the consumer's own surplus (the fixed fee is a pure transfer and cancels),

$$W(k) = \underbrace{\pi(c + k(p - c))}_{= -k s_k(k) \text{ (PS)}} + \underbrace{v(c + k(p - c))}_{= s(k) \text{ (CS)}} = -k s_k(k) + s(k). \quad (20)$$

Evaluate W in the two regimes.

- **Under the two-part tariff (NL):** the per-unit price is c , so $k = 0$, and

$$W(0) = \underbrace{-0 \cdot s_k(0)}_0 + s(0) = s(0) = v(c) \equiv W_{NL}.$$

Welfare is maximized here, since pricing at marginal cost yields the efficient quantity.

- **Under linear pricing (L):** the price is p , so $k = 1$, and

$$W(1) = -1 \cdot s_k(1) + s(1) = s(1) - s_k(1) \equiv W_L.$$

To rank the two, note from (20) that

$$W'(k) = -s_k(k) - k s_{kk}(k) + s_k(k) = -k s_{kk}(k),$$

so $W'(0) = 0$ and $W'(k) < 0$ for $k > 0$ (because $s_{kk} > 0$): welfare peaks at $k = 0$ and falls thereafter. Assuming, as is standard, that welfare is concave in the price (hence in k), the welfare curve looks like Figure 2. In particular $W_{NL} = W(0) > W(1) = W_L$.

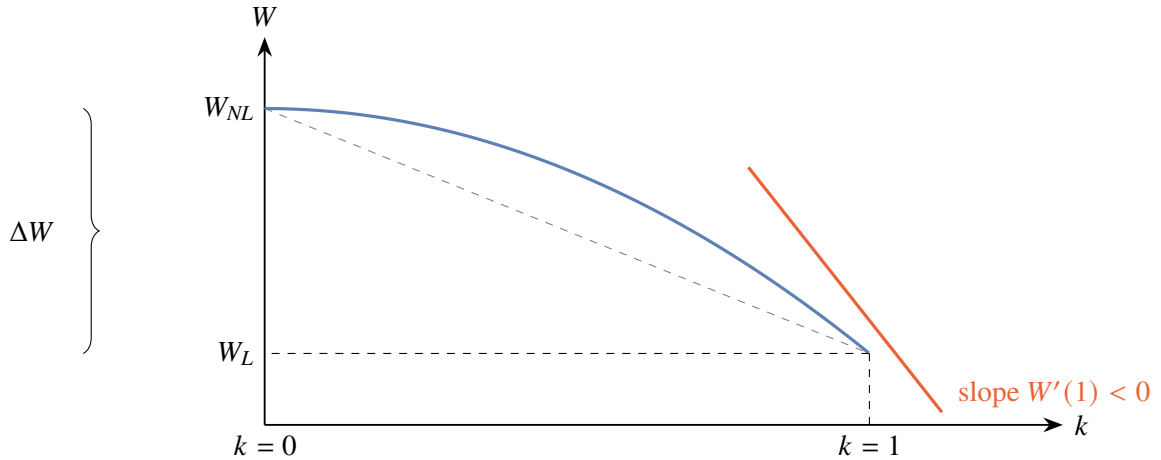


Figure 2: Welfare $W(k)$ is concave and decreasing in k . It is maximized at $k = 0$ (the two-part tariff, marginal-cost pricing) and falls to its linear-pricing value at $k = 1$. The gap is $\Delta W = W_{NL} - W_L$.

Result 2. Welfare is higher under two-part tariffs than under linear pricing, $W_{NL} > W_L$, because two-part tariffs price the marginal unit at cost and so avoid the quantity distortion that a linear price above cost creates.

How large is the welfare gap? Concavity lets us bound $\Delta W \equiv W_{NL} - W_L$. For a concave function the tangent at $k = 1$ lies above the curve, so evaluating that tangent back at $k = 0$ gives $W(0) \leq W(1) - W'(1)$, i.e.

$$\Delta W = W(0) - W(1) \leq -W'(1) = s_{kk}(1), \quad (21)$$

using $W'(1) = -s_{kk}(1)$. Now rewrite $s_{kk}(1)$ with the equilibrium condition (17), $s_{kk}(1) = \pi_L - \frac{1}{\tau}\pi_L^2$,

$$s_{kk}(1) = \pi_L - \frac{1}{\tau}\pi_L^2 = \frac{\pi_L}{\tau}(\tau - \pi_L). \quad (22)$$

Because $\pi_L < \tau$ from (19), the factor $\frac{\pi_L}{\tau} < 1$, while $\tau - \pi_L > 0$. Hence

$$s_{kk}(1) = \underbrace{\frac{\pi_L}{\tau}}_{<1} \underbrace{(\tau - \pi_L)}_{>0} < \tau - \pi_L.$$

Chaining this with the bound (21) and recalling $\tau - \pi_L = \pi_{NL} - \pi_L$,

$$\Delta W \leq s_{kk}(1) < \tau - \pi_L = \pi_{NL} - \pi_L \equiv \Delta\pi. \quad (23)$$

Result 3. The welfare gain from two-part tariffs is smaller than the profit gain, $\Delta W < \Delta\pi$. Firms gain more, when moving from linear to two-part pricing, than society as a whole. That extra gain has to come from somewhere, and the next section shows it comes out of consumers' pockets.

6. Consumer Surplus and the Overall Comparison

Total welfare is consumer surplus plus industry profit, $W = CS + \pi$, so $CS = W - \pi$. Rearranging the welfare-versus-profit comparison (23),

$$\begin{aligned} W_{NL} - W_L &< \pi_{NL} - \pi_L \\ \iff W_{NL} - \pi_{NL} &< W_L - \pi_L \\ &\underbrace{\hspace{1.5cm}}_{CS_{NL}} \quad \underbrace{\hspace{1.5cm}}_{CS_L} \\ \iff CS_{NL} &< CS_L. \end{aligned}$$

Result 4. Consumers are worse off under two-part tariffs than under linear pricing, $CS_{NL} < CS_L$, even though total welfare is higher.

Collecting the three comparisons, moving from linear pricing to two-part tariffs:

- total welfare rises;
- industry profit rises, by more than total welfare;
- consumer surplus falls.

In symbols,

$$W_{NL} > W_L, \quad \pi_{NL} > \pi_L, \quad \text{but} \quad CS_{NL} < CS_L.$$

Two-part tariffs enlarge the pie *and* hand firms a larger slice, so much larger that consumers are left with less than they had under linear pricing. Figure 3 shows the decomposition.

7. Intuition: Why Consumers Prefer Linear Pricing

The comparisons above are driven by how hard firms compete under each pricing regime, which in turn depends on what a firm gains by giving consumers a better deal.

Under **two-part tariffs**, the per-unit price is already at marginal cost, so the only way firm i improves its deal is by cutting the fixed fee. A one-dollar cut in the fee has two effects:

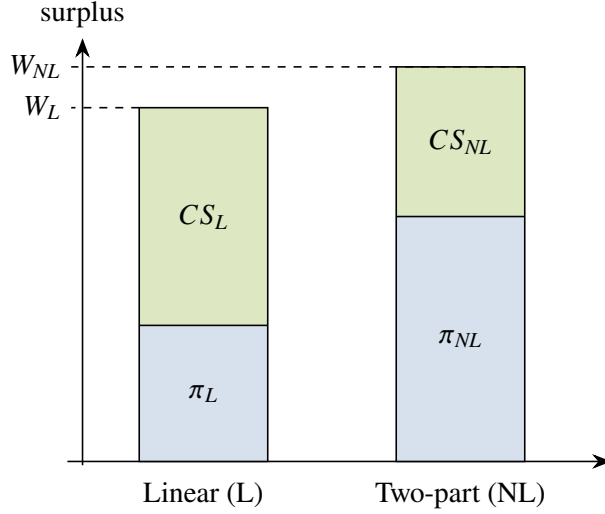


Figure 3: Decomposition of welfare into industry profit and consumer surplus. Moving from linear (L) to two-part (NL) pricing raises total welfare ($W_{NL} > W_L$) and profit ($\pi_{NL} > \pi_L$), but lowers consumer surplus ($CS_{NL} < CS_L$).

- (+) *Business stealing.* A better deal pulls marginal consumers away from the rival, raising the firm's market share $\hat{x}$.
- (-) *Lost revenue.* The firm collects one dollar less from every customer it already had.

Under **linear pricing**, the firm improves its deal by cutting the per-unit price, which has the same two effects *and* a third one:

- (+) *Demand expansion.* A lower price makes each of the firm's own customers buy more units, and the firm earns its margin on the extra volume.

This third effect gives firms a stronger reason to cut prices under linear pricing than to cut fees under two-part tariffs. Competition is therefore fiercer under linear pricing: prices are driven closer to cost, and more surplus is left in consumers' hands. Under two-part tariffs the marginal price is pinned at cost and the fixed fee moves no quantity, so this extra pro-competitive force is absent, firms keep fees high, and consumers end up with less, exactly the ranking $CS_{NL} < CS_L$ found above. The very efficiency of two-part tariffs, pricing each unit at cost, is what removes the demand-expansion channel that would otherwise have disciplined the firms.

8. A Numerical Example with Linear Demand

To put numbers on the four results, take the quadratic gross utility $u(q) = aq - \frac{q^2}{2}$, whose marginal utility is $u'(q) = a - q$. Setting $u'(q) = p$ gives the *linear demand*

$$q(p) = a - p \quad (\text{for } p \leq a),$$

and, evaluating $u(q) - pq$ at $q = a - p$, the indirect utility

$$v(p) = \max_q \left\{ aq - \frac{q^2}{2} - pq \right\} = \frac{1}{2} (a - p)^2.$$

As a check, $v'(p) = -(a - p) = -q(p)$, so Roy's identity (2) holds. Take $a = 2$, $c = 1$, and $\tau = \frac{9}{32}$.

Two-part tariff (NL). From Section 3 the per-unit price is $p^* = c = 1$ and the fee is $m^* = \tau = \frac{9}{32}$, so each firm posts $T(q) = \frac{9}{32} + q$. Every consumer buys $q(c) = a - c = 1$ unit. Industry profit is the fee collected from the unit mass of consumers, total welfare is the efficient surplus, and consumer surplus is that surplus net of the fee:

$$\pi_{NL} = \tau = \frac{9}{32}, \quad W_{NL} = v(c) = \frac{1}{2} (a - c)^2 = \frac{1}{2}, \quad CS_{NL} = v(c) - \tau = \frac{1}{2} - \frac{9}{32} = \frac{7}{32}.$$

Linear pricing (L). With no fee, the symmetric first-order condition from Section 4 is $\tau(a + c - 2p) = (a - p)^2(p - c)$. Writing the markup as $z \equiv p - c$ and inserting $a = 2, c = 1, \tau = \frac{9}{32}$,

$$\frac{9}{32} (1 - 2z) = (1 - z)^2 z \iff 32z^3 - 64z^2 + 50z - 9 = 0 \iff (4z - 1)(8z^2 - 14z + 9) = 0.$$

The quadratic factor has no real root (its discriminant is $14^2 - 4 \cdot 8 \cdot 9 = -92 < 0$), so the unique symmetric equilibrium markup is $z = \frac{1}{4}$, i.e. $p_L = c + \frac{1}{4} = \frac{5}{4}$. Each consumer then buys $q_L = a - p_L = \frac{3}{4}$, and

$$\pi_L = (p_L - c) q_L = \frac{1}{4} \cdot \frac{3}{4} = \frac{6}{32}, \quad CS_L = v(p_L) = \frac{1}{2} \left(\frac{3}{4}\right)^2 = \frac{9}{32}, \quad W_L = \pi_L + CS_L = \frac{15}{32}.$$

Comparison. Table 1 collects the two regimes.

	Two-part tariff (NL)	Linear pricing (L)
Per-unit price p	1	$\frac{5}{4}$
Units per consumer q	1	$\frac{3}{4}$
Industry profit π	$\frac{9}{32} \approx 0.281$	$\frac{6}{32} \approx 0.188$
Consumer surplus CS	$\frac{7}{32} \approx 0.219$	$\frac{9}{32} \approx 0.281$
Total welfare W	$\frac{16}{32} = 0.500$	$\frac{15}{32} \approx 0.469$

Table 1: Equilibrium outcomes under two-part and linear pricing, for $a = 2, c = 1$, and $\tau = \frac{9}{32}$. The profit, consumer-surplus, and welfare rows are written over the common denominator 32 so the numerators can be compared directly.

All four results hold:

$$\pi_{NL} = \frac{9}{32} > \frac{6}{32} = \pi_L, \quad W_{NL} = \frac{16}{32} > \frac{15}{32} = W_L, \quad CS_{NL} = \frac{7}{32} < \frac{9}{32} = CS_L,$$

and the welfare gain $\Delta W = \frac{1}{32}$ is smaller than the profit gain $\Delta \pi = \frac{3}{32}$.

Moving to the two-part tariff pushes the per-unit price down from $\frac{5}{4}$ to marginal cost 1, so each consumer's purchase rises from $\frac{3}{4}$ to a full unit: consumption becomes efficient and total surplus grows by $\frac{1}{32}$. The fixed fee, however, lets firms collect $\frac{3}{32}$ more than under linear pricing, which is $\frac{2}{32}$ more than the entire welfare gain. That extra $\frac{2}{32}$ is exactly the fall in consumer surplus ($\frac{9}{32} - \frac{7}{32}$). Consumption becomes efficient, industry profit rises, and consumers are left worse off.

References

- Armstrong, Mark, and John Vickers (2001). “Competitive Price Discrimination.” *RAND Journal of Economics* 32(4): 579–605. <https://www.jstor.org/stable/2696383>
- Belleflamme, Paul, and Martin Peitz (2015). *Industrial Organization: Markets and Strategies*, 2nd ed. Cambridge: Cambridge University Press (§9.3.2).

Discounts as a Barrier to Entry

A step-by-step exercise in the spirit of Ide, Montero, and Figueroa (2016)

Companion note to Ide, E., J.-P. Montero, and N. Figueroa (2016), “Discounts as a Barrier to Entry,” *American Economic Review* 106(7): 1849–1877. Full references at the end.

Overview. Dominant suppliers commonly offer retailers a discount on every unit they buy, provided they buy only from the dominant supplier: “as long as you buy exclusively from me, you get 10% off the list price on all units; otherwise, you pay the list price.” Antitrust authorities treat these rebates with suspicion, because they look like exclusive dealing, and the post-Chicago literature shows that exclusive dealing can foreclose a more efficient entrant. Ide, Montero, and Figueroa (2016) show that the analogy fails. An exclusive dealing contract commits the retailer before the entrant shows up, while a rebate leaves the retailer free to decide whether to stay exclusive after seeing the entrant’s offer. Without this commitment, a more efficient entrant can always outbid the rebate, so off-list-price rebates cannot block it profitably. This note builds the argument in the three post-Chicago settings that the paper examines, with $v = \frac{3}{2}$, $c_I = 1$, and $\lambda = \frac{1}{2}$ throughout, and with $c_E = 0$ whenever the entrant’s cost is known. Step (a) solves the benchmark without contracts; Step (b) compares exclusive dealing and rebates when the entrant’s cost is uncertain; Step (c) considers two buyers and scale economies; Step (d) considers retailers that compete downstream and distinguishes off-list-price from lump-sum rebates; Step (e) shows that rebates do block inefficient entrants; and Step (f) lists what the toy model keeps and drops.

1. The Toy Model

Consumers and manufacturers. A unit mass of final consumers values a good at $v = \frac{3}{2}$. An incumbent manufacturer, I , produces at marginal cost c_I , which we normalize to $c_I = 1$. A potential entrant, E , produces at marginal cost c_E after paying an entry cost $F \geq 0$. Whenever c_E is known, we normalize it to $c_E = 0$, so the entrant’s cost advantage is $c_I - c_E = 1$ and the examples below differ only in F . Only a share $\lambda = \frac{1}{2}$ of consumers, the contestable demand, regards both products as identical; the remaining consumers buy only from I . Entry is efficient if the entrant’s average cost on the contestable units is below the incumbent’s cost,

$$c_E + \frac{F}{\lambda} < c_I. \quad (1)$$

Retailers and timing. Manufacturers sell through retailers, who have no other costs. On date 1, I offers a contract to the retailer(s); on date 2, E offers a take-or-pay contract for λ units, which is cancelled if E does not enter; on date 3, E decides whether to enter; on date 4, retailers buy according to the contracts they hold, or in the spot market, where I and E compete with nonlinear prices.

Contracts. The paper compares three contracts.

- An exclusive dealing contract (t, w, D) : a payment t from I on date 1, a wholesale price w , and a penalty D that the retailer pays I if she buys from E .
- An off-list-price rebate (r, R) : a list price r per unit, and a discount R on every unit if the retailer buys only from I .
- A lump-sum rebate (r, L) : the list price r and a fixed payment L if the retailer buys only from I , whatever the quantity.

A retailer that switches the contestable λ units to E loses the discount on all units, so the price that E must beat on the contestable units, the effective price of a rebate, is

$$x \equiv r - \frac{R}{\lambda}. \quad (2)$$

The key difference is timing: the penalty D is committed on date 1, while the rebate is lost only if the retailer, after seeing E 's offer on date 2, decides to buy from E .

2. Step (a): The No-Contract Benchmark

Suppose that entry is efficient and I offers no contract. In the spot market, E sells the λ contestable units at a price slightly below c_I , and I sells the remaining $1 - \lambda$ units with a two-part tariff, a unit price c_I and a fixed fee $(1 - \lambda)(v - c_I)$, which extracts the retailer's surplus on these units. The payoffs are

$$\pi_I^{NC} = (1 - \lambda)(v - c_I) = \frac{1}{4}, \quad \pi_E^{NC} = \lambda(c_I - c_E) - F, \quad \pi_B^{NC} = \lambda(v - c_I) = \frac{1}{4}. \quad (3)$$

The retailer keeps $\lambda(v - c_I)$ because she buys the contestable units at c_I instead of v .

Result 1 (no-contract benchmark). Without contracts, the efficient entrant enters, the incumbent earns $\frac{1}{4}$ from the non-contestable consumers, and the retailer earns $\frac{1}{4}$ from the competition between manufacturers on the contestable units.

Every contract that I offers is profitable only if it yields more than $\pi_I^{NC} = \frac{1}{4}$.

3. Step (b): Rent Shifting with an Uncertain Entrant (One Buyer)

Setting. Following Aghion and Bolton (1987), there is one retailer, $F = 0$, and on date 1 neither I nor the retailer knows c_E , which is uniformly distributed on $[0, c_I] = [0, 1]$, so $G(c) = c$ and every entrant is more efficient than I .

Exclusive dealing. A retailer who signed $(0, w, D)$ switches to E 's offer w_E if paying w on $1 - \lambda$ units, D , and w_E on λ units costs her no more than paying w on all units, that is, if $w_E \leq w - \frac{D}{\lambda}$. The entrant therefore enters if and only if $c_E \leq y \equiv w - \frac{D}{\lambda}$. Since the retailer pays w per unit in every case, she signs if $v - w \geq \pi_B^{NC}$, so I sets $w = v - \lambda(v - c_I) = \frac{5}{4}$. Writing $D = \lambda(w - y)$, the incumbent's expected profit is

$$\pi_I^{ED}(y) = (w - c_I) + \lambda(c_I - y)G(y) = \frac{1}{4} + \frac{1}{2}(1 - y)y. \quad (4)$$

The second term is the expected penalty collected from an entrant with cost below y , net of the sales lost. Maximizing over y gives $y^* = c_I - \frac{G(y^*)}{g(y^*)} = \frac{1}{2}$, so $D^* = \lambda(w - y^*) = \frac{3}{8}$, and

$$\pi_I^{ED} = \frac{1}{4} + \frac{1}{8} = \frac{3}{8} > \frac{1}{4}. \quad (5)$$

Result 2 (Aghion and Bolton). With an exclusive contract, the incumbent sets the effective price $y^* = \frac{1}{2}$ and earns $\frac{3}{8}$. Entrants with $c_E \in (\frac{1}{2}, 1)$, half of all entrants, are more efficient than I but stay out, and the expected welfare loss is $\lambda \int_{1/2}^1 (c_I - c) dG(c) = \frac{1}{16}$.

The penalty works as an entry fee: the incumbent extracts part of the entrant's cost advantage, at the price of excluding the entrants whose advantage is small.

Off-list-price rebates. Under a rebate (r, R) , the retailer switches to E if $v - (1 - \lambda)r - \lambda w_E \geq v - (r - R)$, that is, if $w_E \leq x$, so E enters if and only if $c_E \leq x$. The incumbent sells $1 - \lambda$ units at r after entry and all units at $r - R$ otherwise, and substituting $R = \lambda(r - x)$,

$$\pi_I^R(r, x) = (1 - \lambda)(r - c_I) + \lambda(x - c_I)[1 - G(x)]. \quad (6)$$

A retailer who buys from E pays the list price r for the remaining units, so she would rather buy nothing from I if $r > v$; therefore $r \leq v$.

Result 3 (Proposition 1 of the paper). No off-list-price rebate with $x < c_I$ is profitable: for every such rebate, $\pi_I^R < (1 - \lambda)(v - c_I) = \frac{1}{4}$, since $r \leq v$ and the second term of (6) is negative.

Figure 1 depicts the incumbent's expected profit as a function of the effective price it sets, under exclusive dealing (4) and under the best rebate (6) with $r = v$, which becomes $\pi_I^R(v, x) = \frac{1}{4} - \frac{1}{2}(1 - x)^2$. The exclusive contract earns more than the benchmark for every $y < c_I$ and peaks at $y^* = \frac{1}{2}$, while the rebate earns less than the benchmark for every $x < c_I$.

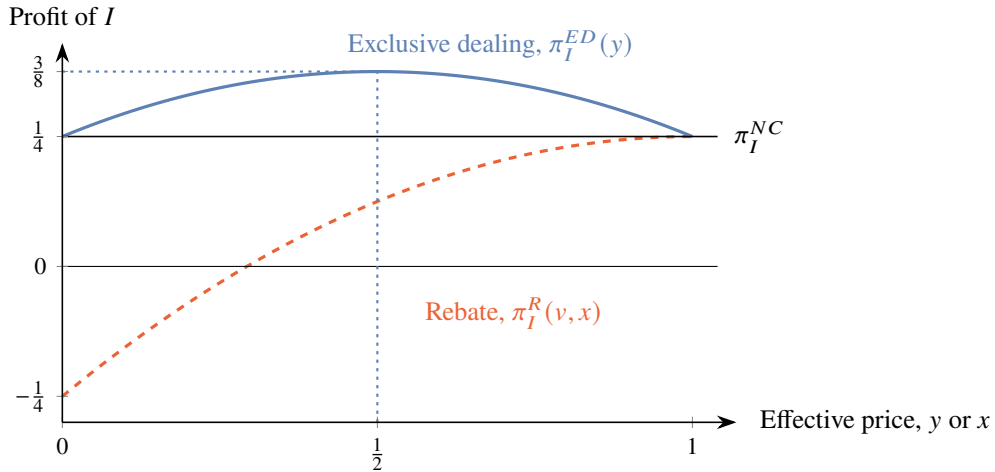


Figure 1: Incumbent's expected profit as a function of the effective price, with an exclusive contract and with an off-list-price rebate ($v = \frac{3}{2}$, $c_I = 1$, $\lambda = \frac{1}{2}$, c_E uniform on $[0, 1]$).

Intuitively, the two contracts look alike: relabeling $w = r - R$ and $D = (1 - \lambda)R$ turns (6) into the exclusive's profit, since giving up the rebate on the $1 - \lambda$ non-contestable units works as a penalty. The difference is the constraint on the list price. With an exclusive, the penalty is committed on date 1, so it does not enter the retailer's later decision, and the only constraint is $w \leq v$. With a rebate, nothing is committed, so the list price the retailer would pay after switching must satisfy

$$r = w + \frac{D}{1 - \lambda} \leq v. \quad (7)$$

The exclusive optimum, $w = \frac{5}{4}$ and $D = \frac{3}{8}$, would require $r = \frac{5}{4} + \frac{3/8}{1/2} = 2 > \frac{3}{2}$, a list price the retailer would never pay.

Figure 2 depicts this comparison in the (D, w) plane. The retailer accepts any exclusive with $w \leq \frac{5}{4}$ (below the horizontal line), while a rebate must also satisfy (7), the region below the blue line. The dashed red lines

are effective prices $w - \frac{D}{\lambda} = y$: the exclusive optimum lies on $y = \frac{1}{2}$, far outside the rebate region, and every rebate that lowers the effective price below c_I forces the incumbent to lower w as well, which costs it more than the rent it extracts.

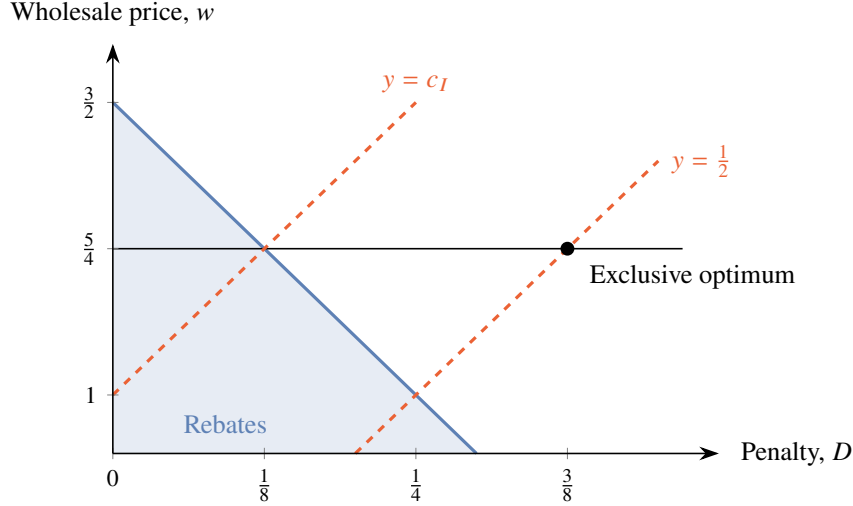


Figure 2: Exclusive contracts and rebates in the (D, w) plane. Rebates must lie in the shaded region, $w + \frac{D}{1-\lambda} \leq v$; dashed lines are effective prices $y = w - \frac{D}{\lambda}$.

4. Step (c): Scale Economies and Two Buyers

Setting. Following Rasmusen et al. (1991) and Segal and Whinston (2000), there are two retailers, B_1 and B_2 , each serving half of the consumers in a separate market. The entrant's cost is known, $c_E = 0$, and its entry cost is $F = \frac{7}{16}$. Entry is efficient, since $\lambda(c_I - c_E) - F = \frac{1}{16} > 0$, but the entrant needs both retailers, since serving one yields at most $\frac{\lambda(v - c_E)}{2} = \frac{3}{8} < F$. The two conditions hold together because $v < 2c_I$.

Exclusive dealing. The incumbent offers B_1 a little more than her benchmark payoff, $\frac{\pi_B^{NC}}{2} = \frac{1}{8}$, and offers B_2 almost nothing. If B_1 signs, E cannot reach the scale it needs, so B_2 expects to pay v anyway and also signs. The incumbent earns $v - c_I - \frac{1}{8} = (1 - \frac{\lambda}{2})(v - c_I) = \frac{3}{8} > \frac{1}{4}$: it pays one buyer to exclude and exploits the other.

Rebates. To replicate this, the incumbent offers B_1 a large rebate (v, R_1) and B_2 almost none. But B_2 has signed nothing, so E can sell to her at a price close to v , and use the profit to outbid B_1 's rebate. B_1 switches if $\frac{\lambda(v - w_{E1})}{2} \geq \frac{R_1}{2}$, and E breaks even when $\frac{\lambda(w_{E1} - c_E)}{2} + \frac{\lambda(v - c_E)}{2} = F$, so to block entry the incumbent needs

$$\frac{R_1}{2} > \lambda(v - c_I) + [\lambda(c_I - c_E) - F] = \frac{1}{4} + \frac{1}{16} = \frac{5}{16}, \quad (8)$$

which leaves it with at most $v - c_I - \frac{5}{16} = \frac{3}{16} < \frac{1}{4}$.

Result 4 (Propositions 2 and 3 of the paper). With scale economies and two buyers:

- exclusive contracts block the efficient entrant and raise the incumbent's profit from $\frac{1}{4}$ to $\frac{3}{8}$;
- rebates that block the entrant leave the incumbent with at most $(1 - \lambda)(v - c_I) - [\lambda(c_I - c_E) - F]$, which is below $\frac{1}{4}$ whenever entry is efficient.

Figure 3 depicts the incumbent's profit from a blocking rebate, $F - \frac{1}{4}$, as a function of the entry cost F . It lies below the benchmark $\frac{1}{4}$ for every F at which entry is efficient, $F < \lambda(c_I - c_E) = \frac{1}{2}$, and above it only when entry is inefficient. Intuitively, to keep B_1 exclusive, the incumbent must give her more than the entrant can offer, and the entrant can pass on to B_1 the surplus it extracts from the unprotected B_2 . When entry is efficient, this payment exceeds what the incumbent gains by exploiting the buyers.

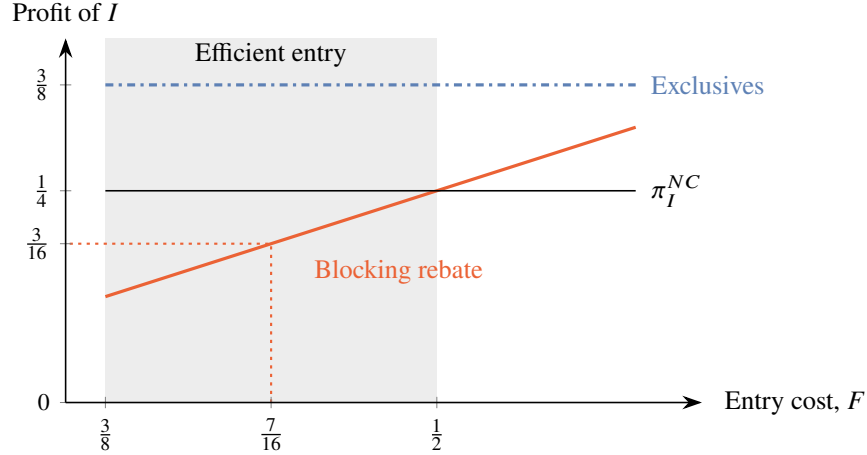


Figure 3: Incumbent's profit from an exclusive contract and from the cheapest rebate that blocks entry, as a function of the entry cost F ($c_E = 0$). The shaded interval has scale economies and efficient entry.

5. Step (d): Retailers Who Compete Downstream

Setting. Following Simpson and Wickelgren (2007) and Asker and Bar-Isaac (2014), two retailers now compete in prices for the same consumers. Since E needs only one retailer to reach the contestable consumers, and price competition passes any cost saving to consumers, retailers earn zero in the benchmark.

Exclusive dealing. A retailer earns zero whether or not she signs, so the incumbent gets both retailers to sign exclusives for an arbitrarily small payment and earns $v - c_I = \frac{1}{2}$ (Proposition 4 of the paper).

Off-list-price rebates. A discount per unit lowers the retailer's marginal cost, so competition passes it to consumers: exclusive retailers sell at $r - R$. To keep E out, the incumbent must set $r - R < c_E + \frac{F}{\lambda}$, below its own cost, so blocking entry with these rebates is a form of limit pricing that loses money (Proposition 5 of the paper).

Lump-sum rebates. A fixed payment L for exclusivity does not change the retailer's marginal cost, so competition does not pass it to consumers. A retailer that switches earns $\lambda(c_I - w_E)$ on the contestable consumers and loses L , so the entrant can win her only if $L < \lambda(c_I - c_E) - F$. Paying each retailer $L = \lambda(c_I - c_E) - F$ blocks entry and yields $v - c_I - 2L$, which exceeds π_I^{NC} if and only if

$$\frac{\lambda(v - c_I)}{2} > \lambda(c_I - c_E) - F. \quad (9)$$

With $c_E = 0$ and $F = \frac{7}{16}$, $L = \frac{1}{16}$ and the incumbent earns $\frac{1}{2} - \frac{1}{8} = \frac{3}{8} > \frac{1}{4}$, so lump-sum rebates exclude an efficient entrant. With $F = 0$, instead, $L = \frac{1}{2}$ and blocking entry costs more than it yields.

Result 5 (Propositions 4–6 of the paper). With retailers who compete downstream:

- exclusive contracts block entry for an arbitrarily small payment;
- off-list-price rebates never block an efficient entrant profitably;
- lump-sum rebates block an efficient entrant profitably if and only if (9) holds, that is, when the entrant's surplus from entry is below half of the incumbent's margin on the contestable consumers.

Figure 4 depicts condition (9) in the (c_E, F) plane. Entry is efficient below the solid line, $F < \lambda(c_I - c_E)$, and lump-sum rebates exclude the entrant in the shaded band, where the entrant's surplus is positive but below $\frac{\lambda(v - c_I)}{2} = \frac{1}{8}$. The dot marks the example $c_E = 0, F = \frac{7}{16}$.

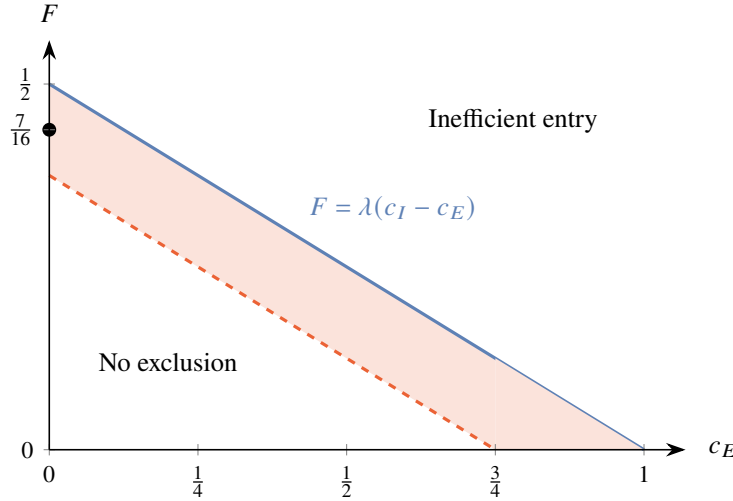


Figure 4: Downstream competition: parameters for which lump-sum rebates profitably exclude an efficient entrant (shaded), with $v = \frac{3}{2}$, $c_I = 1$, $\lambda = \frac{1}{2}$. The dashed line is $F = \lambda(c_I - c_E) - \frac{1}{8}$.

Intuitively, lump-sum rebates work because they reward retailers without lowering retail prices, so retailers internalize part of the incumbent's loss from entry. An off-list-price rebate, in contrast, is competed away in lower retail prices, and the retailers have nothing to lose by switching. The size of the contestable demand matters here: with $c_E = 0$ and $F = \frac{7}{16}$, entry is efficient both for $\lambda = \frac{1}{2}$ and for $\lambda = 1$, but condition (9) holds only for $\lambda = \frac{1}{2}$, since a smaller contestable demand raises the entrant's entry cost per contestable unit. The paper's remark that the size of the contestable demand does not matter applies to off-list-price rebates and to settings without entry costs.

6. Step (e): A Generalized Chicago Critique and Inefficient Entry

The common thread of Steps (b)–(d) is what the paper calls a generalized Chicago critique. Excluding an efficient rival destroys surplus, so if every party could bargain at the same time, exclusion could not occur. Post-Chicago models obtain exclusion because one party, the entrant or final consumers, is absent when the buyer signs, and the buyer is locked in before that party shows up. Exclusive dealing provides this lock-in; off-list-price rebates do not, because the buyer decides whether to stay exclusive only after seeing the entrant's offer. The paper notes that adding an unconditional transfer from the buyer to the incumbent, paid whether or not she later stays exclusive, would restore the commitment, but such transfers have not been documented in rebate cases.

Inefficient entry. Rebates without commitment do block entrants whose entry destroys surplus. Suppose

$c_E = 0$ and $F = \frac{5}{8}$ with one retailer, so E 's average cost on the contestable units is $c_E + \frac{F}{\lambda} = \frac{5}{4}$, above c_I but below v . Without a contract, E would enter, since its profit $\lambda(v - c_E) - F = \frac{1}{8}$ is positive. The incumbent offers the list price $r = v$ and a discount $R = \frac{1}{8}$, so the retailer pays $\frac{11}{8}$ per unit if exclusive, and the effective price on the contestable units is $x = c_E + \frac{F}{\lambda} = \frac{5}{4}$. The entrant cannot beat this price, and the incumbent earns

$$\pi_I = (1 - \lambda)(v - c_I) + \lambda \left(c_E + \frac{F}{\lambda} - c_I \right) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}. \quad (10)$$

Result 6 (inefficient entry). With one retailer and a known entrant's cost, the incumbent profitably blocks entry with an off-list-price rebate if and only if the entrant's average cost on the contestable units exceeds the incumbent's cost, $c_E + \frac{F}{\lambda} > c_I$. The incumbent's gain from blocking, $\lambda(c_E + \frac{F}{\lambda} - c_I)$, equals the welfare loss that entry would cause.

Figure 5 depicts, as a function of the entry cost F with $c_E = 0$, the welfare effect of entry, $\lambda(c_I - c_E) - F = \frac{1}{2} - F$, and the incumbent's gain from a blocking rebate. The gain is zero while entry is efficient, $F < \frac{1}{2}$, and equals the welfare loss from entry above that cost, up to $F = \lambda v = \frac{3}{4}$, where the entrant would not enter even without a contract. Intuitively, the incumbent is the residual claimant of the surplus created by removing an inefficiency, so it can always outbid the entrant for the retailer's loyalty, without any commitment.

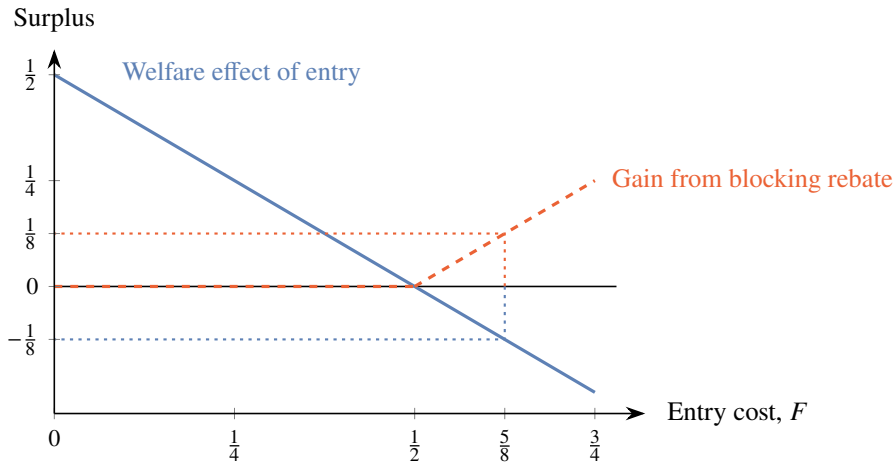


Figure 5: Welfare effect of entry and incumbent's gain from a blocking off-list-price rebate, as a function of the entry cost ($c_E = 0$, one retailer).

7. Step (f): What the Toy Model Keeps and Drops

Table 1 collects the incumbent's profits in the examples of Steps (b)–(e).

What the toy model keeps.

- The role of commitment: exclusive dealing locks the buyer in before the entrant appears, while an off-list-price rebate leaves her free to switch after seeing the entrant's offer.
- The equivalence that fails: relabeling a rebate as an exclusive works for profits but not for the constraint $r \leq v$ on the list price.
- The difference between off-list-price and lump-sum rebates when retailers compete.

Table 1: Incumbent's profit with each contract ($v = \frac{3}{2}$, $c_I = 1$, $\lambda = \frac{1}{2}$; benchmark $\pi_I^{NC} = \frac{1}{4}$).

Setting	Exclusive dealing	Blocking rebate	Entry blocked?
Uncertain entrant, $c_E \sim U[0, 1]$	$\frac{3}{8}$	below $\frac{1}{4}$	Exclusive only
Two buyers, $c_E = 0$, $F = \frac{7}{16}$	$\frac{3}{8}$	$\frac{3}{16}$	Exclusive only
Downstream competition, off-list-price	$\frac{1}{2}$	below 0	Exclusive only
Downstream competition, lump-sum, $F = \frac{7}{16}$	$\frac{1}{2}$	$\frac{3}{8}$	Both
Inefficient entrant, $c_E = 0$, $F = \frac{5}{8}$	—	$\frac{3}{8}$	Rebate (efficiently)

- Rebates that block inefficient entry, which raise welfare.

What it drops.

- Downward-sloping demand and incomplete surplus extraction in the non-contestable segment, which underlie bundling theories of exclusion.
- Asymmetric information about the buyer, contracts that combine rebates with unconditional transfers, and repeated interaction.
- Antitrust implications: the paper argues that the off-list-price rebates in *Michelin II* and *British Airways* could not have blocked a more efficient rival, while the apparently lump-sum rebates in *AMD v. Intel* were potentially exclusionary.

References

- Aghion, P., and Bolton, P. (1987). Contracts as a barrier to entry. *American Economic Review*, 77(3), 388–401.
- Asker, J., and Bar-Isaac, H. (2014). Raising retailers' profits: On vertical practices and the exclusion of rivals. *American Economic Review*, 104(2), 672–686.
- Ide, E., Montero, J.-P., and Figueroa, N. (2016). Discounts as a barrier to entry. *American Economic Review*, 106(7), 1849–1877.
- Rasmusen, E. B., Ramseyer, J. M., and Wiley, J. S., Jr. (1991). Naked exclusion. *American Economic Review*, 81(5), 1137–1145.
- Segal, I. R., and Whinston, M. D. (2000). Naked exclusion: Comment. *American Economic Review*, 90(1), 296–309.
- Simpson, J., and Wickelgren, A. L. (2007). Naked exclusion, efficient breach, and downstream competition. *American Economic Review*, 97(4), 1305–1320.

Forward Trading and Collusion

A step-by-step exercise in the spirit of Liski and Montero (2006)

Companion note to Liski, M., and J.-P. Montero (2006), “Forward Trading and Collusion in Oligopoly,” *Journal of Economic Theory* 131(1): 212–230. Full references at the end.

Overview. In many commodity markets, such as electricity, copper, or coffee, producers sell part of their future output in advance through forward contracts. Allaz and Vila (1993) show that, when firms meet only once, forward trading makes the spot market more competitive: a firm that sells forward commits to producing more, and both firms end up selling forward and earning less, as in a prisoner’s dilemma. For this reason, forward contracting has been advanced as a way to mitigate market power in electricity markets. Liski and Montero (2006) ask whether the same conclusion holds when the same firms meet again and again in both the forward and the spot markets. Their answer is no: the possibility of trading forwards makes it easier for firms to sustain the monopoly outcome. Forward contracts lower the profit from a deviation, because a deviating firm can only take the part of the market that has not been contracted in advance, and, with quantity competition, they make the punishment harsher. This note builds the argument with linear demand $p = 1 - Q$ and zero costs. Step (a) finds the critical discount factors without forward trading; Step (b) solves the one-shot game with forward trading of Allaz and Vila; Step (c) studies collusion with price competition in the spot market; Step (d) studies collusion with quantity competition; Step (e) shows that long forward positions require checking deviations in the forward market; and Step (f) lists what the toy model keeps and drops.

1. The Toy Model

Firms and demand. Two firms produce a homogeneous good at zero marginal cost. In every spot market, the inverse demand is

$$p = 1 - Q, \quad (1)$$

where $Q = q_1 + q_2$ is total output. The paper uses $p = a - Q$ and marginal cost c ; with $a = 1$ and $c = 0$, every profit below is the paper’s profit divided by $(a - c)^2$, so every critical discount factor is the same as in the paper.

Monopoly benchmark. A monopolist produces $q_m = \frac{1}{2}$, charges $p_m = \frac{1}{2}$, and earns $\pi_m = \frac{1}{4}$. If the two firms collude on this outcome and split it equally, each firm earns $\frac{\pi_m}{2} = \frac{1}{8}$ per spot market.

Forward contracts. Before each spot market, a forward market opens in which each firm can sell ($f_i > 0$, a short position) or buy ($f_i < 0$, a long position) contracts for delivery in that spot market. Forward positions are observable, and competitive speculators buy or sell contracts at a forward price equal to the spot price they expect, so there are no arbitrage profits. Once firm i holds a forward position f_i , it sells only $q_i - f_i$ units in the spot market, so its spot payoff is

$$\pi_i^s = p (q_i - f_i), \quad (2)$$

while the revenue from its forward sales was collected when it sold them. Since forwards sell at the expected spot price, firm i ’s total profit from a spot market, counting its forward revenue, is $p q_i$.

Repeated interaction. The two firms meet in an infinite sequence of spot markets, each preceded by a forward market. The discount factor between two consecutive spot markets is $\delta \in (0, 1)$. A firm that deviates

from a collusive plan is punished by its rival in all later periods. The question is how small δ can be for the firms to sustain the monopoly outcome, and how forward trading changes this critical discount factor.

2. Step (a): Collusion without Forward Trading

Quantity competition. In a one-shot Cournot game, firm i maximizes $(1 - q_i - q_j) q_i$, so its best response is $q_i = \frac{1 - q_j}{2}$. The equilibrium is $q_1 = q_2 = \frac{1}{3}$, with price $\frac{1}{3}$ and profit $\frac{1}{9}$ per firm. If the rival produces its collusive output $\frac{q_m}{2} = \frac{1}{4}$, a deviating firm produces $\frac{1 - 1/4}{2} = \frac{3}{8}$, the price falls to $\frac{3}{8}$, and the deviator earns $\frac{9}{64}$. With grim-trigger strategies that revert to the Cournot equilibrium forever, collusion is sustainable if

$$\frac{1/8}{1 - \delta} \geq \frac{9}{64} + \frac{\delta}{1 - \delta} \cdot \frac{1}{9} \iff \delta \geq \frac{9}{17} \approx 0.529. \quad (3)$$

Price competition. In a one-shot Bertrand game, both firms price at marginal cost and earn zero. A firm that undercuts the monopoly price slightly takes the whole market and earns $\pi_m = \frac{1}{4}$ in the deviation period, so collusion is sustainable if $\frac{1/8}{1 - \delta} \geq \frac{1}{4}$, that is, $\delta \geq \frac{1}{2}$.

Result 1 (no forward trading). The monopoly outcome can be sustained:

- with quantity competition and Cournot reversion, if and only if $\delta \geq \frac{9}{17} \approx 0.529$;
- with price competition and Bertrand reversion, if and only if $\delta \geq \frac{1}{2}$.

Reverting to the Cournot equilibrium is not the harshest credible punishment. With an optimal penal code (Abreu, 1986), firms punish a deviation with one period of high output, $z = \frac{5}{12}$ each, and then return to collusion; the critical discount factor falls to $\frac{9}{32} \approx 0.281$. We return to this benchmark in Step (f).

3. Step (b): Forward Trading in a One-Shot Market

The spot market. Given forward positions f_i and f_j , firm i chooses q_i to maximize its spot payoff (2), $(1 - q_i - q_j)(q_i - f_i)$. The first-order condition, $1 - 2q_i - q_j + f_i = 0$, gives the best response

$$q_i = \frac{1 - q_j + f_i}{2}. \quad (4)$$

A forward sale shifts firm i 's best response outward: the firm no longer cares about the price it receives on the f_i units it has already sold, so it behaves as if it were larger. Solving the two best responses,

$$q_i = \frac{1 + 2f_i - f_j}{3}, \quad p = \frac{1 - f_i - f_j}{3}. \quad (5)$$

Figure 1 depicts the best responses (4) without forward positions (solid) and when both firms sell $f = \frac{1}{5}$ forward (dashed). Both best responses shift outward by $\frac{f}{2}$, and the equilibrium moves from the Cournot point C , with $q_i = \frac{1}{3}$, to point AV , with $q_i = \frac{2}{5}$.

The forward market. Anticipating (5), and since forwards sell at the expected spot price, firm i chooses f_i to maximize its total profit,

$$p q_i = \frac{(1 - f_i - f_j)(1 + 2f_i - f_j)}{9}. \quad (6)$$

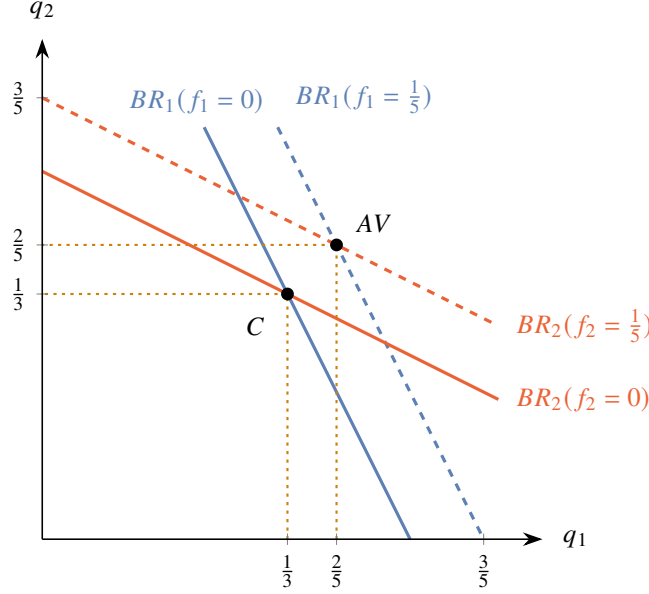


Figure 1: Spot-market best responses without forward positions (solid) and with forward sales $f_1 = f_2 = \frac{1}{5}$ (dashed).

The first-order condition, $-(1 + 2f_i - f_j) + 2(1 - f_i - f_j) = 0$, gives $f_i = \frac{1 - f_j}{4}$, so in the symmetric equilibrium $f_i = \frac{1}{5}$. Substituting into (5), $q_i = \frac{2}{5}$ and $p = \frac{1}{5}$, and each firm earns $\frac{2}{25} = 0.08 < \frac{1}{9} \approx 0.111$.

Result 2 (Allaz and Vila, one forward round). When a single forward market precedes the spot market, each firm sells $\frac{1}{5}$ forward and produces $\frac{2}{5}$; the price falls from $\frac{1}{3}$ to $\frac{1}{5}$, and each firm's profit falls from $\frac{1}{9}$ to $\frac{2}{25}$.

Intuitively, selling forward works as a commitment to produce more, which leads the rival to produce less, as in a Stackelberg game. Each firm therefore wants to sell forward, but when both do it, both produce more and both earn less than without a forward market, a prisoner's dilemma.

Several forward rounds. Suppose now that N forward markets open before the spot market. In the last round, a firm that already holds P_i from earlier rounds chooses its final position F_i , and its payoff is $p(q_i - P_i)$, since only the units it sells in this round earn the price it can still affect. The first-order condition gives $F_i = \frac{1 - F_j + 3P_i}{4}$, so the last round's equilibrium is $F_i = \frac{1 + 4P_i - P_j}{5}$. In the earlier round, firm i anticipates this response and maximizes $p q_i = \frac{(1 - P_i - P_j)(2 + 3P_i - 2P_j)}{25}$, whose first-order condition gives $P_i = \frac{1}{7}$ in a symmetric equilibrium. With $N = 2$ rounds, each firm therefore ends up selling $\frac{2}{7}$ forward, produces $\frac{3}{7}$, and the price is $\frac{1}{7}$. Repeating these steps for any N ,

$$F(N) = \frac{N}{3 + 2N}, \quad q(N) = \frac{1 + N}{3 + 2N}, \quad p(N) = \frac{1}{3 + 2N}, \quad \pi(N) = \frac{1 + N}{(3 + 2N)^2}. \quad (7)$$

Result 3 (Allaz and Vila, N forward rounds). Each firm's profit $\pi(N)$ in (7) decreases in the number of forward rounds and tends to zero, the competitive outcome, as N grows.

Figure 2 depicts $\pi(N)$ for $N = 1, \dots, 12$, together with the Cournot profit $\frac{1}{9}$ and the collusive profit $\frac{1}{8}$. Every additional forward round gives each firm another chance to commit to a larger output, so the market approaches the competitive outcome. This path will serve as the punishment in Step (d).

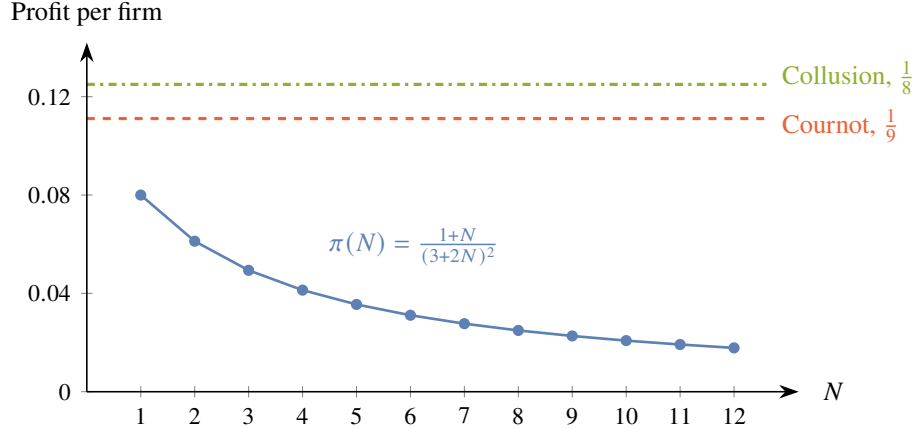


Figure 2: Profit per firm when N forward rounds precede the spot market.

4. Step (c): Collusion with Price Competition in the Spot Market

The collusive plan. Firms now meet repeatedly. Along the collusive path, each firm sells $x \frac{q_m}{2} = \frac{x}{4}$ units forward one period ahead, with $0 \leq x \leq 1$, and both firms charge $p_m = \frac{1}{2}$ in the spot market. A share x of the monopoly output is thus contracted in advance, and each firm sells the rest, $(1-x) \frac{q_m}{2}$, in the spot market. If any firm deviates, both price at marginal cost forever after.

Deviations. A deviation in the forward market never pays: speculators see it immediately and anticipate competitive pricing in the next spot market, so they pay at most the marginal cost, zero, for the deviator's contracts. In the spot market, the contracted output $x q_m = \frac{x}{2}$ is already committed, so a deviating firm faces the residual demand $1 - \frac{x}{2} - p$. Undercutting slightly is no longer optimal; the deviator sets

$$p_d = \arg \max_p p \left(1 - \frac{x}{2} - p\right) = \frac{2-x}{4}, \quad \text{earning} \quad \pi_d = \frac{(2-x)^2}{16}. \quad (8)$$

Critical discount factor. At the opening of a spot market, a firm that follows the plan earns its uncontracted share of monopoly profit in that market, $(1-x) \frac{1}{8}$, plus $\frac{1}{8}$ in every later market. The firm does not deviate if

$$\frac{1-x}{8} + \frac{\delta}{1-\delta} \cdot \frac{1}{8} \geq \frac{(2-x)^2}{16}, \quad (9)$$

which holds if and only if

$$\delta \geq \delta_B(x) \equiv \frac{x^2 - 2x + 2}{x^2 - 2x + 4}. \quad (10)$$

Result 4 (Proposition 1 of the paper). With price competition in the spot market, the critical discount factor $\delta_B(x)$ decreases in the share x of output sold forward, from $\delta_B(0) = \frac{1}{2}$ to $\delta_B(1) = \frac{1}{3}$.

Figure 3 depicts $\delta_B(x)$. Collusion can be sustained above the curve. For $x < 0$, when firms buy forwards (long positions), the curve is flat at $\frac{1}{2}$: a deviator can undercut and take the whole spot market, and speculators close their positions at the same price whether a deviation occurs or not (Proposition 2 of the paper).

Intuitively, forward trading has two opposing effects. On the one hand, it lowers the payoff from following the plan in the current period, since the contracted share x of current sales has already been paid for; this is the pro-competitive effect. On the other hand, it lowers the payoff from deviating, since the contracted

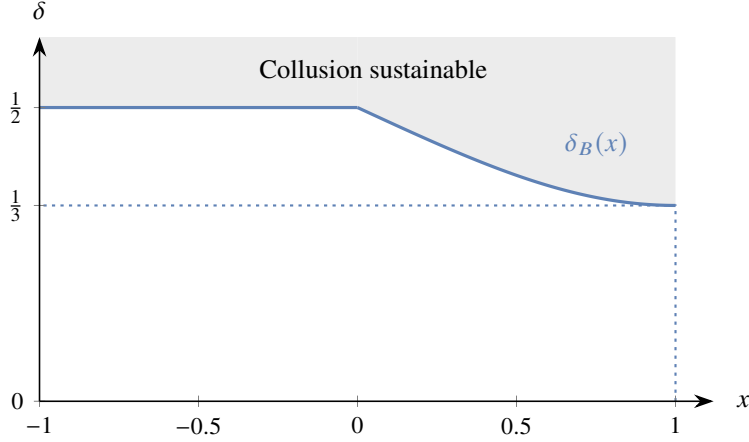


Figure 3: Critical discount factor with price competition in the spot market, as a function of the share x of monopoly output sold forward one period ahead ($x < 0$: long positions).

units already serve part of the market and the deviator can only capture the residual demand; this is the pro-collusive effect. With price competition, the second effect dominates for every $x \in [0, 1]$.

5. Step (d): Collusion with Quantity Competition in the Spot Market

The collusive plan. Along the collusive path, each firm sells $\frac{x}{4}$ units forward one period ahead, now with $-1 \leq x \leq 1$, and supplies another $(1-x)\frac{1}{4}$ in the spot market, so that each firm's total output is $\frac{1}{4}$ and the price is $\frac{1}{2}$. If any firm deviates, both play the Allaz–Vila equilibrium thereafter. Since firms can trade in every forward market for delivery in any later spot market, the spot market that opens N periods after a deviation is preceded by N forward openings, and each firm earns $\pi(N)$ in (7) there.

Deviations. In the spot market, the rival produces $\frac{1}{4}$ in total, of which $\frac{x}{4}$ was sold forward. A deviating firm chooses its total output Q_i to maximize $\left(1 - \frac{1}{4} - Q_i\right) \left(Q_i - \frac{x}{4}\right)$, which gives

$$Q_i = \frac{3+x}{8}, \quad p_d = \frac{3-x}{8}, \quad \pi_d = \frac{(3-x)^2}{64}. \quad (11)$$

Since π_d decreases in x , forward sales again reduce the profit from a deviation. With $x = 0$, $\pi_d = \frac{9}{64}$, as in Step (a). For $x \geq 0$, a deviation in the forward market does not pay: against a rival who sells $\frac{x}{4}$, the best forward sale is $\frac{1-x/4}{4}$, which earns $\frac{(1-x/4)^2}{8} \leq \frac{1}{8}$ in the next spot market and triggers the punishment.

A first look: one forward round per period. Suppose, only for this paragraph, that firms could trade forwards only in the period right before delivery, so the punishment is the one-round equilibrium of Result 2 forever, with profit $\frac{2}{25}$ per period. With $x = 0$, collusion is sustainable if

$$\frac{1/8}{1-\delta} \geq \frac{9}{64} + \frac{\delta}{1-\delta} \cdot \frac{2}{25} \iff \delta \geq \frac{25}{97} \approx 0.258. \quad (12)$$

The deviation profit is the same as without forward trading, $\frac{9}{64}$; the critical discount factor falls from $\frac{9}{17}$ to $\frac{25}{97}$ only because the punishment profit falls from $\frac{1}{9}$ to $\frac{2}{25}$.

The full punishment. When forward positions accumulate after a deviation, the punishment profit falls to zero over time (Figure 2). At the opening of a spot market, the firm does not deviate if

$$\frac{1-x}{8} + \frac{\delta}{1-\delta} \cdot \frac{1}{8} \geq \frac{(3-x)^2}{64} + \sum_{N=1}^{\infty} \delta^N \frac{1+N}{(3+2N)^2}. \quad (13)$$

The left-hand side increases in δ faster than the right-hand side, so (13) holds if and only if $\delta \geq \delta_C(x)$, where $\delta_C(x)$ solves (13) with equality. The sum has no simple closed form, so we solve for $\delta_C(x)$ numerically.

Result 5 (Proposition 3 of the paper, short positions). With quantity competition in the spot market and $0 \leq x \leq 1$:

- $\delta_C(0) = 0.237$, $\delta_C(\frac{1}{2}) = 0.391$, and $\delta_C(1) = 0.512$;
- $\delta_C(x)$ increases in x , and $\delta_C(x) < \frac{9}{17}$ for every $x \in [0, 1]$.

The paper reports $\delta_C(0) = 0.238$; solving (13) to more digits gives 0.2372. Figure 4 depicts $\delta_C(x)$ for $x \in [0, 1]$ (solid), together with the pure-spot cutoff $\frac{9}{17}$ with Cournot reversion and the cutoff $\frac{9}{32}$ with an optimal penal code. Every point between the solid curve and the line at $\frac{9}{17}$ is a discount factor at which collusion can be sustained with forward trading but not without it. The part of the figure for $x < 0$ is the subject of Step (e).

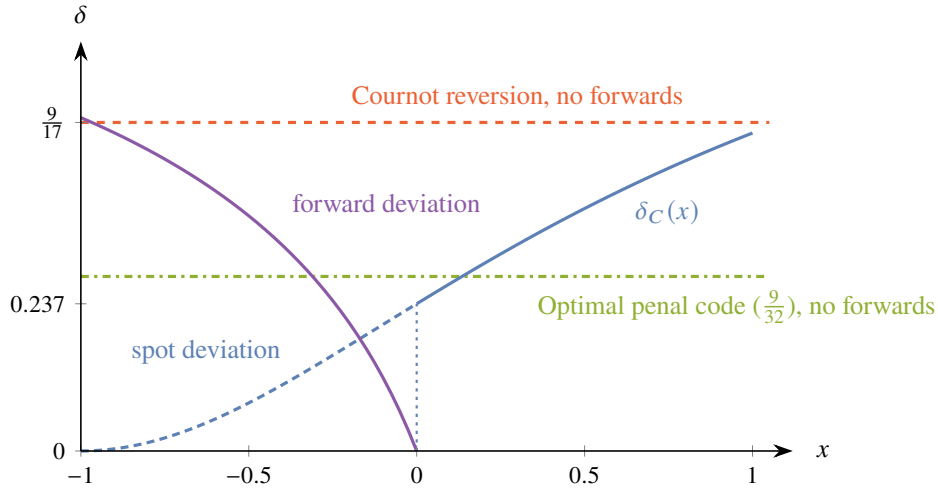


Figure 4: Critical discount factor with quantity competition in the spot market. For $x < 0$, the dashed curve is the cutoff against a spot-market deviation and the purple curve the cutoff against a forward-market deviation (Step (e)).

Intuitively, the two effects of Step (c) are still present, but a third one appears: the punishment is harsher than Cournot reversion, because after a deviation each firm keeps selling forward to commit to a larger output. This third effect does not depend on x , and it is the reason why $\delta_C(0) = 0.237$ lies below $\frac{9}{17}$ even though nobody sells forward along the collusive path: the threat of falling into a situation of heavy forward contracting is what deters a deviation. In contrast with price competition, a larger x now raises the critical discount factor, because with quantity competition forward sales lower the payoff from following the plan, $\frac{1-x}{8}$, by more than they lower the deviation profit, $\frac{(3-x)^2}{64}$.

Figure 5 compares the two sides of the no-deviation conditions (9) and (13) at $\delta = 0.4$, as functions of x . With price competition (panel (a)), the payoff from deviating falls faster than the payoff from following the plan, so collusion requires enough forward sales, $x \geq 1 - \frac{1}{\sqrt{3}} \approx 0.423$. With quantity competition (panel (b)), the payoff from following the plan falls faster, so collusion requires few forward sales, $x \leq 0.534$.

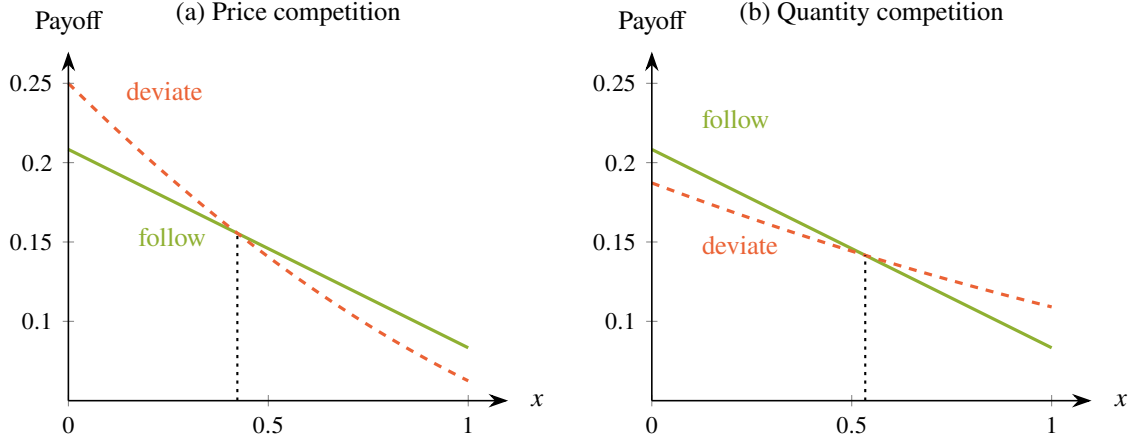


Figure 5: Payoff from following the collusive plan and from deviating in the spot market, as a function of the share x sold forward, for $\delta = 0.4$.

6. Step (e): Long Positions and Deviations in the Forward Market

When firms buy forwards along the collusive path ($x < 0$), the spot-market deviation becomes less attractive still: with $x = -1$, each firm buys its entire output in advance, and $\pi_d = \frac{1}{4}$ equals the payoff from following the plan in that period, so condition (13) holds for every δ (the dashed part of the blue curve in Figure 4). Long positions are not exotic: Mahenc and Salanié (2004) show that, with price competition in a one-shot market, firms buy part of their own output forward to soften competition. The paper concludes that $\delta_C(-1) = 0$. Its argument that a deviation in the forward market never pays, however, relies on the deviation profit $\frac{(1-x/4)^2}{8}$ being at most $\frac{1}{8}$, which holds only for $x \geq 0$.

With long positions, a firm can deviate in the forward market by selling forwards instead of buying them. Its rival has committed to buying $\frac{|x|}{4}$, which makes the rival less aggressive in the next spot market, and the deviator earns $\frac{(1-x/4)^2}{8} > \frac{1}{8}$ there, followed by the punishment path. The firm does not deviate in this way if

$$\frac{1}{1-\delta} \cdot \frac{1}{8} \geq \frac{(1-x/4)^2}{8} + \sum_{N=1}^{\infty} \delta^N \frac{1+N}{(3+2N)^2}. \quad (14)$$

Result 6 (long positions). With quantity competition in the spot market and $x < 0$:

- condition (14) fails for small δ , so the critical discount factor is at least the cutoff of (14), the purple curve in Figure 4;
- at $x = -1$, this cutoff is 0.537, above the pure-spot cutoff $\frac{9}{17}$;
- taking both deviations into account, the lowest cutoff, about 0.181, is reached with a small long position, $x \approx -0.17$.

The cutoff in Result 6 is a lower bound, since it considers only forward sales for the next spot market; a deviator could also sell forwards for later markets. The qualitative message of the paper survives against the deviations considered here: forward trading still lowers the critical discount factor below $\frac{9}{17}$ for every x above about -0.97 . But firms cannot rely on large long positions to sustain collusion at any discount factor, because a firm that is expected to buy forwards has an incentive to sell them instead. The general lesson is that a collusive plan with two markets must be checked against deviations in both of them.

7. Step (f): What the Toy Model Keeps and Drops

Table 1 collects the critical discount factors of Steps (a)–(e).

Table 1: Critical discount factors for sustaining the monopoly outcome ($p = 1 - Q$, zero costs).

Setting	Critical δ
Price competition, no forwards	$\frac{1}{2}$
Price competition, all output sold forward one period ahead ($x = 1$)	$\frac{1}{3}$
Quantity competition, no forwards, Cournot reversion	$\frac{9}{17} \approx 0.529$
Quantity competition, no forwards, optimal penal code	$\frac{9}{32} \approx 0.281$
Quantity competition, one forward round per period, $x = 0$	$\frac{25}{97} \approx 0.258$
Quantity competition, forwards, $x = 0$	0.237
Quantity competition, forwards, $x = 1$	0.512
Quantity competition, forwards, $x = -1$ (forward-market deviation)	at least 0.537

What the toy model keeps.

- The two effects of forward trading on collusion: fewer uncontracted sales along the collusive path (pro-competitive) and a smaller market for a deviator plus, with quantity competition, a harsher punishment (pro-collusive).
- The contrast between price and quantity competition: collusion calls for selling forward with price competition, but for selling little or nothing with quantity competition.
- The result that forward trading can help collusion even when no forwards are traded along the equilibrium path, because of the punishment that forward trading makes possible.

What it drops.

- Optimal punishments with forward trading. Without forward trading, an optimal penal code lowers the cutoff to $\frac{9}{32}$, which is still above $\delta_C(0) = 0.237$; the paper does not derive the optimal penal code with forward trading.
- Price setting in the forward market. If firms set the prices of their forward contracts, the punishment becomes competitive pricing, and with $x = 0$ the cutoff falls to $\frac{1}{9}$, since $\frac{1/8}{1-\delta} \geq \frac{9}{64}$ if and only if $\delta \geq \frac{1}{9}$.
- Contracting more than one period ahead, hedging demand from risk-averse firms or buyers, and more than two firms.
- Regulation. If a regulator requires firms to keep large forward sales, the pro-competitive effect can dominate. Forward contracting is therefore not, on its own, a remedy for market power in repeated markets such as electricity.

References

- Abreu, D. (1986). Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory*, 39(1), 191–225.
- Allaz, B., and Vila, J.-L. (1993). Cournot competition, forward markets and efficiency. *Journal of Economic Theory*, 59(1), 1–16.
- Liski, M., and Montero, J.-P. (2006). Forward trading and collusion in oligopoly. *Journal of Economic Theory*, 131(1), 212–230.
- Mahenc, P., and Salanié, F. (2004). Softening competition through forward trading. *Journal of Economic Theory*, 116(2), 282–293.

EconS 594 – Ph.D. Industrial Organization

Fall 2026

Collusion, Leniency Programs, and Whistleblower Rewards

Setup

Consider two firms selling a homogeneous product and competing à la Bertrand. They face a linear demand function $Q(p) = 1 - p$, where Q denotes aggregate output and p represents the market price. Both firms have the same marginal cost, normalized to $c = 0$. Firms interact in an infinitely repeated game with common discount factor $\delta \in (0, 1)$ and use grim-trigger strategies: any deviation from a collusive agreement triggers permanent reversion to the static Bertrand–Nash equilibrium.

We use the following notation throughout:

- π^m : monopoly profit. Solving $\max_{p \geq 0} p(1 - p)$ yields $p^m = \frac{1}{2}$ and $\pi^m = \frac{1}{4}$.
- $\pi^c = \frac{\pi^m}{2} = \frac{1}{8}$: each firm's collusive profit (equal split).
- $\pi^d = \pi^m = \frac{1}{4}$: deviation profit (undercut the collusive price by ε and capture all sales).
- $\pi^n = 0$: Nash equilibrium profit (Bertrand with symmetric costs).

Part (a). No fines

When the competition authority (CA) does not monitor collusion, each firm compares the present value of colluding forever with the payoff from a one-shot deviation followed by permanent Nash punishment. Collusion is sustainable if

$$\frac{\pi^c}{1 - \delta} \geq \pi^d + \frac{\delta}{1 - \delta} \pi^n.$$

Substituting $\pi^c = \frac{1}{8}$, $\pi^d = \frac{1}{4}$, and $\pi^n = 0$, and rearranging:

$$\frac{1}{8} \geq (1 - \delta) \frac{1}{4} \quad \implies \quad \delta \geq \bar{\delta} = \frac{1}{2}.$$

Interpretation. When firms are sufficiently patient ($\delta \geq 1/2$), the long-run gain from collusion outweighs the short-run temptation to deviate. This threshold, $\bar{\delta} = 1/2$, is the benchmark we compare against when enforcement tools are introduced below.

Part (b). Antitrust audits

Suppose the CA audits the industry in every period with probability $\alpha \in [0, 1]$. If collusion is detected, each firm pays a fine $F > 0$. Both colluding and deviating firms face the expected fine αF in the current period (the cartel was active that period). The sustainability condition becomes

$$\frac{1}{1 - \delta} (\pi^c - \alpha F) \geq (\pi^d - \alpha F) + \frac{\delta}{1 - \delta} \pi^n.$$

Since $\pi^c = \frac{1}{8}$, $\pi^d = \frac{1}{4}$, and $\pi^n = 0$:

$$\frac{1}{8} - \alpha F \geq (1 - \delta) \left(\frac{1}{4} - \alpha F \right).$$

Solving for δ :

$$\delta \geq \bar{\delta}_A(\alpha, F) \equiv \frac{1}{2(1 - 4\alpha F)}.$$

This threshold falls below one, $\bar{\delta}_A < 1$, only when $\alpha F < \frac{1}{8}$, that is, when

$$F < \frac{1}{8\alpha}.$$

If instead $F \geq \frac{1}{8\alpha}$, then $\bar{\delta}_A \geq 1$ and no discount factor $\delta \in (0, 1)$ sustains collusion: the expected fine alone deters the cartel, leaving nothing for a leniency or whistleblower program to add. We therefore maintain

$$F < \frac{1}{8\alpha}$$

for the rest of the handout, so that collusion is sustainable in this audit benchmark and the enforcement tools below have a cartel to break.

Comparative statics.

$$\frac{\partial \bar{\delta}_A}{\partial \alpha} = \frac{2F}{(1 - 4\alpha F)^2} > 0, \quad \frac{\partial \bar{\delta}_A}{\partial F} = \frac{2\alpha}{(1 - 4\alpha F)^2} > 0.$$

Both a more severe fine (F) and a more likely audit (α) raise the minimal discount factor, making collusion harder to sustain. Since $\bar{\delta}_A \geq \bar{\delta} = \frac{1}{2}$ for all $\alpha F \geq 0$, audits unambiguously hinder collusion.

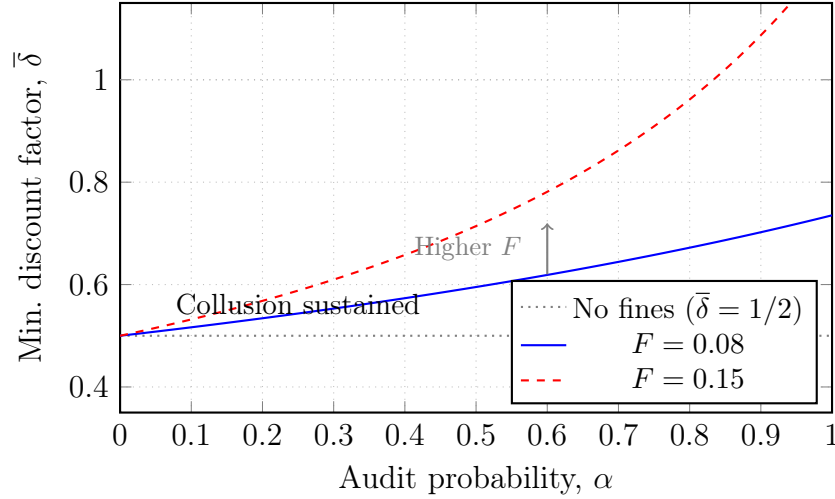


Figure 1. Minimal discount factor under audits, $\bar{\delta}_A(\alpha, F) = \frac{1}{2(1 - 4\alpha F)}$. An increase in F shifts the curve upward, shrinking the region where collusion can be sustained ($\delta \geq \bar{\delta}_A$).

Part (c). Leniency programs

The CA now offers a leniency program: a firm that provides evidence about collusion receives a reward R , where $R > -\alpha F$. The reward can be negative (a reduced sentence, $0 > R > -\alpha F$) or positive (a cash bounty plus no jail time, $R > 0$). A deviating firm can simultaneously deviate and report the cartel, collecting R instead of facing the expected fine. Recall from Part (b) that we maintain $F < \frac{1}{8\alpha}$, so that collusion is sustainable to begin with and the program has a cartel to deter.

The sustainability condition becomes

$$\frac{1}{1 - \delta} (\pi^c - \alpha F) \geq (\pi^d + R) + \frac{\delta}{1 - \delta} \pi^n.$$

The colluding firm still faces the expected fine αF , but the deviating firm now receives R instead. Substituting and simplifying:

$$\frac{1}{8} - \alpha F \geq (1 - \delta) \left(\frac{1}{4} + R \right).$$

Solving for δ :

$$\delta \geq \bar{\delta}_L(\alpha, F, R) \equiv \frac{1 + 8(R + \alpha F)}{2(1 + 4R)}.$$

Comparison with audits. We have $\bar{\delta}_L \geq \bar{\delta}_A$ if and only if

$$\frac{1 + 8(R + \alpha F)}{2(1 + 4R)} \geq \frac{1}{2(1 - 4\alpha F)}.$$

Cross-multiplying:

$$[1 + 8(R + \alpha F)](1 - 4\alpha F) \geq 1 + 4R.$$

Expanding and simplifying yields $4(R + \alpha F)(1 - 8\alpha F) \geq 0$, which holds since $R + \alpha F > 0$ and, because $\alpha F < \frac{1}{8}$ from Part (b), $1 - 8\alpha F > 0$. Therefore $\bar{\delta}_L \geq \bar{\delta}_A$: leniency programs hinder collusion beyond audits alone.

Comparative statics.

$$\frac{\partial \bar{\delta}_L}{\partial R} = \frac{2(1 - 8\alpha F)}{(1 + 4R)^2}, \quad \frac{\partial \bar{\delta}_L}{\partial \alpha} = \frac{4F}{1 + 4R} > 0, \quad \frac{\partial \bar{\delta}_L}{\partial F} = \frac{4\alpha}{1 + 4R} > 0.$$

A more severe or more likely fine increases $\bar{\delta}_L$, hindering collusion. Since $F < \frac{1}{8\alpha}$ from Part (b), we have $1 - 8\alpha F > 0$, so a more generous reward R also increases $\bar{\delta}_L$, making collusion harder to sustain.

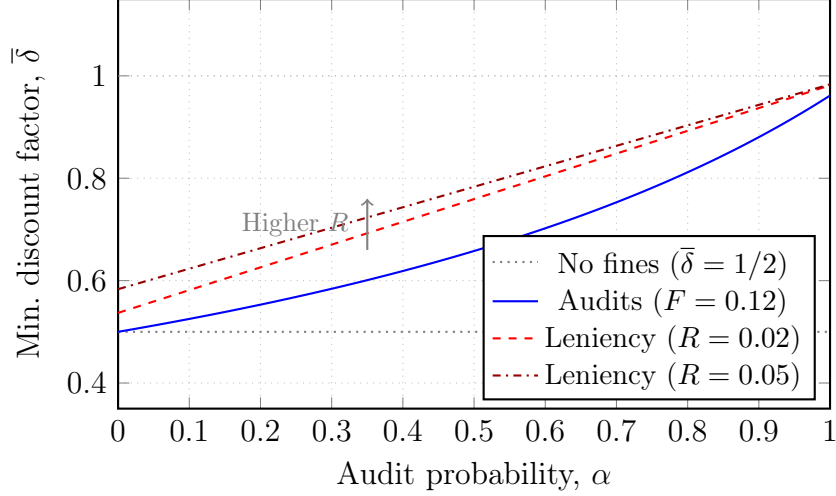


Figure 2. Minimal discount factor under audits vs. leniency (with $F = 0.12$). The leniency curve lies above the audit curve for all α , and an increase in R shifts it further upward.

Part (d). Whistleblower programs

This part extends the analysis following Aubert, Rey, and Kovacic (2006). That paper distinguishes between leniency programs (which reduce the fine paid by a cartel *firm* that self-reports) and whistleblower programs (which pay a bounty to *employees* who report the cartel to the CA).

Model. Suppose each firm employs $n \geq 1$ workers who observe whether the firm participates in the cartel. The CA offers a whistleblower bounty $B > 0$ to any employee who reports evidence of collusion. If an employee reports, the CA obtains hard evidence, convicts the cartel, and the employee collects B .

To prevent its n employees from reporting, each colluding firm must pay each worker a “bribe” of at least B per period. The total cost of silencing all employees is therefore nB per period per firm, as in Aubert et al. (2006). A deviating firm must also pay the bribe in the current period, since its employees observed past collusive communication and could still report to the CA. After the cartel collapses (Nash reversion), there is no collusive evidence to report, so no bribes are needed.

Collusion condition. With the whistleblower program in place (alongside audits), per-period collusive profits become $\pi^c - \alpha F - nB$, deviation profits become $\pi^d - \alpha F - nB$, and

Nash-reversion profits remain $\pi^n = 0$. The sustainability condition is

$$\frac{1}{1-\delta}(\pi^c - \alpha F - nB) \geq (\pi^d - \alpha F - nB) + \frac{\delta}{1-\delta}\pi^n.$$

Substituting and simplifying:

$$\frac{1}{8} - \alpha F - nB \geq (1-\delta)\left(\frac{1}{4} - \alpha F - nB\right).$$

Solving for δ :

$$\delta \geq \bar{\delta}_W(\alpha, F, n, B) \equiv \frac{1}{2(1 - 4\alpha F - 4nB)}.$$

This expression has the same structure as the audit threshold $\bar{\delta}_A$ but with αF replaced by $\alpha F + nB$. It is well defined when $\alpha F + nB < \frac{1}{8}$.

Comparative statics.

$$\frac{\partial \bar{\delta}_W}{\partial B} = \frac{2n}{(1 - 4\alpha F - 4nB)^2} > 0, \quad \frac{\partial \bar{\delta}_W}{\partial n} = \frac{2B}{(1 - 4\alpha F - 4nB)^2} > 0.$$

A larger bounty (B) and more informed employees (n) both raise $\bar{\delta}_W$, making collusion harder to sustain. The mechanism operates through the cost of keeping employees silent: higher B or more employees forces each firm to spend more on bribes, which erodes the per-period return from collusion.

Comparison with audits. Since $nB > 0$, the denominator in $\bar{\delta}_W$ is smaller than that in $\bar{\delta}_A$, so $\bar{\delta}_W > \bar{\delta}_A$ for all parameter values: whistleblower programs provide additional deterrence beyond audits alone.

Comparison with leniency. The ranking between $\bar{\delta}_W$ and $\bar{\delta}_L$ depends on the magnitude of the total bribe nB relative to the leniency reward R . Comparing:

$$\bar{\delta}_W = \frac{1}{2(1 - 4\alpha F - 4nB)} \quad \text{vs.} \quad \bar{\delta}_L = \frac{1 + 8(R + \alpha F)}{2(1 + 4R)}.$$

The whistleblower program is more effective ($\bar{\delta}_W \geq \bar{\delta}_L$) if and only if

$$1 + 4R \geq [1 + 8(R + \alpha F)](1 - 4\alpha F - 4nB) \quad \Longleftrightarrow$$

$$nB \geq \frac{\left(\frac{1}{8} - \alpha F\right)(\alpha F + R)}{\frac{1}{8} + \alpha F + R}.$$

Under the maintained assumption $\alpha F < \frac{1}{8}$, the right-hand side is positive, so the whistleblower program dominates leniency only if the total bribe nB exceeds this threshold.

Numerical illustration. Set $F = 0.12$, $R = 0.04$, $n = 2$, $B = 0.02$ (so $nB = 0.04$). At $\alpha = 0$, the threshold equals $\frac{(1/8)(0.04)}{1/8 + 0.04} = \frac{0.005}{0.165} \approx 0.030$, which is less than $nB = 0.04$. As α increases, the threshold decreases further. Hence $\bar{\delta}_W > \bar{\delta}_L$ for all α under these parameters.

Full ranking. Combining the results across parts (a)–(d):

$$\bar{\delta}_W \geq \bar{\delta}_L \geq \bar{\delta}_A \geq \bar{\delta} = \frac{1}{2}.$$

Each enforcement tool shrinks the set of discount factors $[\bar{\delta}, 1)$ that sustain collusion. The whistleblower program provides the strongest deterrence (when nB is large enough), followed by leniency, audits, and no enforcement.

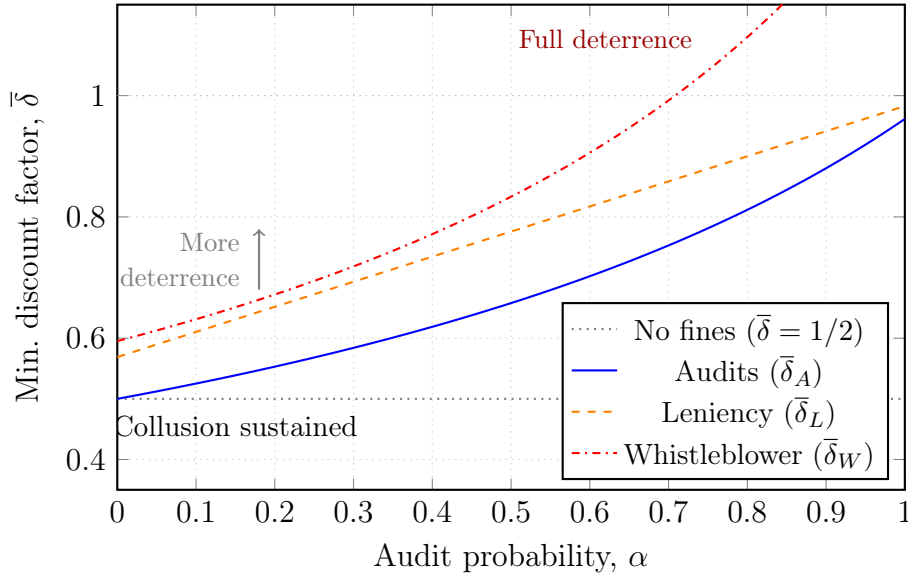


Figure 3. Minimal discount factor under all four regimes ($F = 0.12$, $R = 0.04$, $n = 2$, $B = 0.02$). For a given α , collusion is sustained for $\delta \geq \bar{\delta}$; above $\bar{\delta} = 1$, collusion is deterred for all discount factors. An increase in F , R , n , or B shifts the respective curve upward.

Interpretation. Aubert, Rey, and Kovacic (2006) emphasize that whistleblower programs attack cartel stability from a different margin than leniency: rather than rewarding a firm

for betraying its partner, they reward an *individual* for betraying the firm. This forces the cartel to internalize the cost of keeping employees silent (nB per period), which reduces the per-period return from collusion even when the firms themselves would never self-report. The two instruments are complementary:

- leniency programs destabilize the cartel agreement between firms, by making deviation more attractive through the reward R ;
- whistleblower rewards raise the operating cost of the cartel, by forcing silence payments to n employees.

These programs in practice

The two instruments in this exercise are used by competition authorities around the world. The modern leniency program began in the United States, where the Department of Justice adopted a corporate leniency policy in 1978 and, after it went almost unused, rewrote it in 1993 to grant automatic amnesty to the first firm that reports a cartel (the reward R of Part (c)). The European Commission issued its first Leniency Notice in 1996, revising it in 2002 and 2006. Adoption then spread to developing and emerging economies: Brazil enacted a leniency program in 2000, South Africa in 2004, and India in 2009, each granting the first cooperating firm full or near-full immunity. The whistleblower bounty B of Part (d) is also real. South Korea introduced individual informant rewards in 2002, paying an employee who reports a cartel up to 3 billion won, and the United Kingdom's Competition and Markets Authority raised its cartel whistleblower reward to £250,000 in 2023.

Two cases show the mechanism at work. In the international vitamins cartel, Rhône-Poulenc (later Aventis) escaped criminal fines in 1999 by being the first member to cooperate under the U.S. program, while its partners F. Hoffmann-La Roche and BASF pleaded guilty and paid \$500 million and \$225 million. In the air-cargo cartel, Lufthansa's leniency application set off a worldwide investigation that led the European Commission to fine 11 carriers about 799 million euros in 2010.

References

Aubert, C., P. Rey, and W.E. Kovacic (2006). “The Impact of Leniency and Whistleblowing Programs on Cartels.” *International Journal of Industrial Organization*, 24(6), pp. 1241–1266.

European Commission (2006). *Notice on Immunity from Fines and Reduction of Fines in Cartel Cases*, OJ C 298, 8 December 2006.

U.S. Department of Justice, Antitrust Division (1993). *Corporate Leniency Policy*.

Communicating Quality: Disclosure or Price Signaling

A step-by-step exercise in the spirit of Daughety and Reinganum (2008)

Companion note to Daughety, A. F., and J. F. Reinganum (2008), “Communicating Quality: A Unified Model of Disclosure and Signalling,” *RAND Journal of Economics* 39(4): 973–989. Full references at the end.

Overview. A firm that knows the quality of its product can tell consumers about it in two ways. It can *disclose* the quality through a credible channel, such as an independent audit, at a cost, or it can *signal* the quality through an action that consumers observe, such as a high price. The two literatures have developed separately. Disclosure models assume that marginal cost does not depend on quality, so a firm that does not disclose can only pool with others; signaling models assume that higher quality is costlier to produce and ignore disclosure. Daughety and Reinganum (2008) put the two together: when higher quality is costlier to produce, the alternative to disclosure is not silence but a revealing price, so disclosure and signaling are substitutes, and the firm picks the less costly channel. The main results are that the lowest type signals without distortion, that higher types distort their price upward and so earn less, that types above a cutoff disclose, and therefore that profit is U-shaped in quality; that voluntary disclosure is socially *insufficient*, whereas the classical pooling analysis finds it excessive; and that a decentralized subsidy can restore efficiency. This note builds the model with a linear demand and a uniform prior, so that every number can be computed. Section 1 sets up the model; Step (a) solves the full-information benchmark, which is the outcome under disclosure; Step (b) solves the separating signaling equilibrium; Step (c) shows how signaling costs profit and how the loss grows with quality; Step (d) solves the firm’s choice between the two channels; Step (e) compares it with the social optimum and describes mandatory disclosure and a subsidy; Step (f) contrasts the result with the classical model in which quality does not affect cost; and Step (g) lists what the toy model keeps and drops.

1. The Toy Model

Quality. A monopolist produces a good of quality $\theta \in [\underline{\theta}, \bar{\theta}] = [0, 1]$, which it observes and consumers do not. Quality is the probability that a unit is completely satisfactory, and consumers’ prior is uniform on the unit interval. Step (f) uses the same type space and prior.

Consumers. A consumer’s utility from q units is $(\alpha - (1 - \tilde{\theta})\delta)q - \frac{\beta}{2}q^2 + I - pq$, where $\tilde{\theta}$ is the perceived quality, p is the price, and I is income. A satisfactory unit yields a linear coefficient α and a disappointing unit yields $\alpha - \delta$, so a higher perceived quality shifts demand out, and demand is

$$q(p, \tilde{\theta}) = \frac{A(\tilde{\theta}) - p}{\beta}, \quad A(\tilde{\theta}) \equiv \alpha - (1 - \tilde{\theta})\delta. \quad (1)$$

Perceived quality equals θ if the firm discloses. Otherwise consumers form it from the price, by a belief function $B(p)$ that is part of a perfect Bayesian equilibrium.

The firm. A firm of type θ produces at constant marginal cost $k\theta$, so higher quality is costlier, and earns

$$\pi(p, \theta, \tilde{\theta}) = \frac{(p - k\theta)(A(\tilde{\theta}) - p)}{\beta}. \quad (2)$$

Once it knows θ , the firm chooses its price and whether to disclose, which costs $D > 0$ (for example, an audit). Disclosure is truthful, so the perceived quality is θ .

Parameter values. $\alpha = \frac{5}{2}$, $\beta = 1$, $\delta = 1$, and $k = \frac{1}{2}$, so $A(\bar{\theta}) = \frac{3}{2} + \bar{\theta}$ and $q = \frac{3}{2} + \bar{\theta} - p$. The choices $\beta = \delta = 1$ are normalizations. The paper assumes $0 < \underline{\theta} < \bar{\theta} < 1$, so the unit type space is a limiting case, which the paper itself uses in its Section 4. The value $\alpha = \frac{5}{2}$ keeps the intercept $A(\bar{\theta}) = \frac{3}{2} + \bar{\theta}$ simple while satisfying the paper's assumptions: quality is socially valuable, $\delta > k$ (Assumption 1); every type can earn a positive margin at some price, $A(\underline{\theta}) = \frac{3}{2} > k\bar{\theta} = \frac{1}{2}$ (Assumption 2); and a slightly stronger version used in the proofs, $A(\underline{\theta}) = \frac{3}{2} > k\bar{\theta} + k(\bar{\theta} - \underline{\theta}) = 1$ (Assumption A3). Step (f) reuses α, β , and δ with $k = 0$. Figure 1 depicts demand (1) for three perceived qualities: each unit of perceived quality raises the intercept $A(\bar{\theta})$ by $\delta = 1$, so the same price sells more when consumers believe quality is higher, which is why the firm cares about perceptions.

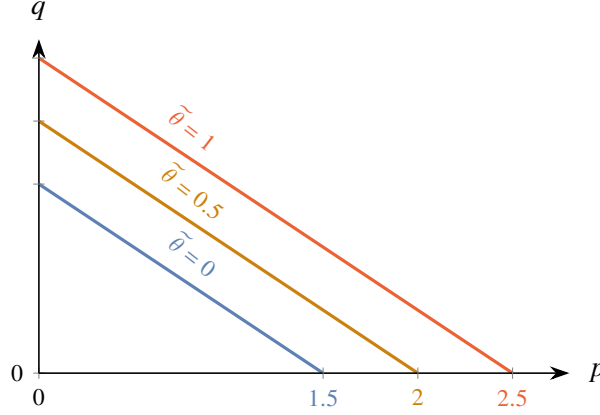


Figure 1: Demand $q = \frac{3}{2} + \bar{\theta} - p$ for three perceived qualities; the intercepts on the price axis are $A(\bar{\theta}) = \frac{3}{2} + \bar{\theta}$.

2. Step (a): Full Information

Suppose consumers observe θ . The firm maximizes (2) with $\bar{\theta} = \theta$, so

$$P^f(\theta) = \frac{A(\theta) + k\theta}{2}, \quad \pi^f(\theta) = \frac{(A(\theta) - k\theta)^2}{4\beta}, \quad CS^f(\theta) = \frac{(A(\theta) - P^f(\theta))^2}{2\beta}, \quad (3)$$

the price, profit, and consumer surplus. In the toy, $P^f(\theta) = \frac{3}{4}(1 + \theta)$, $\pi^f(\theta) = \frac{(3+\theta)^2}{16}$, and $CS^f(\theta) = \frac{(3+\theta)^2}{32}$. All three rise with quality: a unit of quality raises the demand intercept by δ and marginal cost by $k < \delta$. A firm that discloses pays D and charges $P^f(\theta)$, earning $\pi^f(\theta) - D$. At the two ends of the interval, $P^f = 0.75$ and $\pi^f = 0.5625$ at $\theta = 0$, and $P^f = 1.5$ and $\pi^f = 1$ at $\theta = 1$.

3. Step (b): The Separating Signaling Equilibrium

Now suppose disclosure is too costly to be used, and consumers infer quality from the price. A *separating* equilibrium is a strictly increasing price function $P^s(\theta)$ and beliefs $B^*(p)$ such that each type prefers its price to any other, $\pi(P^s(\theta), \theta, \theta) \geq \max_p \pi(p, \theta, B^*(p))$, and beliefs are correct, $B^*(P^s(\theta)) = \theta$.

The differential equation. A type that charges p is perceived as $B(p)$, so its profit is $\pi(p, \theta, B(p))$. The first-order condition in p is

$$(A(B(p)) - p) + (p - k\theta)(\delta B'(p) - 1) = 0.$$

In equilibrium $B(P^s(\theta)) = \theta$ and $B'(p) = \frac{1}{P^s(\theta)}$, so P^s solves

$$\frac{dp}{d\theta} = \frac{\delta(p - k\theta)}{2p - A(\theta) - k\theta} = \frac{p - \frac{1}{2}\theta}{2p - \frac{3}{2} - \frac{3}{2}\theta} \quad \text{in the toy.} \quad (4)$$

The numerator is positive. The denominator is positive if and only if $p > P^f(\theta)$, since $2P^f = A + k\theta$. So the equilibrium price rises with quality, has a slope above δ (because $p - k\theta > 2p - A - k\theta$ when $p < A$), and lies above the full-information price wherever it is defined. The lowest type has no one below it to deter, so it has no reason to distort: the initial condition is $P^s(\underline{\theta}) = P^f(\underline{\theta}) = 0.75$, where the denominator is zero and the slope is infinite.

The solution. The paper solves (4) implicitly:

$$(A(\theta) - p)^{\delta-k} \left(2p - k\theta + \frac{k(\alpha - \delta)}{\delta - k} \right)^\delta = K, \quad \text{that is,} \quad \sqrt{\frac{3}{2} + \theta - p} \left(2p - \frac{1}{2}\theta + \frac{3}{2} \right) = K \quad (5)$$

in the toy. Differentiating the logarithm of the left side along (4) gives zero, so the left side is constant along the solution, and K is set by the boundary condition at $\theta = 0$: $K = \sqrt{0.75} \cdot 3 = \frac{3}{2}\sqrt{3} = 2.598$. Solving for p at each θ gives the schedule of Table 1.

Result 1 (Proposition 1 of the paper). There is a unique separating perfect Bayesian equilibrium:

- the lowest type charges its full-information price, $P^s(\underline{\theta}) = P^f(\underline{\theta})$;
- every higher type charges more than its full-information price, and P^s is strictly increasing and concave, given by (5);
- beliefs are $B^*(p) = (P^s)^{-1}(p)$ for prices on the schedule, $\underline{\theta}$ below it, and $\bar{\theta}$ above it.

Intuitively, a higher price is a costly signal because it cuts the quantity sold, and a lower-quality type, which has the lower cost $k\theta$, would gain from being taken for a higher type. To deter that, the price schedule must climb faster than the value of quality to consumers, δ , and the distortion accumulates with quality. In the toy, $P^s(1) = 2.282$ against $P^f(1) = 1.5$, and quantity at the top falls from 1 to $A - P^s = 0.218$.

Figure 2 depicts the two price functions against quality. They coincide at $\theta = 0$, the signaling price rises steeply at first and then flattens (concave), and the gap, which is the price distortion, widens with quality.

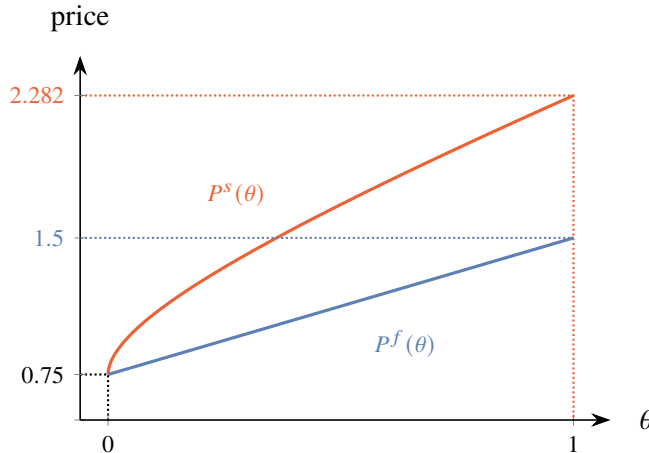


Figure 2: Signaling price $P^s(\theta)$ and full-information price $P^f(\theta) = \frac{3}{4}(1 + \theta)$ against quality, with $\delta = 1$ and $k = \frac{1}{2}$.

The equilibrium price is the best reply of each type to the beliefs. Figure 3 depicts, for the type $\theta = 0.5$, its profit against the price it charges. The blue curve is the profit if consumers perceived the true quality at every price, which peaks at the full-information price $P^f(0.5) = 1.1250$ with height $\pi^f = 0.7656$. The red curve is the profit under the equilibrium beliefs $B^*(p)$: consumers infer a higher quality from a higher price, which raises demand along the curve. The red curve lies below the blue one at lower prices, where consumers underestimate quality, and above it at higher prices, where they overestimate it; the two cross at $P^s(0.5) = 1.6833$, where beliefs are correct, and the red curve peaks there with height $\pi^s = 0.4539$. Under these beliefs the type does not gain from any other price.

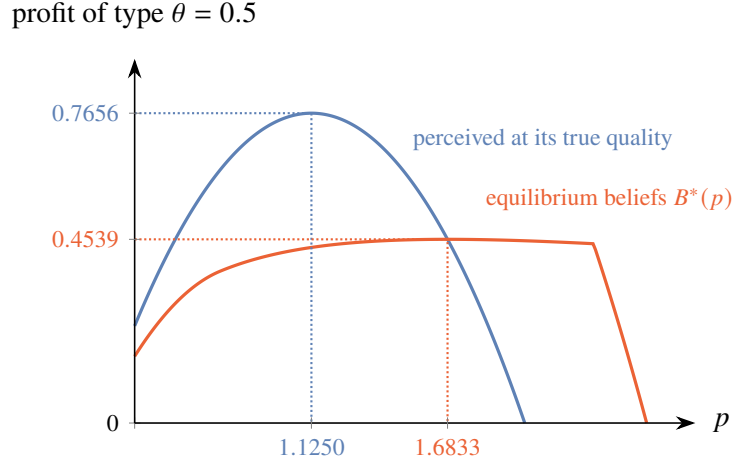


Figure 3: Profit of type $\theta = 0.5$ against its price, if perceived at its true quality at every price (blue) and under the equilibrium beliefs (red).

4. Step (c): The Cost of Signaling

Substituting the equilibrium price, the profit of type θ under signaling is $\pi^s(\theta) = \frac{(P^s - k\theta)(A - P^s)}{\beta}$. Beliefs are correct in equilibrium and the first-order condition holds, so by the envelope theorem only the direct effect of quality on cost remains,

$$\frac{d\pi^s}{d\theta} = -\frac{k(A(\theta) - P^s(\theta))}{\beta} = -k q^s(\theta) < 0, \quad (6)$$

while the full-information profit rises, $\frac{d\pi^f}{d\theta} = \frac{(\delta - k)(A - k\theta)}{2\beta} > 0$. Signaling profit *falls* with quality because of the price distortion, and the two coincide at the lowest type, $\pi^s(\underline{\theta}) = \pi^f(\underline{\theta})$. The profit loss from having to signal, $\Delta\pi \equiv \pi^f - \pi^s$, is therefore strictly increasing in quality,

$$\frac{d\Delta\pi}{d\theta} = \frac{(\delta - k)(A - k\theta)}{2\beta} + k q^s(\theta) > 0. \quad (7)$$

Result 2 (Proposition 2 of the paper). The difference $\pi^f(\theta) - \pi^s(\theta)$ is strictly increasing in θ , from zero at the lowest type.

Table 1 reports the two prices and the two profits. The loss rises from 0 at $\theta = 0$ to 0.6115 at $\theta = 1$, and signaling profit falls from 0.5625 to 0.3885 while full-information profit rises from 0.5625 to 1.

θ	P^f	P^s	π^f	π^s	$\Delta\pi = \pi^f - \pi^s$
0.0	0.7500	0.7500	0.5625	0.5625	0.0000
0.2	0.9000	1.2613	0.6400	0.5094	0.1306
0.4	1.0500	1.5520	0.7225	0.4705	0.2520
0.6	1.2000	1.8093	0.8100	0.4388	0.3712
0.8	1.3500	2.0505	0.9025	0.4119	0.4906
1.0	1.5000	2.2820	1.0000	0.3885	0.6115

Table 1: Prices and profits under full information (f) and under signaling (s), with $\alpha = \frac{5}{2}$, $\beta = 1$, $\delta = 1$, and $k = \frac{1}{2}$.

5. Step (d): The Choice between Signaling and Disclosure

A type strictly prefers to disclose if and only if $\pi^f(\theta) - D > \pi^s(\theta)$, that is, if and only if $\Delta\pi(\theta) > D$. By Result 2, there are three cases:

- if $D \geq \Delta\pi(\bar{\theta}) = 0.6115$, every type signals;
- if $D = 0$, every type discloses;
- if $0 < D < \Delta\pi(\bar{\theta})$, there is a marginal type θ^V with $\Delta\pi(\theta^V) = D$: types below it signal and types above it disclose. The marginal type is indifferent and, as in the paper, is assumed to disclose.

A nondisclosing price above $P^s(\theta^V)$ is off the equilibrium path, and consumers treat it as coming from the type $B^*(p)$; the incentive constraints then guarantee that no disclosing type gains by signaling instead. The firm's overall profit is

$$\Pi(\theta) = \begin{cases} \pi^s(\theta) & \text{if } \theta < \theta^V, \\ \pi^f(\theta) - D & \text{if } \theta \geq \theta^V, \end{cases} \quad (8)$$

which is continuous. It falls on $[\underline{\theta}, \theta^V]$, because π^s falls, and rises on $[\theta^V, \bar{\theta}]$, because π^f rises: profit is *U-shaped* in quality. The highest type earns more than the lowest if $\pi^f(\bar{\theta}) - D > \pi^s(\underline{\theta}) = \pi^f(\underline{\theta})$, that is, if $D < D^{\text{mod}} \equiv \pi^f(\bar{\theta}) - \pi^f(\underline{\theta})$, a *moderate* disclosure cost.

Result 3 (voluntary disclosure and U-shaped profit). For $0 < D < \Delta\pi(\bar{\theta})$:

- types below θ^V signal and types above it disclose, where $\Delta\pi(\theta^V) = D$;
- overall profit $\Pi(\theta)$ is U-shaped in θ , with its minimum at θ^V ;
- if $D < D^{\text{mod}}$, the highest type earns the most; if $D^{\text{mod}} < D < \Delta\pi(\bar{\theta})$, the lowest type does.

In the toy, $D^{\text{mod}} = 1 - 0.5625 = 0.4375$. For $D = 0.3$ (moderate), $\theta^V = 0.480$, the minimum profit is $\Pi(\theta^V) = 0.4571$, and $\Pi(\bar{\theta}) = 0.7000 > \Pi(\underline{\theta}) = 0.5625$. For $D = 0.5$ (between D^{mod} and 0.6115), $\theta^V = 0.816$, the minimum is 0.4099, and $\Pi(\bar{\theta}) = 0.5000 < \Pi(\underline{\theta})$. Disclosure at moderate cost partly resolves a problem of many price-signaling models: there, higher-quality firms earn less, so firms would want to produce low quality, whereas here the highest type earns the most.

Figure 4 depicts the profits for $D = 0.3$ in panel (a) and $D = 0.5$ in panel (b). The red curve is signaling profit π^s and the blue curves are full-information profit π^f (solid) and net of the disclosure cost (dashed).

The firm earns the higher of the red curve and the dashed blue curve, shown in green, and the two cross at θ^V . In panel (a) the right end of the green curve is above its left end, and in panel (b) it is below.

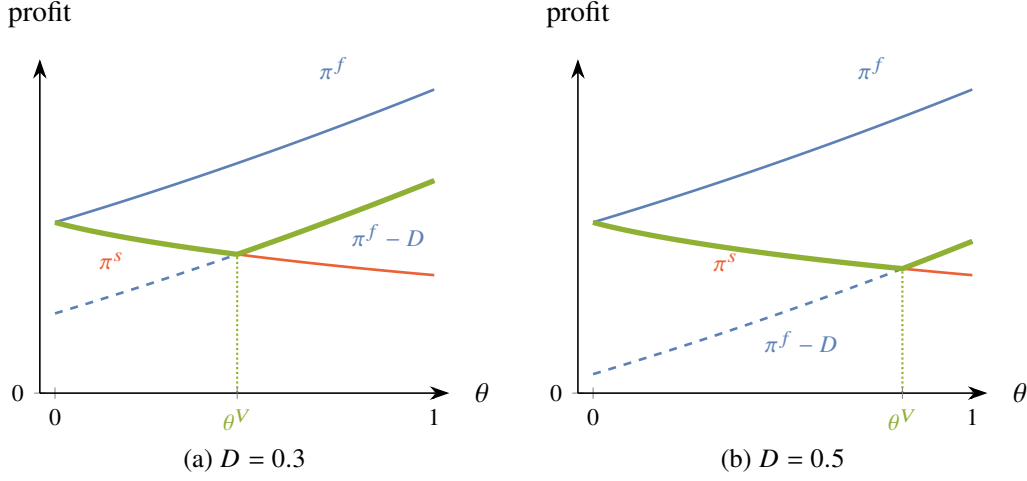


Figure 4: Full-information profit π^f , its value net of the disclosure cost, signaling profit π^s , and the firm's overall profit Π (green) against quality.

6. Step (e): Social Efficiency, Mandatory Disclosure, and a Subsidy

The firm weighs only its own profit when it chooses its channel. Disclosure also lowers the price from the signaling schedule to the full-information line and raises output, which benefits consumers. Define consumer surplus under signaling as $CS^s(\theta) = \frac{(A - P^s)^2}{2\beta}$ and the surplus gain from disclosure as $\Delta CS(\theta) \equiv CS^f(\theta) - CS^s(\theta)$, which is strictly increasing in θ (Proposition 3 of the paper). Table 2 reports the numbers. Let $W(\hat{\theta})$ be total surplus when types below $\hat{\theta}$ signal and types above it disclose,

$$W(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} (CS^s + \pi^s) dG(\theta) + \int_{\hat{\theta}}^{\bar{\theta}} (CS^f + \pi^f - D) dG(\theta), \quad (9)$$

so that

$$W'(\hat{\theta}) = [D - \Delta\pi(\hat{\theta}) - \Delta CS(\hat{\theta})] g(\hat{\theta}). \quad (10)$$

The bracket equals $D > 0$ at $\underline{\theta}$ and decreases in $\hat{\theta}$, by Result 2 and Proposition 3. When $0 < D < \Delta\pi(\bar{\theta})$, the case the paper studies, it equals $-\Delta CS(\theta^V) < 0$ at θ^V , so surplus is maximized at the unique θ^W with $\Delta\pi(\theta^W) + \Delta CS(\theta^W) = D$. The firm's own cutoff solves $\Delta\pi(\theta^V) = D$. Since $\Delta CS > 0$, $\theta^W < \theta^V$.

Result 4 (Section 3 of the paper).

- Voluntary disclosure is socially insufficient: types in $[\theta^W, \theta^V)$ signal although surplus would be higher if they disclosed.
- Mandatory disclosure lowers prices and raises output, but it makes types below θ^W pay D although surplus is higher if they signaled.
- A subsidy $s(\theta; D) = \pi^s(\theta) - (\pi^f(\theta) - D)$ for $\theta \in [\theta^W, \theta^V]$ and zero otherwise, paid to a type that obtains an audit, induces the efficient channel with the smallest payments that achieve it and leaves the firm's overall profit $\Pi(\theta)$ unchanged.

In the toy with $D = 0.3$, $\theta^V = 0.480$ and $\theta^W = 0.152$. Figure 5 depicts the two conditions: θ^V is where the red curve $\Delta\pi$ reaches D , and θ^W is where the green curve $\Delta\pi + \Delta CS$ does. The subsidy equals $D - \Delta\pi(\theta)$: at θ^W it equals $\Delta CS(\theta^W) = 0.1995$, the consumer gain that the firm does not internalize, and it falls to zero at θ^V . Table 3 reports average surplus per type, with the uniform prior. For $D = 0.3$, voluntary disclosure gives 0.8170, mandatory disclosure gives 0.8563, and the optimum gives 0.8728, against 0.5271 if every type must signal. For $D = 0.5$ the numbers are 0.6204, 0.6563, and 0.7217, so the gap between the voluntary outcome and the optimum is larger when disclosure is costlier.

θ	CS^f	CS^s	ΔCS
0.0	0.2813	0.2813	0.0000
0.2	0.3200	0.0962	0.2238
0.4	0.3613	0.0606	0.3007
0.6	0.4050	0.0423	0.3627
0.8	0.4513	0.0311	0.4201
1.0	0.5000	0.0238	0.4762

Table 2: Consumer surplus under full information and under signaling.

gain from disclosure

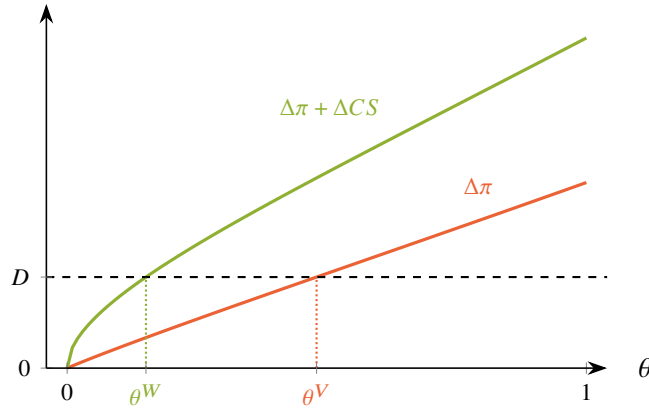


Figure 5: Profit gain $\Delta\pi$ and total gain $\Delta\pi + \Delta CS$ from disclosing, against quality, with the disclosure cost $D = 0.3$.

7. Step (f): The Classical Disclosure Model, Where Quality Does Not Affect Cost

The classical disclosure literature finds that a monopolist discloses too *much*. The paper attributes the difference to two assumptions: that cost does not depend on quality, so that nondisclosure can only pool, and that demand is inelastic. Set $k = 0$ and keep everything else unchanged: θ uniform on $[0, 1]$ and $\alpha = \frac{5}{2}$, $\beta = \delta = 1$, so that $A(\theta) = \frac{3}{2} + \theta$ as before. All types below a cutoff θ^P pool and charge a common price, and consumers' perceived quality is the average of the pool, $\frac{\theta^P}{2}$. The pooled price and profit are $P^f(\frac{\theta^P}{2})$ and $\pi^f(\frac{\theta^P}{2})$, because cost does not vary with type.

	$D = 0.3$	$D = 0.5$
All types signal	0.5271	0.5271
Voluntary disclosure (types above θ^V)	0.8170	0.6204
Mandatory disclosure (all types)	0.8563	0.6563
Socially optimal (types above θ^W)	0.8728	0.7217

Table 3: Average total surplus per type, with a uniform prior on $[0, 1]$.

The firm's cutoff. The marginal type θ^{PV} is indifferent between pooling and disclosing, so $D = \pi^f(\theta) - \pi^f(\frac{\theta}{2})$. With $A(\theta) = \frac{3}{2} + \theta$ and $\pi^f = \frac{A^2}{4}$,

$$D = \frac{A(\theta)^2 - A(\theta/2)^2}{4} = \frac{3\theta}{8} + \frac{3\theta^2}{16}. \quad (11)$$

The planner's cutoff. With $k = 0$, a pooled type consumes $q^P = \frac{A(\theta^P/2)}{2}$ at the pooled price and gains the utility of its true quality, so average surplus over the pool equals the surplus at the pool's average quality: the term that is linear in $\theta - \frac{\theta^P}{2}$ integrates to zero over the pool. Total surplus is $W^P(\theta^P) = \frac{3}{8} \left[\theta^P A(\frac{\theta^P}{2})^2 + \int_{\theta^P}^1 A(\theta)^2 d\theta \right] - D(1 - \theta^P)$, and $\frac{dW^P}{d\theta^P} = 0$ gives

$$D = \frac{3}{8} \left[A(\theta)^2 - A(\frac{\theta}{2})^2 - \theta A(\frac{\theta}{2}) \right] = \frac{3\theta^2}{32}. \quad (12)$$

Figure 6 depicts both cutoffs against D . For the same D , (11) exceeds (12) at every $\theta > 0$, so $\theta^{PV} < \theta^{PW}$: too many types disclose, as in the classical result. Solving the two equations gives $\theta^{PV} = \sqrt{1 + \frac{16D}{3}} - 1$ and $\theta^{PW} = \sqrt{\frac{32D}{3}}$. For example, at $D = 0.06$ the firm's cutoff is $\theta^{PV} = 0.149$ and the planner's is $\theta^{PW} = 0.8$. The planner's cutoff reaches one at $D = \frac{3}{32} = 0.094$, beyond which no disclosure is socially optimal at all.

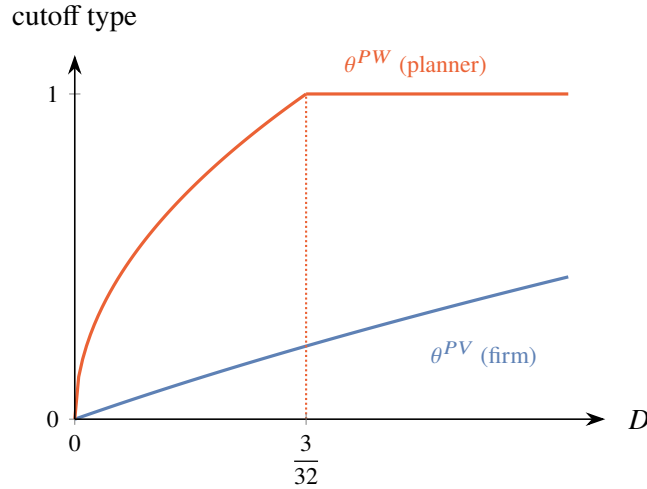


Figure 6: Cutoff type above which types disclose, chosen by the firm (θ^{PV}) and by the planner (θ^{PW}), against the disclosure cost D , with $k = 0$ and θ uniform on $[0, 1]$.

Two comments. First, this reproduces the paper's contrast: with a cost that does not depend on quality, voluntary disclosure is excessive, whereas with $k > 0$ in Step (e) it is insufficient. In the signaling model disclosure lowers a distorted price and raises output, a social gain that the firm does not capture; in the pooling model disclosure by a high type mainly redistributes, and it lowers the pool's average quality, which hurts the types that remain in it. Second, Proposition 4(ii) of the paper adds that the ranking reverses, with insufficient disclosure, when D exceeds $\frac{2(\alpha-\delta)^2}{3\beta}$, which is $\frac{3}{2}$ here. The reversal could not be reproduced: for the planner's problem in the paper's equation (6), with each pooled type treated as the pool's mean, as the paper states, the toy gives $\theta^{PV} < \theta^{PW}$ at every D below $\frac{9}{16}$, the cost at which the firm's cutoff reaches one, and both cutoffs equal one above it. The same holds for any parameters that satisfy Assumption 2, which with $k = 0$ and $\underline{\theta} = 0$ requires $\alpha > \delta$: the firm's condition is $D = \frac{4\delta\theta(\alpha-\delta)+3\delta^2\theta^2}{16\beta}$, the planner's is $D = \frac{3\delta^2\theta^2}{32\beta}$, and the first exceeds the second by $\frac{\delta\theta[8(\alpha-\delta)+3\delta\theta]}{32\beta} > 0$. The threshold in Proposition 4(ii) is the value of D at which the two conditions cross: they are equal at $\theta = \frac{8(\delta-\alpha)}{3\delta}$, where $D = \frac{2(\alpha-\delta)^2}{3\beta}$, but this crossing is positive only if $\alpha < \delta$, a case that Assumption 2 rules out.

8. Step (g): What the Toy Model Keeps and Drops

The toy model keeps the main conclusions of the paper:

1. When higher quality is costlier, the unique separating equilibrium has the lowest type undistorted and every higher type charging more than its full-information price (Step (b), Proposition 1).
2. Signaling profit falls with quality, full-information profit rises, and the gap grows, so only high types disclose (Steps (c) and (d), Proposition 2).
3. Overall profit is U-shaped in quality, and for a moderate disclosure cost the highest type earns the most (Step (d)).
4. Voluntary disclosure is socially insufficient, mandatory disclosure is excessive, and a decentralized subsidy achieves the optimum without changing the firm's profit (Step (e); Section 3 of the paper).
5. When cost does not depend on quality, nondisclosure means pooling and voluntary disclosure is excessive (Step (f)).

The toy fixes a linear-quadratic specification, a uniform prior, and the parameters of Section 1. Relative to Daughety and Reinganum (2008), the firm is a monopolist and price is the only signal; the paper notes that other signals used alongside price, such as advertising, could raise signaling profit and so raise the cutoff θ^V . Disclosure is a perfect, verifiable report bought from a competitive auditor, and a noisy report that keeps the same support would be worth nothing, because the separating price schedule depends only on the support of beliefs. The type space is exogenous: the paper speculates that a firm that chooses its quality by R&D would sample until it draws a low or a high quality, and that its type space could become the union of two intervals. An extension to oligopoly is hard, because each firm's incentive constraint is also a best reply to its rival's price schedule. Each of these features can be relaxed one at a time.

References

Daughety, A. F., and J. F. Reinganum (2008). "Communicating Quality: A Unified Model of Disclosure and Signalling." *RAND Journal of Economics*, 39(4), pp. 973–989.

Signaling a Clean Product with Its Price

A step-by-step exercise in the spirit of Mahenc (2008)

Companion note to Mahenc, P. (2008), “Signaling the Environmental Performance of Polluting Products to Green Consumers,”
International Journal of Industrial Organization 26(1): 59–68. Full references at the end.

Overview. Many products claim to be “green,” but consumers cannot check how much pollution a product generates, and a firm’s claim is cheap to make. Mahenc (2008) asks whether the price can do the job that the claim cannot. His answer extends the classic result that a high price can signal high quality (Milgrom and Roberts, 1986) to a setting where consumers care about the *total* pollution associated with a product, so that demand and emissions are linked. When a cleaner product is more costly to produce, a clean firm signals with a price above its full-information level, and, because it sells less, it also pollutes less than under full information. When a cleaner product is cheaper to produce, as in the Porter hypothesis, the distortions go the other way. This note builds the mechanism with uniform tastes and linear emissions, so that every equilibrium has closed form. Step (a) solves the full-information benchmark; Step (b) states the firm’s incentives when consumers cannot observe cleanness; Step (c) finds the separating equilibrium with two types when cleaner products are costlier; Step (d) reverses the cost ranking; Step (e) solves the continuum-of-types model; Step (f) asks whether a label that reveals cleanness raises welfare; and Step (g) lists what the toy model keeps and drops.

1. The Toy Model

Emissions. A monopolist produces q units of a good. Production generates pollution

$$e = \frac{q}{\theta}, \quad (1)$$

where $\theta > 0$ is the environmental performance of the product: a higher θ means a cleaner product. This is the emission function $e = \frac{t(q)}{\theta}$ of the paper with $t(q) = q$.

Consumers. A unit mass of consumers have basic willingness to pay r , uniformly distributed on $[0, 1]$, and each buys at most one unit. A buyer bears a private damage $v(e) = e$ from the pollution of the product (a health cost, or the guilt from buying a dirty good), so a buyer of type r obtains $r - e - p$. The global damage of pollution falls on buyers and non-buyers alike and therefore does not affect who buys; it enters only the welfare analysis of Step (f).

Demand. The marginal buyer $\tilde{r}$ satisfies $\tilde{r} - e - p = 0$, and the buyers are those with $r \geq \tilde{r}$, so $q = 1 - \tilde{r}$. Substituting (1), $\tilde{r} - \frac{1-\tilde{r}}{\theta} - p = 0$, and solving,

$$D(p, \theta) = \alpha(\theta) (1 - p), \quad \alpha(\theta) \equiv \frac{\theta}{1 + \theta}. \quad (2)$$

Demand is the standard linear demand $1 - p$ scaled down by the factor $\alpha(\theta) < 1$: every extra buyer adds pollution that lowers the willingness to pay of all buyers, and this effect is weaker for a cleaner product. Pollution is then

$$e = \frac{D(p, \theta)}{\theta} = \frac{1 - p}{1 + \theta}, \quad (3)$$

which falls with the price and with cleanness. The inverse demand is $p = 1 - q - \frac{q}{\theta}$, and consumer surplus is $CS = \frac{q^2}{2}$.

Costs. The cost of production is $c(q, \theta) = m(\theta) q$, with a constant marginal cost $m(\theta)$. The paper's main case is that a cleaner product is more costly to produce, $m'(\theta) > 0$, so that $c_{q\theta} > 0$.

Numerical example. A dirty type $\theta_L = 1$ and a clean type $\theta_H = 2$, so $\alpha_L = \frac{1}{2}$ and $\alpha_H = \frac{2}{3}$. In Steps (a) to (c) the marginal costs are $m_L = 0$ and $m_H = \frac{1}{4}$.

Figure 1 depicts demand (2) and pollution (3) for the two types, with the price on the horizontal axis. Both demands vanish at $p = 1$, but the clean type's demand is higher at every lower price, because it scales the standard demand by $\alpha_H = \frac{2}{3}$ instead of $\alpha_L = \frac{1}{2}$. Pollution is lower for the clean type at every price, because each unit sold pollutes less, and it falls with the price for both types because fewer units are sold.

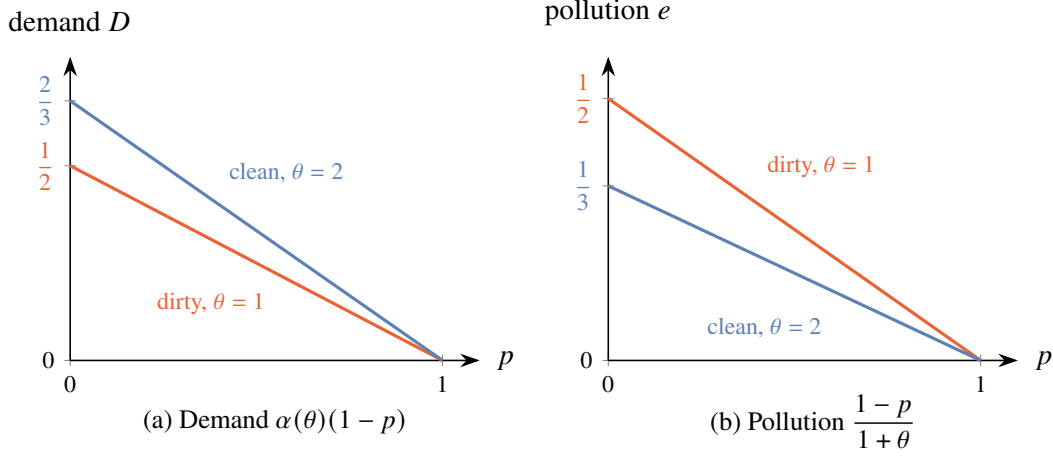


Figure 1: Demand and pollution against the price for the dirty type ($\theta_L = 1$) and the clean type ($\theta_H = 2$).

2. Step (a): Full Information

If consumers observe θ , the firm maximizes $(p - m)\alpha(\theta)(1 - p)$, so

$$p^*(\theta) = \frac{1 + m(\theta)}{2}, \quad q^*(\theta) = \frac{\alpha(\theta)(1 - m(\theta))}{2}, \quad \pi^*(\theta) = \frac{\alpha(\theta)(1 - m(\theta))^2}{4}. \quad (4)$$

The price elasticity of demand is $\eta = \frac{p}{1-p}$, and at p^* the Lerner index satisfies

$$\frac{p^* - m}{p^*} = \frac{1 - m}{1 + m} = \frac{1}{\eta(p^*)}, \quad (5)$$

the standard result that Lemma 1 of the paper records. Cleanness scales demand by the factor $\alpha(\theta)$, which rises with θ , without changing the elasticity at a given price, so it raises profit through the size of the market. In the example, the dirty type charges $p_L^* = \frac{1}{2}$, sells $\frac{1}{4}$, earns $\pi_L^* = \frac{1}{8}$, and generates $e_L^* = \frac{1}{4}$; the clean type charges $p_H^* = \frac{5}{8}$, sells $\frac{1}{4}$, earns $\pi_H^* = \frac{3}{32}$, and generates $e_H^* = \frac{1}{8}$.

Figure 2 depicts the profit of each type against its price. The dirty type's profit $\frac{1}{2} p (1 - p)$ is zero at $p = 0$ and at $p = 1$ and peaks at $p_L^* = \frac{1}{2}$ with height $\frac{1}{8}$. The clean type's profit $\frac{2}{3} (p - \frac{1}{4})(1 - p)$ is zero at its marginal cost $p = \frac{1}{4}$ and at $p = 1$, and peaks at $p_H^* = \frac{5}{8}$ with height $\frac{3}{32}$. The clean type charges more because its marginal cost is higher, and it earns less because its higher cost outweighs its larger market. These two curves, and their crossings with the profit levels in Step (c), are the ingredients of Figure 4.

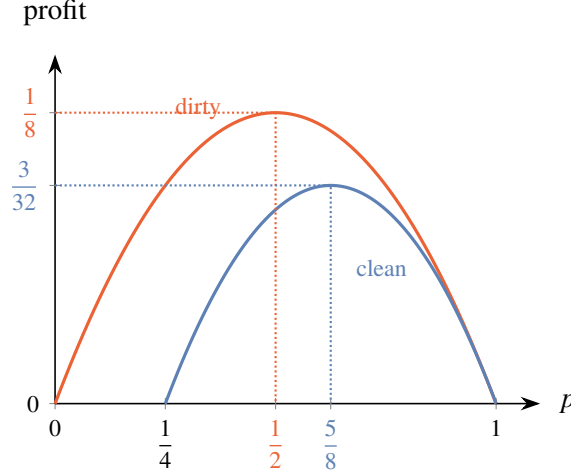


Figure 2: Profit of each type against its price under full information, with $(m_L, m_H) = (0, \frac{1}{4})$.

3. Step (b): Hidden Cleanness and the Firm's Incentives

Now consider that only the firm observes θ . Consumers see the price and form an inference $\hat{\theta}$, and they buy according to $D(p, \hat{\theta})$. A firm of type θ that is perceived as $\hat{\theta}$ earns

$$\pi(p, \theta, \hat{\theta}) = \alpha(\hat{\theta}) (p - m(\theta))(1 - p). \quad (6)$$

We look for separating perfect Bayesian equilibria, in which different types charge different prices and consumers infer the type from the price. Two properties of (6) drive the analysis.

Every type wants to be perceived as clean. Since α is increasing, $\pi_{\hat{\theta}} > 0$ whenever $p > m(\theta)$ (Condition 2 of the paper). The worst inference for any type is therefore the dirtiest one.

Single crossing. The perceived type multiplies profit, so a type's ranking of prices for a given perception depends only on $\ln[(p - m)(1 - p)]$, whose slope in p is

$$\frac{\partial}{\partial p} \ln[(p - m)(1 - p)] = \frac{1}{p - m} - \frac{1}{1 - p}, \quad \frac{\partial}{\partial m} \left[\frac{1}{p - m} - \frac{1}{1 - p} \right] = \frac{1}{(p - m)^2} > 0. \quad (7)$$

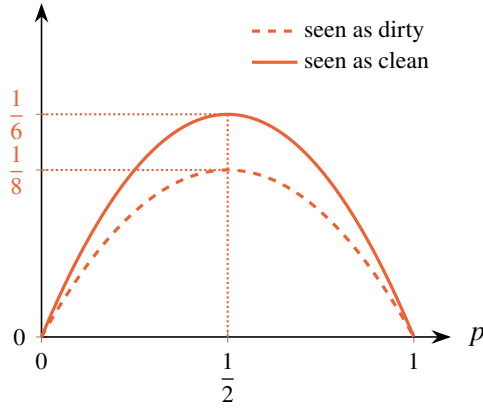
A firm with a higher marginal cost gains more, in proportional terms, from raising its price: a price increase raises its thinner margin by a larger proportion, while the proportional loss of sales, $\frac{1}{1-p}$ per unit of price, is the same for every type. When cleaner products are more costly ($m' > 0$), a clean firm is more willing than a dirty firm to trade a higher price for a better perception. In terms of the paper's Condition 4, the marginal rate of substitution between price and perceived type satisfies

$$\frac{\partial}{\partial \theta} \left[-\frac{\pi_{\hat{\theta}}}{\pi_p} \right] = \frac{c_{q\theta} D_{\hat{\theta}} D}{\pi_p^2}, \quad (8)$$

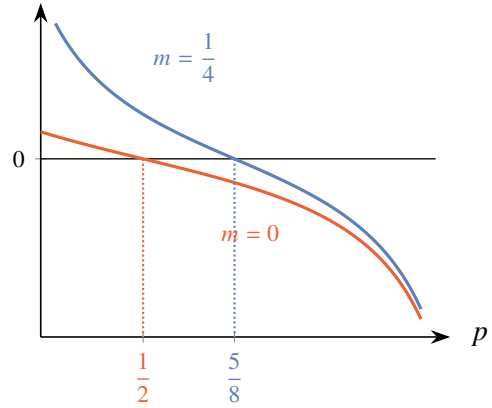
and since $D_{\hat{\theta}} = \frac{1-p}{(1+\theta)^2} > 0$, the sign of the single-crossing condition is the sign of $c_{q\theta}$.

Figure 3 depicts both properties. Panel (a) shows the dirty type's profit when it is perceived as dirty and when it is perceived as clean: the second curve is the first one multiplied by $\frac{\alpha_H}{\alpha_L} = \frac{4}{3}$, so it lies above it at every price and the two peak at the same price. Panel (b) shows the slope (7) of $\ln[(p - m)(1 - p)]$ for $m = 0$ and $m = \frac{1}{4}$. The slope is higher for the costlier type at every price, and it is zero at the price that maximizes $(p - m)(1 - p)$, which is $\frac{1}{2}$ for $m = 0$ and $\frac{5}{8}$ for $m = \frac{1}{4}$. At a price between $\frac{1}{2}$ and $\frac{5}{8}$, for example, the dirty type's profit falls in proportional terms while the clean type's profit is still rising.

profit of the dirty type



(a) Perception raises profit

slope of $\ln[(p-m)(1-p)]$ 

(b) Single crossing

Figure 3: Profit of the dirty type ($m = 0$) when perceived as dirty (dashed, α_L) and as clean (solid, α_H), and the slope of $\ln[(p-m)(1-p)]$ for $m = 0$ and $m = \frac{1}{4}$.

4. Step (c): Two Types when Cleaner Products Are Costlier

Let consumers believe that the firm is clean if it charges p_H and dirty otherwise. In a separating equilibrium the dirty type is recognized, so it charges its full-information price $p_L^* = \frac{1}{2}$ and earns $\frac{1}{8}$: there is no distortion at the bottom (Lemma 2 of the paper). The clean type's price is determined by two incentive constraints. Each one compares a profit that depends on the price, which is a curve in p_H , with a payoff that does not depend on it, which is a horizontal line, and Figure 4 is built from exactly these four objects.

The dirty type must not mimic. Charging p_H would make the dirty type look clean, and it would earn $\alpha_H(p_H - m_L)(1 - p_H) = \frac{2}{3}p_H(1 - p_H)$, a curve, instead of its equilibrium profit $\frac{1}{8}$, a line. Separation requires

$$\frac{2}{3}p_H(1 - p_H) \leq \frac{1}{8}. \quad (9)$$

The clean type must prefer to separate. If the clean type charges any other price, consumers believe it is dirty, and its best such deviation is its full-information price against the dirty type's demand, which earns $\alpha_L \frac{(1-m_H)^2}{4} = \frac{9}{128}$, a line. Charging p_H earns $\alpha_H(p_H - m_H)(1 - p_H) = \frac{2}{3}(p_H - \frac{1}{4})(1 - p_H)$, a curve. Separation requires

$$\frac{2}{3}\left(p_H - \frac{1}{4}\right)(1 - p_H) \geq \frac{9}{128}. \quad (10)$$

Rearranging the conditions. Each condition is a quadratic inequality in p_H . Moving everything to one side of (9), and multiplying (10) by $\frac{3}{2}$ to get $(p_H - \frac{1}{4})(1 - p_H) \geq \frac{27}{256}$ before doing the same,

$$\begin{aligned} (9) \quad &\iff p_H^2 - p_H + \frac{3}{16} \geq 0 \iff \left(p_H - \frac{1}{4}\right)\left(p_H - \frac{3}{4}\right) \geq 0 \\ &\iff p_H \leq \frac{1}{4} \text{ or } p_H \geq \frac{3}{4}, \end{aligned} \quad (11)$$

$$\begin{aligned} (10) \quad &\iff p_H^2 - \frac{5}{4}p_H + \frac{91}{256} \leq 0 \iff \left(p_H - \frac{7}{16}\right)\left(p_H - \frac{13}{16}\right) \leq 0 \\ &\iff \frac{7}{16} \leq p_H \leq \frac{13}{16}. \end{aligned} \quad (12)$$

The four roots are the prices at which a curve crosses its line, and they are the numbers marked on the horizontal axis of Figure 4(a).

Constructing Figure 4(a). Students can reproduce the panel in six steps, using the numbers of Table 1.

1. Draw the price p on the horizontal axis, from 0 to 1, and profit on the vertical axis, from 0 to about 0.2.
2. Plot the profit that the dirty type would earn if it were seen as clean, $\frac{2}{3}p(1-p)$. It is a downward parabola that is zero at $p = 0$ and $p = 1$ and peaks at $p = \frac{1}{2}$ with height $\frac{1}{6}$.
3. Draw the horizontal line at the dirty type's equilibrium profit, $\frac{1}{8}$. The parabola crosses it where $p(1-p) = \frac{3}{16}$, that is, at $p = \frac{1}{4}$ and $p = \frac{3}{4}$, and lies above it in between, where the dirty type would mimic.
4. Plot the profit of the clean type if it is seen as clean, $\frac{2}{3}(p - \frac{1}{4})(1-p)$, for $p \geq \frac{1}{4}$. It is zero at $p = \frac{1}{4}$ and $p = 1$ and peaks at $p = \frac{5}{8}$ with height $\frac{3}{32}$, its full-information profit.
5. Draw the horizontal line at the clean type's profit if it is seen as dirty, $\frac{9}{128}$. The curve crosses it where $(p - \frac{1}{4})(1-p) = \frac{27}{256}$, at $p = \frac{7}{16}$ and $p = \frac{13}{16}$, and lies above it in between, where the clean type prefers to separate.
6. Shade the prices at which the dirty type's curve is below its line and the clean type's curve is above its line, which gives $[\frac{3}{4}, \frac{13}{16}]$. The prices $p \leq \frac{1}{4}$ also satisfy (11), but there the clean type's curve is below its line.

Object in Figure 4(a)	Formula	Zeros or crossings	Peak
Dirty type, seen as clean (red curve)	$\frac{2}{3}p(1-p)$	0 and 1	$\frac{1}{6}$ at $p = \frac{1}{2}$
Dirty type's equilibrium profit (red line)	$\frac{1}{8}$	$\frac{1}{4}$ and $\frac{3}{4}$	—
Clean type, seen as clean (blue curve)	$\frac{2}{3}(p - \frac{1}{4})(1-p)$	$\frac{1}{4}$ and 1	$\frac{3}{32}$ at $p = \frac{5}{8}$
Clean type, seen as dirty (blue line)	$\frac{9}{128}$	$\frac{7}{16}$ and $\frac{13}{16}$	—

Table 1: Ingredients of Figure 4(a). The crossings are those of each curve with the line of the same color.

Combining (9) and (10), the separating prices for the clean type are $p_H \in [\frac{3}{4}, \frac{13}{16}]$. The low prices $p_H \leq \frac{1}{4}$ that would also deter the dirty type are useless to the clean type, since they leave it no margin over its marginal cost $\frac{1}{4}$. Among the separating equilibria, the Intuitive Criterion of Cho and Kreps (1987) selects the least costly one, $p_H = \frac{3}{4}$: if the clean type charged $p_H > \frac{3}{4}$, a deviation to a price in $(\frac{3}{4}, p_H)$ would not pay for the dirty type even if consumers believed it was clean, but it would pay for the clean type, so consumers should attribute the deviation to the clean type.

Result 1 (Proposition 1 of the paper, two types). When cleaner products are more costly to produce:

- the clean type signals with a price above its full-information level, $\frac{3}{4} > \frac{5}{8}$;
- the dirty type charges its full-information price, $\frac{1}{2}$;
- the clean type sells less, $\frac{1}{6}$ instead of $\frac{1}{4}$, and pollutes less, $\frac{1}{12}$ instead of $\frac{1}{8}$.

The clean type pays for its signal: its profit falls from $\frac{3}{32}$ to $\frac{1}{12}$, a loss of $\frac{1}{96}$. Pollution falls by one third, because the signal works by restricting output and output is what pollutes.

In Figure 4(a), the dirty type mimics at the prices where the red curve lies above the dotted red line, the clean type separates at the prices where the blue curve lies above the dotted blue line, and the separating prices, shaded, satisfy both conditions.

5. Step (d): When Cleaner Products Are Cheaper

Porter and van der Linde (1995) argue that cleaner products can be cheaper to produce. Swap the marginal costs: $m_L = \frac{1}{4}$ and $m_H = 0$, so that $c_{q\theta} < 0$. Under full information the dirty type charges $\frac{5}{8}$ and earns $\frac{9}{128}$, and the clean type charges $\frac{1}{2}$, sells $\frac{1}{3}$, earns $\frac{1}{6}$, and generates $e_H^* = \frac{1}{6}$. The constraint that the dirty type must not mimic compares its curve $\frac{2}{3}(p - \frac{1}{4})(1 - p)$ with its full-information profit $\frac{9}{128}$, and the constraint that the clean type must prefer to separate compares its curve $\frac{2}{3}p(1 - p)$ with the profit $\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$ that it earns if it is seen as dirty. They become

$$\frac{2}{3}\left(p_H - \frac{1}{4}\right)(1 - p_H) \leq \frac{9}{128} \iff p_H \leq \frac{7}{16} \text{ or } p_H \geq \frac{13}{16}, \quad (13)$$

$$\frac{2}{3}p_H(1 - p_H) \geq \frac{1}{8} \iff \frac{1}{4} \leq p_H \leq \frac{3}{4}, \quad (14)$$

so the separating prices are $p_H \in [\frac{1}{4}, \frac{7}{16}]$. Panel (b) of Figure 4 is built with the same six steps, with the roles of the two types exchanged: the red curve is the same parabola as the blue curve of panel (a) and its line is $\frac{9}{128}$, and the blue curve is the same parabola as the red curve of panel (a), with the line $\frac{1}{8}$, so the crossings $\frac{1}{4}, \frac{3}{4}$ and $\frac{7}{16}, \frac{13}{16}$ appear again, now for the other type. The least costly separating price for the clean type is now the highest one in the interval, $p_H = \frac{7}{16}$, because its profit $\frac{2}{3}p(1 - p)$ rises up to $p = \frac{1}{2}$.

Result 2 (cleaner is cheaper). When cleaner products are cheaper to produce:

- the clean type signals with a price below its full-information level, $\frac{7}{16} < \frac{1}{2}$;
- the dirty type charges its full-information price, $\frac{5}{8}$;
- the clean type sells more, $\frac{3}{8}$ instead of $\frac{1}{3}$, and pollutes more, $\frac{3}{16}$ instead of $\frac{1}{6}$.

Intuitively, the clean firm now has the wider margin, so a price cut lowers its margin by a smaller proportion than it lowers the dirty firm's, while both gain the same proportion of sales. A price cut is therefore the signal that the dirty firm finds too costly to copy, and the direction of the signal follows the sign of $c_{q\theta}$, as (8) shows.

Table 2 reports the full-information and signaling outcomes in both cases.

Case	Type	p^*	p^s	q^*	q^s	e^*	e^s	π^*	π^s
Cleaner is costlier	Dirty	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
$(m_L, m_H) = (0, \frac{1}{4})$	Clean	$\frac{5}{8}$	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{3}{32}$	$\frac{1}{12}$
Cleaner is cheaper	Dirty	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{9}{128}$	$\frac{9}{128}$
$(m_L, m_H) = (\frac{1}{4}, 0)$	Clean	$\frac{1}{2}$	$\frac{7}{16}$	$\frac{1}{3}$	$\frac{3}{8}$	$\frac{1}{6}$	$\frac{3}{16}$	$\frac{1}{6}$	$\frac{21}{128}$

Table 2: Price, output, pollution, and profit under full information (*) and in the least costly separating equilibrium (s), with $\theta_L = 1$ and $\theta_H = 2$.

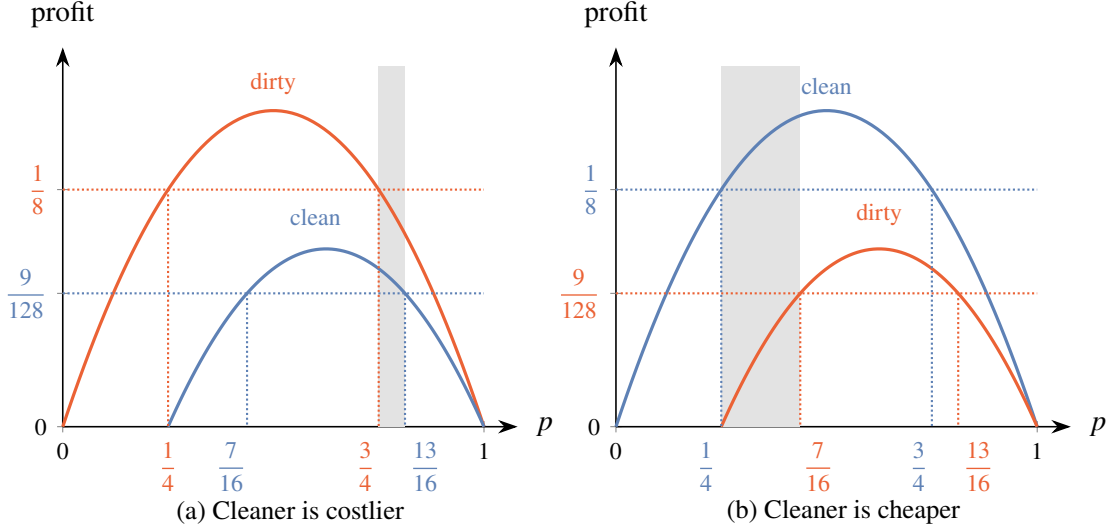


Figure 4: Profit from charging p and being perceived as clean, for each type, and each type's payoff outside the clean type's price (dotted). Shaded: separating prices for the clean type. Types $\theta_L = 1$ and $\theta_H = 2$; marginal costs $(m_L, m_H) = (0, \frac{1}{4})$ in panel (a) and $(\frac{1}{4}, 0)$ in panel (b).

6. Step (e): A Continuum of Types

The paper works with a continuum of types $\theta \in [\underline{\theta}, \bar{\theta}]$ and the method of Mailath (1987). Let $\theta \in [1, 2]$ with $m(\theta) = \frac{\theta-1}{4}$, which reproduces the two types of Step (c) at the end points. A separating equilibrium is a strictly monotone price function $\varphi(\theta)$, and a consumer who sees p infers $\hat{\theta} = \varphi^{-1}(p)$.

Local incentive compatibility. Type θ chooses which type $\hat{\theta}$ to imitate, earning $\Pi(\theta, \hat{\theta}) \equiv \pi(\varphi(\hat{\theta}), \theta, \hat{\theta})$. Truth-telling requires $\Pi_{\hat{\theta}}(\theta, \theta) = 0$, that is, $\pi_p \varphi'(\theta) + \pi_{\hat{\theta}} = 0$, which gives the differential equation

$$\varphi'(\theta) = -\frac{\pi_{\hat{\theta}}}{\pi_p} = \frac{(\varphi - m(\theta))(1 - \varphi)}{\theta(1 + \theta)(2\varphi - 1 - m(\theta))}, \quad (15)$$

where we used $\frac{\alpha'(\theta)}{\alpha(\theta)} = \frac{1}{\theta(1+\theta)}$ and $\pi_p = \alpha(\theta)(1 + m(\theta) - 2\varphi)$. The initial condition is $\varphi(1) = p^*(1) = \frac{1}{2}$: the dirtiest type is recognized as the dirtiest, so it has no reason to distort (Lemma 2 of the paper).

The price exceeds the full-information price. Along a separating equilibrium $\varphi' > 0$ and $\pi_{\hat{\theta}} > 0$, so (15) requires $\pi_p < 0$, that is, $\varphi(\theta) > \frac{1+m(\theta)}{2} = p^*(\theta)$ for every $\theta > 1$. Each type prices above the level at which its profit stops rising, which is what makes the price costly enough to deter the types just below it.

Result 3 (Proposition 1 of the paper). With a continuum of types and $c_{q\theta} > 0$:

- the separating price $\varphi(\theta)$ is strictly increasing;
- it exceeds $p^*(\theta)$ for every type except the dirtiest;
- every type except the dirtiest pollutes less than under full information.

Equation (15) has no closed-form solution, but it is easy to integrate numerically (as a function $\theta(\varphi)$, whose slope is zero at the initial point). Figure 5 depicts the solution. In panel (a), the signaling price rises from 0.500 at $\theta = 1$ to 0.714 at $\theta = 1.5$ and 0.765 at $\theta = 2$, while the full-information price $p^*(\theta) = \frac{\theta+3}{8}$ rises only from 0.500 to 0.625. In panel (b), pollution under signaling falls from 0.250 to 0.078, against 0.125

under full information for the cleanest type. The price function has an infinite slope at $\theta = 1$: the types just above the dirtiest one must move their prices away from $\frac{1}{2}$ quickly, because near p^* a small price change costs almost nothing.

The cleanest type charges 0.765 in the continuum, more than the $\frac{3}{4}$ it charges with two types. With two types, the clean firm only needs to deter the dirtiest firm, whose profit is undistorted. With a continuum, each type must deter the types just below it, which already charge distorted prices, and the distortions accumulate along the chain. A discrete chain of types confirms this: the top price is 0.750 with two types, 0.762 with 11 equally spaced types, and approaches 0.765 as the grid becomes finer.

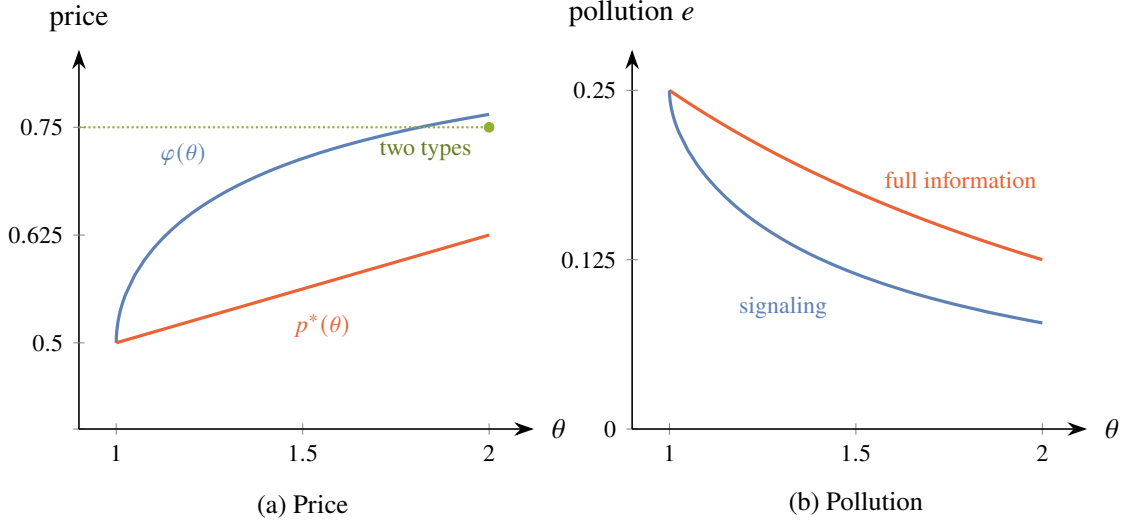


Figure 5: Separating price and pollution with a continuum of types $\theta \in [1, 2]$ and $m(\theta) = \frac{\theta-1}{4}$, against full information.

7. Step (f): Should a Label Reveal Cleanness?

The paper notes that a labeling program that resolves consumers' uncertainty removes one distortion but can worsen another: by ending the need to signal, it can raise pollution. The toy model makes the trade-off explicit. Let the global damage of pollution be δe , borne by everyone, and measure welfare as $CS + \pi - \delta e$, with $CS = \frac{q^2}{2}$. A label that reveals θ moves the clean type from its signaling outcome to its full-information outcome and leaves the dirty type unchanged.

Cleaner is costlier. With the label the clean type sells $\frac{1}{4}$ instead of $\frac{1}{6}$. Consumer surplus rises from $\frac{1}{72}$ to $\frac{1}{32}$ and profit from $\frac{1}{12}$ to $\frac{3}{32}$, a gain of $\frac{1}{36}$ in $CS + \pi$, while pollution rises by $\frac{1}{8} - \frac{1}{12} = \frac{1}{24}$. The label raises welfare if and only if

$$\frac{1}{36} > \frac{\delta}{24} \iff \delta < \frac{2}{3}. \quad (16)$$

Cleaner is cheaper. With the label the clean type raises its price from $\frac{7}{16}$ to $\frac{1}{2}$ and sells $\frac{1}{3}$ instead of $\frac{3}{8}$. $CS + \pi$ falls by $\frac{7}{576}$, because the signaling price cut had moved the monopoly output closer to the efficient level, while pollution falls by $\frac{3}{16} - \frac{1}{6} = \frac{1}{48}$. The label raises welfare if and only if

$$\frac{\delta}{48} > \frac{7}{576} \iff \delta > \frac{7}{12}. \quad (17)$$

Result 4 (labels). A label that reveals cleanness raises welfare:

- when cleaner products are costlier, if and only if the pollution damage is low, $\delta < \frac{2}{3}$;
- when cleaner products are cheaper, if and only if the pollution damage is high, $\delta > \frac{7}{12}$.

Signaling distorts the clean firm's output, and the direction of the distortion determines whether removing it helps the environment or the market. This illustrates the paper's point that a single instrument can mitigate one distortion and worsen another; with three distortions (the externality, market power, and asymmetric information), the paper calls for three instruments: a Pigouvian tax on emissions, a tax or subsidy that offsets the price distortion, and a lump-sum transfer.

8. Step (g): What the Toy Model Keeps and Drops

The toy model keeps the paper's main conclusions:

1. Every type wants to look clean (Condition 2), and the dirtiest type charges its full-information price (Lemma 2, Step c).
2. The direction of the signal follows the sign of $c_{q\theta}$: a price above the full-information level when cleaner products are costlier, below it when they are cheaper (Proposition 1 and its discussion, Steps c and d).
3. Signaling lowers pollution when cleaner products are costlier, from $\frac{1}{8}$ to $\frac{1}{12}$ for the clean type with two types, and for every type above the dirtiest with a continuum (Steps c and e).
4. Full disclosure can raise pollution, so a label alone need not raise welfare (Step f).

The linear-uniform specification adds two features to the general analysis. First, cleanness enters demand as the multiplicative factor $\alpha(\theta) = \frac{\theta}{1+\theta}$, which makes single crossing a one-line calculation (7). Second, the comparison between two types and a continuum shows that the distortion of the cleanest type grows with the number of types it must separate from, from 0.750 to 0.765.

Relative to Mahenc (2008), the toy model fixes $v(e) = e$, $t(q) = q$, a uniform distribution of basic willingness to pay on $[0, 1]$, and a constant marginal cost; the paper allows a convex private damage, a convex emission function, and a general cost $c(q, \theta)$. As in the paper, the firm is a monopolist, cleanness is exogenous, the price is the only signal, and the analysis focuses on separating equilibria. The paper's conclusion suggests advertising and green investment as additional signals; each of these features, and competition between a clean and a dirty firm, can be added one at a time.

References

- Cho, I.-K., and D. M. Kreps (1987). "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics*, 102(2), pp. 179–221.
- Mahenc, P. (2008). "Signaling the Environmental Performance of Polluting Products to Green Consumers." *International Journal of Industrial Organization*, 26(1), pp. 59–68.

Mailath, G. J. (1987). "Incentive Compatibility in Signaling Games with a Continuum of Types." *Econometrica*, 55(6), pp. 1349–1365.

Milgrom, P., and J. Roberts (1986). "Price and Advertising Signals of Product Quality." *Journal of Political Economy*, 94(4), pp. 796–821.

Porter, M. E., and C. van der Linde (1995). "Toward a New Conception of the Environment-Competitiveness Relationship." *Journal of Economic Perspectives*, 9(4), pp. 97–118.

Penetration or Skimming Pricing for Credence Products?

A step-by-step exercise in the spirit of Mahenc (2025)

Companion note to Mahenc, P. (2025), “Penetration or Skimming Pricing for Credence Products?,” *Journal of Economic Dynamics and Control* 180: 105196. Full reference at the end.

Overview. A firm that adds a *credence attribute* to its product—say, sustainable sourcing or a lower carbon footprint—cannot prove the claim at the point of sale. Consumers learn the truth only after a *verification phase* in which certifiers, NGOs, or regulators do their slow work. Until then, the price is the only signal. Mahenc (2025) asks: what price path credibly reveals that a new entrant is socially responsible? His answer depends on competition. Facing conventional incumbents (an *occupied* market), the responsible entrant signals with *penetration pricing*—a low introductory price that rises later. In an *untapped* market, it signals with *skimming pricing*—a high introductory price that falls later. In both cases the firm conspicuously wastes money to prove its type, an instance of Veblen’s law. This note builds the mechanism in the simplest possible setting: a two-period game with the verification lag collapsed to a single period. Seven steps follow: (a) the model, (b) the full-information benchmark, (c) the signaling problem, (d) the occupied market and penetration pricing, (e) the untapped market and skimming pricing, (f) comparative statics and the verification lag, and (g) a numerical example.

1. The Model

Consumers. A unit mass of consumers each buy at most one unit. A unit with *social performance* $e \in [0, 1]$ generates $1 - e$ units of a harmful by-product; the conventional product has $e = 0$. Consumers differ in their *aversion* x to the harm, with x distributed uniformly on $[0, 1]$. A consumer of aversion x who buys a unit of performance e at price p gets

$$u = v - (1 - e)x - p, \quad (1)$$

where v is the gross value of the product category. A higher e lowers the harm term $(1 - e)x$, so consumers pay more for a greener product.

The entrant and costs. An entrant introduces a product of performance e at a constant unit cost c_e , against a conventional alternative of unit cost c_0 . Write the *cost premium* of responsibility as

$$\varphi \equiv c_e - c_0 > 0. \quad (2)$$

A fixed setup cost $F > 0$ is sunk once the firm enters; because it is paid whether or not the firm signals, it cancels from every comparison below and we drop it.

Two scenarios. The entrant faces either

- an *untapped market*: no competitor, so the entrant is a monopolist; or
- an *occupied market*: perfectly competitive incumbents already sell the conventional product at price c_0 .

The two settings differ only in the outside option consumers have, and that difference is what flips the sign of the price signal.

Normalization. To keep the algebra light we set

$$v = 1, \quad c_0 = 0, \quad (3)$$

so the conventional unit cost (and competitive price) is 0 and the responsible unit cost is $c_e = \varphi$. Throughout we assume $\varphi < e$ (the responsible product is socially desirable) and $v - c_0 \geq 1$ (the occupied market is fully covered); both hold under (3) for the parameters we use.

Timing and the verification lag. Social responsibility is the entrant's private information. In Mahenc's full model it takes T trading periods for consumers to learn the truth; here we take the simplest case $T = 1$. The game then has two stages:

Stage 1. Verification period ($t = 1$). The entrant posts a price; consumers buy and try to infer the type from the price.

Stage 2. Full information ($t \geq 2$). The type becomes public, and the entrant charges the full-information price forever after.

Players discount the future at rate $\delta \in (0, 1)$, so the value of the full-information continuation, seen from $t = 1$, is $\frac{\delta}{1 - \delta}$ times the per-period full-information profit. Section 6 restores general T in one line.

2. Step (a): The Full-Information Benchmark

Suppose the type is known. We record the responsible entrant's optimal price and profit in each scenario; these are the post-verification payoffs the signal is trying to protect.

Untapped market. The indifferent consumer solves $0 = v - (1 - e)x - p$, giving demand $D_u(p) = \frac{v - p}{1 - e} = \frac{1 - p}{1 - e}$. The responsible entrant maximizes $(p - \varphi)D_u(p)$, yielding

$$\hat{p}_r^u = \frac{v + c_e}{2} = \frac{1 + \varphi}{2}, \quad \hat{\pi}_r^u = \frac{(v - c_e)^2}{4(1 - e)} = \frac{(1 - \varphi)^2}{4(1 - e)}. \quad (4)$$

A conventional monopolist would instead charge $\hat{p}_c^u = \frac{v + c_0}{2} = \frac{1}{2}$ and earn $\hat{\pi}_c^u = \frac{(v - c_0)^2}{4} = \frac{1}{4}$.

Occupied market. A consumer is indifferent between the conventional product at c_0 and the new product at p when $v - x - c_0 = v - (1 - e)x - p$, i.e. at $x = \frac{p - c_0}{e} = \frac{p}{e}$. Consumers above this cutoff buy the responsible product, giving demand $D_o(p) = 1 - \frac{p - c_0}{e} = 1 - \frac{p}{e}$. Maximizing $(p - \varphi)D_o(p)$,

$$\hat{p}_r^o = \frac{e + c_0 + c_e}{2} = \frac{e + \varphi}{2}, \quad \hat{\pi}_r^o = \frac{(e - \varphi)^2}{4e}. \quad (5)$$

A conventional entrant here is forced by competition to price at $c_0 = 0$ and earns a *zero* margin: $\hat{\pi}_c^o = 0$. This zero-profit outside option is the engine of the whole occupied-market result.

3. Step (b): The Signaling Problem

Now let the type be private in period 1. The entrant is responsible (r) or conventional (c); consumers hold a prior μ_0 that it is responsible and update to a posterior μ after seeing the price. Consumers who believe the product is responsible with probability μ treat its perceived performance as μe , so demand under belief μ is

$$D(p, \mu) = \begin{cases} 1 - \frac{p - c_0}{\mu e} & \text{occupied market,} \\ \frac{v - p}{1 - \mu e} & \text{untapped market.} \end{cases} \quad (6)$$

For any price, demand rises with μ : being *perceived* as responsible is always worth more, whatever the truth. That is exactly why a conventional entrant is tempted to mimic, and why the responsible entrant needs a signal a mimic would not copy.

Separating logic. We look for a *separating* equilibrium in which the responsible type charges a signaling price $\underline{p}$ in period 1, is believed ($\mu = 1$), and then earns the full-information profit forever; the conventional type charges its own full-information price and is believed conventional ($\mu = 0$). Two conditions must hold.

- **Incentive compatibility (IC): the conventional type must not want to mimic $\underline{p}$.** Its gain from being taken as responsible for one period must not exceed what it earns honestly. IC pins down the *direction* of the price distortion.
- **Individual rationality (IR): the responsible type must prefer to signal** rather than accept being mistaken for conventional. IR pins down *when* signaling is worthwhile.

Among all separating prices, the cheapest one—the least distorted, hence most profitable for the responsible type—is the *Riley outcome*, the unique signal that survives the intuitive criterion (Cho and Kreps, 1987). We take that price in each scenario.

4. Step (c): Occupied Market, Penetration Pricing

IC forces a low price. Suppose the responsible type is expected to charge $\underline{p}$. If the conventional type mimics it and is believed responsible, it earns the flow profit

$$\pi_c(\underline{p}, 1) = (\underline{p} - c_0) D_o(\underline{p}, 1) = \underline{p} \left(1 - \frac{\underline{p}}{e}\right),$$

which is strictly positive for any $\underline{p} \in (0, e)$ and peaks at $\underline{p} = \frac{e}{2}$. But honest behavior earns the conventional type only $\hat{\pi}_c^o = 0$. So it would gladly mimic *any* interior price: high prices cannot separate. To deter mimicry the responsible type must drive the price down to where the conventional type gains nothing,

$$\underline{p} \left(1 - \frac{\underline{p}}{e}\right) \leq 0 \iff \underline{p} \leq c_0 = 0. \quad (7)$$

(A distortion the other way, $\underline{p} \geq e$, would choke demand to zero, so no one could learn.) The cheapest such price—the highest one that still separates—is $\underline{p} = c_0$: the responsible entrant matches the competitive price.

Penetration. In the occupied market the responsible entrant signals by charging the competitive price $\underline{p} = c_0$, then raises the price to $\hat{p}_r^o$ once its type is public. The introductory price is below the full-information price: a penetration path.

The signal is a conspicuous loss. Selling at $\underline{p} = c_0 = 0$ while bearing unit cost $c_e = \varphi$, the responsible type serves the whole market ($D_o(0, 1) = 1$) at a per-period loss

$$\pi_r(c_0, 1) = (c_0 - c_e) \cdot 1 = -\varphi. \quad (8)$$

This is Veblen’s “conspicuous waste”: the responsible type burns φ per period to prove itself. A conventional firm, already earning zero, would never add a loss on top, so the loss is exactly what makes the signal credible.

IR: when is the loss worth bearing? The responsible type accepts the period-1 loss φ only if the discounted full-information profit it unlocks covers it:

$$\frac{\delta}{1 - \delta} \hat{\pi}_r^o \geq \varphi \iff \frac{\delta}{1 - \delta} \cdot \frac{(e - \varphi)^2}{4e} \geq \varphi. \quad (9)$$

Solving for δ gives a critical discount factor

$$\tilde{\delta} = \frac{4e\varphi}{(e + \varphi)^2} = \frac{4r}{(1 + r)^2}, \quad r \equiv \frac{\varphi}{e} \in (0, 1), \quad (10)$$

where the second form uses the premium-to-performance ratio $r = \varphi/e$: dividing numerator and denominator of the first form by e^2 leaves $\tilde{\delta}$ a function of r alone. The map $r \mapsto \frac{4r}{(1 + r)^2}$ rises on $(0, 1)$ and reaches 1 at $r = 1$, that is, at $\varphi = e$; there $\tilde{\delta} = 1$, so no $\delta < 1$ can separate. Desirability ($\varphi < e$, i.e. $r < 1$) keeps us strictly below that boundary.

Result 1 (occupied market). Penetration pricing separates the types if and only if $\delta \geq \tilde{\delta}$. Signaling is worthwhile only for a patient entrant, because it must survive a real, out-of-pocket loss during the verification phase before collecting the reward.

5. Step (d): Untapped Market, Skimming Pricing

IC forces a high price. Now there is no competitor. A conventional type that mimics $\underline{p}$ and is believed responsible earns

$$\pi_c(\underline{p}, 1) = (\underline{p} - c_0) D_u(\underline{p}, 1) = \underline{p} \frac{1 - \underline{p}}{1 - e},$$

while honestly it earns the conventional monopoly profit $\hat{\pi}_c^u = \frac{1}{4}$. Mimicry is deterred when

$$\underline{p} \frac{1 - \underline{p}}{1 - e} \leq \frac{1}{4} \iff \underline{p} \leq \frac{1 - \sqrt{e}}{2} \text{ or } \underline{p} \geq \frac{1 + \sqrt{e}}{2}. \quad (11)$$

The responsible type's profit is single-peaked at its full-information price $\hat{p}_r^u = \frac{1 + \varphi}{2}$, which lies inside the forbidden interval $\left(\frac{1 - \sqrt{e}}{2}, \frac{1 + \sqrt{e}}{2}\right)$ precisely when $\varphi < \sqrt{e}$. That inequality is not a separate assumption: it follows from desirability. Because $e \in (0, 1)$ we have $\sqrt{e} > e$, so

$$\varphi < e < \sqrt{e}, \quad (12)$$

and $\varphi < \sqrt{e}$ holds automatically. The optimum therefore always sits inside the interval, and the cheapest separating price is its *upper* edge,

$$\underline{p} = \frac{v + c_0}{2} + \frac{(v - c_0)\sqrt{e}}{2} = \frac{1 + \sqrt{e}}{2} > \hat{p}_r^u. \quad (13)$$

Skimming. In the untapped market the responsible entrant signals by charging the high price $\underline{p} = \frac{1 + \sqrt{e}}{2}$, then lowers it to $\hat{p}_r^u$ once its type is public. The introductory price is above the full-information price: a skimming path.

The signal is forgone sales, not a loss. At the skimming price the responsible type still earns a *positive* margin; it simply serves fewer consumers than at $\hat{p}_r^u$, skimming the cream of the market most willing to pay for responsibility. The high price is one a conventional type cannot afford to match: its product is worth less to consumers, so the same high price would cost it too many sales relative to its comfortable monopoly profit. Again the signal works by wasting something the mimic values more—here, sales volume.

IR holds for all δ . Because signaling here means a *reduced* profit rather than an outright loss, the responsible type is willing to signal regardless of how patient it is.

Result 2 (untapped market). Skimming pricing separates the types for every δ . Unlike penetration, it needs no patience threshold: both types have a profitable future, so the only question is the direction of the period-1 distortion, and IC sends it upward.

6. Step (e): Comparative Statics and the Verification Lag

The occupied-market threshold (10) carries the paper's comparative statics. Differentiating,

$$\frac{\partial \tilde{\delta}}{\partial e} = \frac{4\varphi(\varphi - e)}{(e + \varphi)^3} < 0, \quad \frac{\partial \tilde{\delta}}{\partial \varphi} = \frac{4e(e - \varphi)}{(e + \varphi)^3} > 0 \quad (\text{since } \varphi < e). \quad (14)$$

Result 3 (comparative statics). Signaling is easier (a lower $\tilde{\delta}$):

- when the product is more differentiated (higher e);
- when the cost premium of responsibility is smaller (lower φ).

A greener product commands a larger future profit worth protecting, and a cheaper premium makes the period-1 loss smaller; both lower the patience requirement. The ratio form (10) shows the same thing directly: $\tilde{\delta}$ rises in $r = \varphi/e$, and higher e or lower φ both lower r .

The verification lag. With a general verification length T , the responsible type repeats the penetration price for T periods, so the separation condition (9) becomes $\frac{\delta^T}{1 - \delta} \hat{\pi}_r^o \geq \varphi \frac{1 - \delta^T}{1 - \delta}$, which reduces to

$$\delta \geq \tilde{\delta}^{1/T}. \quad (15)$$

The threshold $\tilde{\delta}^{1/T}$ rises toward 1 as $T \rightarrow \infty$: a longer verification lag makes penetration harder to sustain and demands a more patient entrant.

Result 4 (greenwashing).

- In the occupied market, the entrant cannot separate when $\delta < \tilde{\delta}^{1/T}$, and pooling prices (greenwashing) take over the early periods.
- In the untapped market, the entrant separates for any T and any δ .

Anything that speeds verification (faster certification, active NGOs) lowers the bar for honest signaling; anything that slows it (fake certification, information lobbying) protects greenwashing.

The two scenarios are summarized in Table 1, and the two price paths in Figure 1.

7. Step (f): A Numerical Example

Take

$$v = 1, \quad c_0 = 0, \quad e = \frac{1}{4}, \quad \varphi = \frac{1}{8},$$

so the responsible unit cost is $c_e = \varphi = \frac{1}{8}$. Both scenarios are socially desirable, $\varphi < e$, since $\frac{1}{8} < \frac{1}{4}$; and by (12) this already delivers the upward skimming distortion, as $\varphi < e < \sqrt{e}$ reads $\frac{1}{8} < \frac{1}{4} < \frac{1}{2}$. The premium-to-performance ratio is $r = \frac{\varphi}{e} = \frac{1}{2}$.

	Occupied market (competition)	Untapped market (monopoly)
Signaling strategy	Penetration (low price)	Skimming (high price)
Signaling price $\underline{p}$	c_0	$\frac{v + c_0}{2} + \frac{(v - c_0)\sqrt{e}}{2}$
Bias vs. full information	Downward	Upward
Period-1 sacrifice	A loss, $-\varphi$	Forgone sales (profit still > 0)
Separation requires	$\delta \geq \tilde{\delta} = \frac{4e\varphi}{(e + \varphi)^2}$	every δ
Real-world echoes	Costco/Kroger organic; BYD EVs; free samples	Tesla Model S; luxury eco-brands

Table 1: The two scenarios of Mahenc (2025). Competition flips the sign of the price signal: a responsible entrant penetrates a market held by conventional incumbents but skims an untapped one.

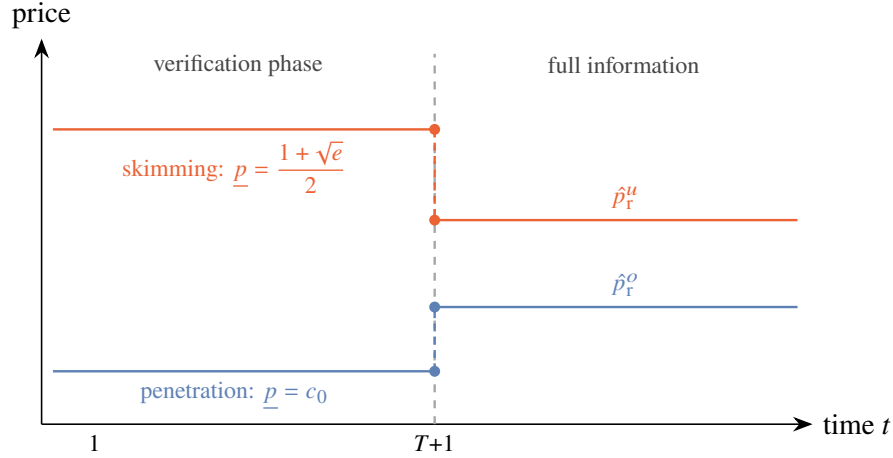


Figure 1: The two signaling price paths. In the occupied market the responsible entrant *penetrates*: it holds the low competitive price c_0 through the verification phase, then raises the price to $\hat{p}_r^o$. In the untapped market it *skims*: it holds a high price $\underline{p} = \frac{1+\sqrt{e}}{2}$, then lowers it to $\hat{p}_r^u$. Both distortions shrink to zero once the type is public.

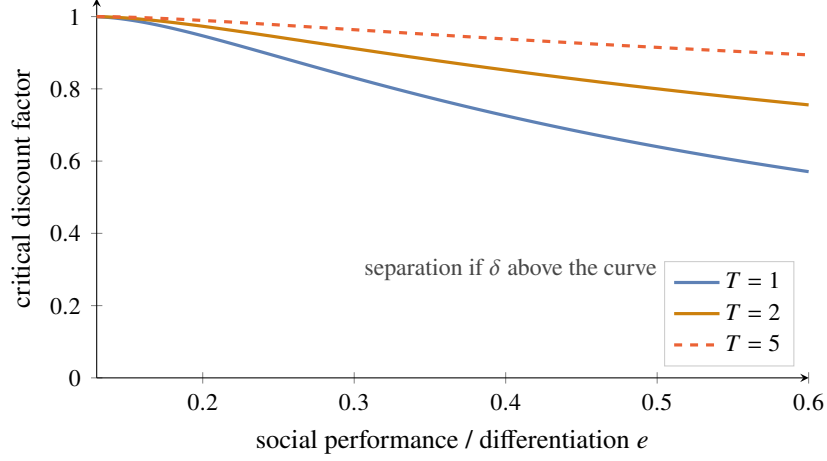


Figure 2: The occupied-market threshold $\tilde{\delta}^{1/T} = [4e\varphi/(e + \varphi)^2]^{1/T}$ against differentiation e , at $\varphi = \frac{1}{8}$, for three verification lengths. It falls with e (more differentiation makes signaling easier) and rises with T (a longer verification lag makes it harder), approaching 1 as $T \rightarrow \infty$. Penetration separates only for discount factors above the relevant curve; the untapped market needs no such condition.

Occupied market (penetration). From (5),

$$\hat{p}_r^o = \frac{e + \varphi}{2} = \frac{3}{16}, \quad \hat{\pi}_r^o = \frac{(e - \varphi)^2}{4e} = \frac{1}{64}.$$

The entrant signals by charging the competitive price $\underline{p} = c_0 = 0$ during verification, taking a loss of $\varphi = \frac{1}{8}$ per period, then raises the price to $\frac{3}{16}$. The critical discount factor, in ratio form with $r = \frac{1}{2}$, is

$$\tilde{\delta} = \frac{4r}{(1+r)^2} = \frac{4 \cdot \frac{1}{2}}{\left(\frac{3}{2}\right)^2} = \frac{2}{\frac{9}{4}} = \frac{8}{9}.$$

So penetration separates the types only if $\delta \geq \frac{8}{9}$: the entrant must be quite patient.

Untapped market (skimming). From (4) and (13),

$$\hat{p}_r^u = \frac{1 + \varphi}{2} = \frac{9}{16}, \quad \underline{p} = \frac{1 + \sqrt{e}}{2} = \frac{1 + \frac{1}{2}}{2} = \frac{3}{4}.$$

The entrant signals by charging the high price $\frac{3}{4}$ during verification, then lowers it to $\frac{9}{16}$. Since $\frac{3}{4} > \frac{9}{16}$, the introductory price is biased *upward*. A conventional mimic would earn $\underline{p} \frac{1 - \underline{p}}{1 - e} = \frac{3}{4} \cdot \frac{1/4}{3/4} = \frac{1}{4}$, exactly its honest monopoly profit $\hat{\pi}_c^u = \frac{1}{4}$: it is just deterred. Skimming separates for every δ .

The same performance level $e = \frac{1}{4}$ and premium $\varphi = \frac{1}{8}$ thus produce *opposite* price signals—a rock-bottom 0 under competition, a premium $\frac{3}{4}$ under monopoly—purely because of the competitive environment.

8. What the Toy Model Keeps and Drops

Mahenc’s central messages all survive the simplification:

1. **Competition flips the signal.** Penetration (a low, rising price) in the occupied market; skimming (a high, falling price) in the untapped market (Results 1 and 2).
2. **Signaling is conspicuous waste.** The responsible type sacrifices profit during verification—an out-of-pocket loss under competition, forgone sales under monopoly—and the mimic cannot afford to copy it.
3. **Patience matters only under competition.** Penetration needs $\delta \geq \tilde{\delta}$; skimming works for all δ .
4. **Comparative statics.** Signaling is easier with more differentiation and a smaller cost premium, and harder the longer verification takes—with greenwashing filling the gap when the entrant is too impatient (Results 3 and 4).

What the toy model drops, relative to Mahenc (2025): (i) the general verification length T is collapsed to $T = 1$, so the trigger-strategy and never-dissuaded-once-convinced machinery that sustains repeated separation is replaced by a single period (Section 6 restores the $\tilde{\delta}^{1/T}$ scaling in one line); (ii) the multi-period pooling/greenwashing phase and its timing are only sketched, not derived; and (iii) the equilibrium-selection argument (intuitive criterion, Riley outcome) is stated rather than proved. Each can be reinstated at the cost of the dynamic bookkeeping that makes the full paper dense. For teaching, the two-period version already delivers the penetration -versus-skimming dichotomy and the Veblen logic behind it.

References

- Mahenc, Philippe (2025). “Penetration or Skimming Pricing for Credence Products?” *Journal of Economic Dynamics and Control* 180: 105196. <https://doi.org/10.1016/j.jedc.2025.105196>
- Bagwell, Kyle, and Michael H. Riordan (1991). “High and Declining Prices Signal Product Quality.” *American Economic Review* 81(1): 224–239.
- Cho, In-Koo, and David M. Kreps (1987). “Signaling Games and Stable Equilibria.” *Quarterly Journal of Economics* 102(2): 179–221.

Label Confusion and the Groucho Effect

A step-by-step exercise in the spirit of Harbaugh et al. (2011)

Companion note to Harbaugh, R., J. W. Maxwell, and B. Roussillon (2011), “Label Confusion: The Groucho Effect of Uncertain Standards,” *Management Science* 57(9): 1512–1527. Full reference at the end.

Overview. An eco-label certifies that a product meets a pass-fail standard, but consumers rarely know how demanding that standard is. A label then carries news about two unknowns at once: the product is at least as good as the standard, and the standard is no harder than the product. Harbaugh et al. (2011) call the second inference the *Groucho effect*, after Groucho Marx’s quip that he would not join any club willing to accept him as a member. When a product with a poor reputation displays a label, consumers conclude that the label must be easy to get. This note builds the mechanism in the simplest setting that keeps it intact: the standard is drawn uniformly on $[0, 1]$, which is the paper’s own leading example. Under this normalization a firm of quality q passes with probability q , so the label reweights the prior distribution of quality by q , and the equilibrium cutoffs have closed forms. Step (a) sets up the benchmark with a known standard; Step (b) derives the joint updating behind the Groucho effect; Step (c) measures how much a label reveals; Step (d) finds the labeling and non-labeling equilibria; Step (e) adds firm reputations; Step (f) studies label proliferation and focal labels; Step (g) turns to several firms and the strategic choice among labels; and Step (h) lists what the toy model keeps and drops.

1. The Toy Model

Quality. A firm sells a product of environmental quality $Q \in [0, 1]$, drawn from a cumulative distribution F with full support, mean $\mu \equiv E[Q]$, and variance $\sigma^2 \equiv \text{Var}(Q)$. The firm observes the realization q ; consumers know only F . Quality is fixed, so the firm makes no investment in it.

The label. A single eco-label has a standard $S \in [0, 1]$, drawn uniformly and independently of Q . The firm observes the realization s . If $q \geq s$ the firm may certify at a fixed cost $c > 0$ (certification fees, testing, documentation); if $q < s$ it cannot certify. The standard is *certain* if consumers also observe s , and *uncertain* if they know only that S is uniform.

Payoffs. The firm earns consumers’ expected quality given what they observe, minus c if it certifies. (The paper notes that its results hold whenever the firm’s payoff is a strictly increasing function of expected quality.)

The key normalization. Because S is uniform, a firm of quality q meets the standard with probability

$$\Pr(S \leq q) = q. \quad (1)$$

Figure 1 depicts (1). Panel (a) plots the density of the standard, $g(s) = 1$ on $[0, 1]$, and shades its part to the left of a quality level q : the shaded area is the probability that the standard is no harder than q . Panel (b) plots this probability against q , which gives the 45-degree line: a firm of higher quality is more likely to meet the standard, and a firm of quality 1 meets it for sure. A label therefore reweights each quality level q by its pass probability q .

The numerical examples below also take F uniform, so that $\mu = \frac{1}{2}$, $E[Q^2] = \frac{1}{3}$, and $\sigma^2 = \frac{1}{12}$.

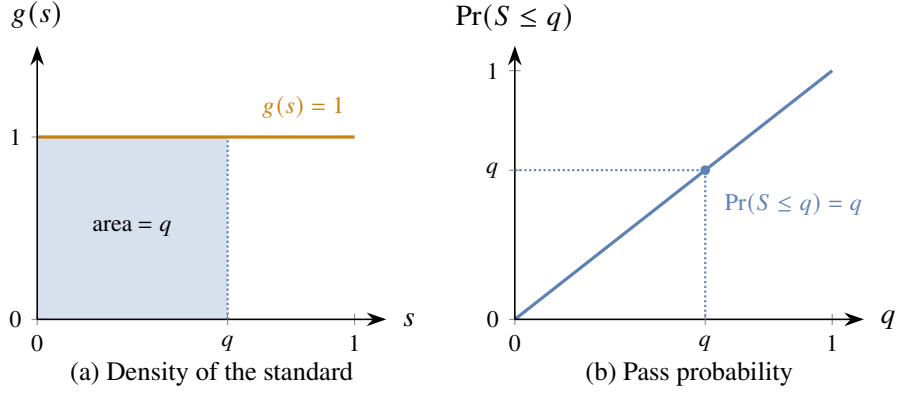


Figure 1: Probability that a uniform standard is no harder than quality q (shown for $q = 0.6$).

2. Step (a): The Benchmark with a Known Standard

Consider first a certain standard s , and assume that every firm able to certify does so. Consumers who see a label learn $Q \geq s$, and consumers who see no label learn $Q < s$. Write $\bar{q}(s) \equiv E[Q \mid Q \geq s]$ and $\underline{q}(s) \equiv E[Q \mid Q < s]$. Figure 2 depicts the two events on the quality interval $[0, 1]$ for a given s . With uniform F , quality is spread evenly over each event, so the expected quality in an event is the midpoint of its interval: $\underline{q}(s)$ is the midpoint of $[0, s]$ and $\bar{q}(s)$ is the midpoint of $[s, 1]$.

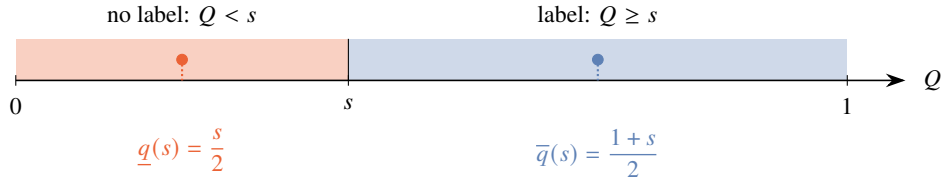


Figure 2: Expected quality given $Q < s$ and given $Q \geq s$, with uniform quality (shown for $s = 0.4$).

In formulas,

$$\bar{q}(s) = \frac{1+s}{2}, \quad \underline{q}(s) = \frac{s}{2}. \quad (2)$$

Averaging over the possible standards,

$$E[\bar{q}(S)] = \int_0^1 \frac{1+s}{2} ds = \frac{3}{4}, \quad E[\underline{q}(S)] = \int_0^1 \frac{s}{2} ds = \frac{1}{4}. \quad (3)$$

Figure 3 depicts the two integrals in (3): the shaded area under $\bar{q}(s)$ in panel (a) equals $\frac{3}{4}$, and the shaded area under $\underline{q}(s)$ in panel (b) equals $\frac{1}{4}$. Because the standard has density 1 on $[0, 1]$, each area is the average of the line over the standards.

These are the posterior means consumers would form if the conditional distribution of the standard did not change with the firm's labeling decision. Step (b) shows that it does change.

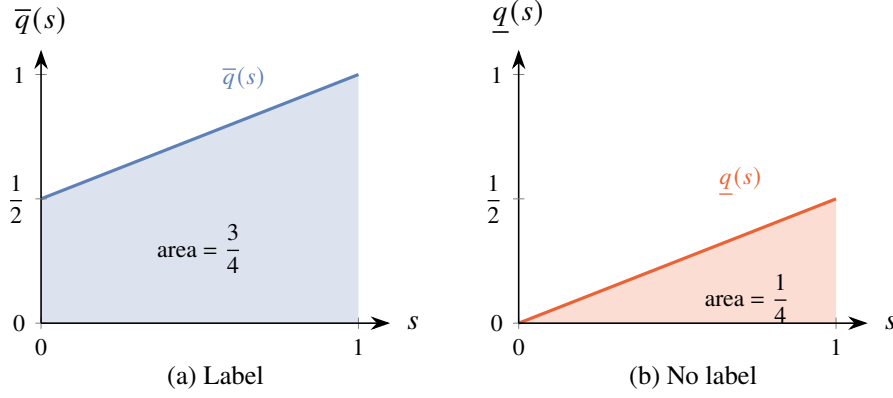


Figure 3: Expected quality given a known standard s , with uniform quality; shaded areas are $E[\bar{q}(S)]$ and $E[\underline{q}(S)]$.

3. Step (b): Joint Updating and the Groucho Effect

When the standard is uncertain, a label tells consumers only that $Q \geq S$ for some unobserved S . By (1), the expected quality of a labeled product is

$$E[Q | Q \geq S] = \frac{E[Q \Pr(S \leq Q | Q)]}{\Pr(S \leq Q)} = \frac{E[Q \cdot Q]}{E[Q]} = \frac{E[Q^2]}{\mu}, \quad (4)$$

and, by the same steps, the expected quality of an unlabeled product is

$$E[Q | Q < S] = \frac{E[Q(1 - Q)]}{1 - \mu} = \frac{\mu - E[Q^2]}{1 - \mu}. \quad (5)$$

With uniform F , $E[Q^2] = \frac{1}{3}$ and $\mu = \frac{1}{2}$, so these equal $\frac{2}{3}$ and $\frac{1}{3}$. Figure 4 depicts the reweighting behind the two means. A label multiplies the prior density $f(q)$ by the pass probability q and divides by μ , the probability of a label, which gives the density $2q$; no label multiplies it by $1 - q$ and divides by $1 - \mu$, which gives $2(1 - q)$. The first density puts more weight on high qualities, so its mean lies above the prior mean $\frac{1}{2}$, and the second puts more weight on low qualities, so its mean lies below it.

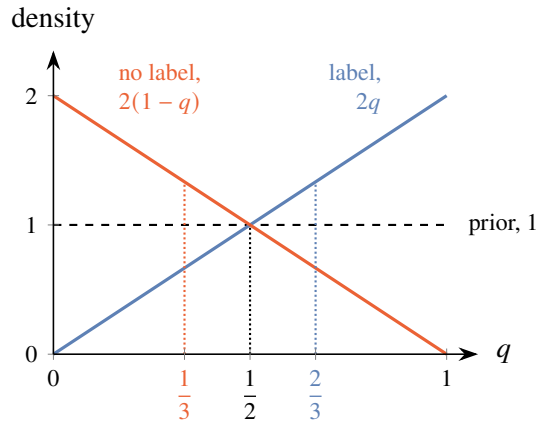


Figure 4: Density of quality before the label decision is observed (prior) and after a label or no label, with uniform quality and a uniform standard.

Consumers also revise their estimate of the standard. Conditional on a label, the density of S is

$$g(s \mid Q \geq S) = \frac{1 - F(s)}{\mu}, \quad (6)$$

which is decreasing in s : easy standards are over-represented among labeled products, because more firms clear them. Integrating,

$$E[S \mid Q \geq S] = \frac{E[Q^2]}{2\mu} = \frac{1}{3}, \quad E[S \mid Q < S] = \frac{1 - E[Q^2]}{2(1 - \mu)} = \frac{2}{3}, \quad (7)$$

where the second equality in each expression uses uniform F . Meeting the standard is good news about quality and bad news about the standard: the estimate of S falls from $\frac{1}{2}$ to $\frac{1}{3}$. Harbaugh et al. (2011) call this downward revision the *Groucho effect*. Failing to meet the standard works the other way: the estimate of S rises to $\frac{2}{3}$, the *reverse Groucho effect*.

Figure 5 depicts the revision of the standard. Conditional on a label, the density of S is $2(1 - s)$, from (6); conditional on no label, it is $F(s)/(1 - \mu) = 2s$, which is increasing in s because hard standards are over-represented among unlabeled products. The means of the two densities are the estimates $\frac{1}{3}$ and $\frac{2}{3}$ in (7).

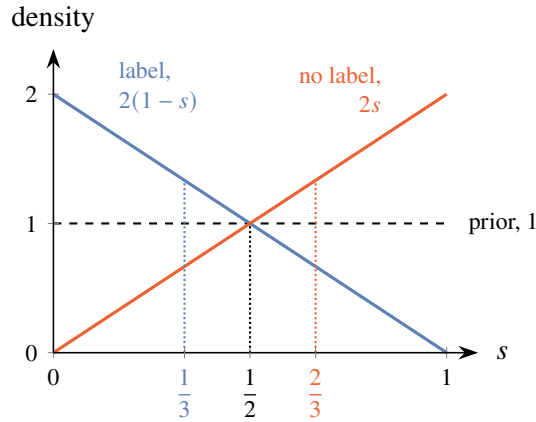


Figure 5: Density of the standard before the label decision is observed (prior) and after a label or no label, with uniform quality and a uniform standard.

Figure 6 depicts the two cases in the unit square, with the standard on the horizontal axis and quality on the vertical axis, for uniform F . In panel (a), with an uncertain standard, the prior point $(\frac{1}{2}, \frac{1}{2})$ moves to the centroid of the upper triangle $\{Q \geq S\}$ (shaded blue) after a label and to the centroid of the lower triangle $\{Q < S\}$ (shaded red) without one; both moves are diagonal, because the label shifts the estimates of Q and S in opposite directions. In panel (b), with a known standard s , the estimate of s cannot move, so the full update falls on quality: the posterior mean jumps to the upper line $\bar{q}(s)$ or the lower line $\underline{q}(s)$ of (2).

Comparing the two panels of Figure 6 with the averages in (3),

$$\underbrace{E[\underline{q}(S)]}_{\frac{1}{4}} < \underbrace{E[Q \mid Q < S]}_{\frac{1}{3}} < \underbrace{\mu}_{\frac{1}{2}} < \underbrace{E[Q \mid Q \geq S]}_{\frac{2}{3}} < \underbrace{E[\bar{q}(S)]}_{\frac{3}{4}}. \quad (8)$$

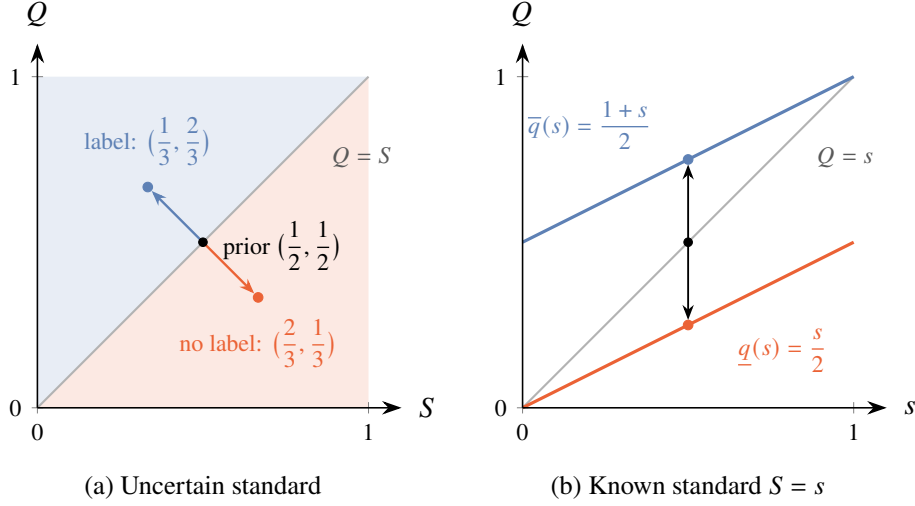


Figure 6: Updated estimates of quality and of the standard after a label and after no label, with uniform quality and a uniform standard.

Result 1 (moderation). Compared with a known standard, with an uncertain standard:

- a label is weaker good news;
- the absence of a label is weaker bad news.

The paper proves the ranking (8) for any F and G (inequality (6) of the paper); the uniform numbers make the size of the gap visible.

Figure 7 places the five numbers of (8) on the unit interval.

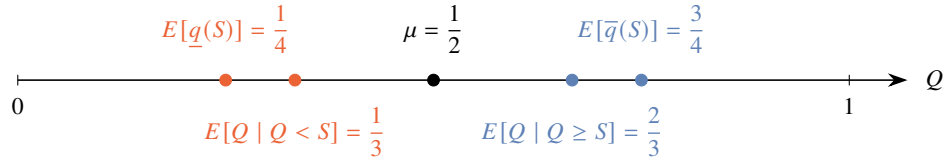


Figure 7: Expected quality after a label or no label, with a known and with an uncertain standard, for uniform quality and a uniform standard.

4. Step (c): How Much Does a Label Reveal?

Measure the information in a label by the share of the prior variance of quality that it removes,

$$R^2 \equiv 1 - \frac{E[(Q - \hat{Q})^2]}{\sigma^2}, \quad (9)$$

where $\hat{Q}$ is consumers' posterior mean. The numerator is the expected mean-squared error (MSE) that the paper uses, so a higher R^2 is a lower MSE. Without a label the MSE is $\sigma^2 = \frac{1}{12}$. With a known standard s the MSE equals $\frac{s^3 + (1-s)^3}{12}$, whose average over s is $\frac{1}{24}$, so $R^2 = \frac{1}{2}$. With an uncertain standard the two posterior means are $\frac{2}{3}$ and $\frac{1}{3}$, each reached with probability $\frac{1}{2}$, so the MSE is $\frac{1}{3} - (\frac{1}{2} \cdot \frac{4}{9} + \frac{1}{2} \cdot \frac{1}{9}) = \frac{1}{18}$ and $R^2 = \frac{1}{3}$.

Result 2 (Proposition 1 of the paper). A label is less informative on average when its standard is uncertain. A sum-of-squares identity shows why: the expected MSE with an uncertain standard minus the expected MSE with a known standard equals

$$\int_0^1 \left[F(s)(E[Q | Q < S] - \underline{q}(s))^2 + (1 - F(s))(E[Q | Q \geq S] - \bar{q}(s))^2 \right] dG(s) > 0, \quad (10)$$

because consumers who do not know s must use a single posterior mean in each event, while consumers who know s use the best mean for each s . With uniform F and G the integral equals $\frac{1}{18} - \frac{1}{24} = \frac{1}{72}$. Figure 8 depicts the expected MSE against the standard s for consumers who know s (blue curve) and for consumers who use the posterior means $\frac{2}{3}$ and $\frac{1}{3}$ whatever s is (red curve). The shaded area between the curves is the integral in (10), and the dotted lines mark the two averages $\frac{1}{24}$ and $\frac{1}{18}$. Consumers who know s place the posterior means at the midpoints of the two events, which minimizes the error at every s , so the blue curve lies below the red one everywhere. The red curve rises above the dashed line $\frac{1}{12}$, the MSE without a label, for standards below about 0.09 and above about 0.91, where consumers who do not know s misread the label; on average over s , the label still lowers the MSE from $\frac{1}{12}$ to $\frac{1}{18}$.

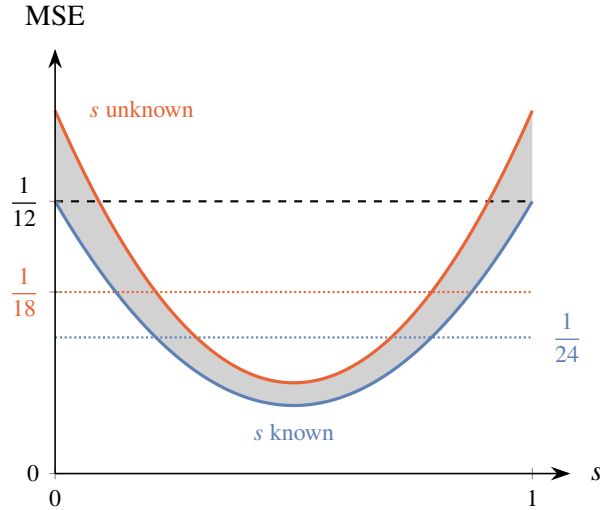


Figure 8: Expected MSE of consumers' estimate of quality against the standard s , with uniform quality; the shaded area is the integral in (10), and the dashed line is the MSE without a label.

Table 1 reports the uniform values.

Standard	$E[Q \text{label}]$	$E[Q \text{no label}]$	$E[S \text{label}]$	$E[S \text{no label}]$	R^2
Known (average over s)	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Uncertain	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$

Table 1: Posterior means and informativeness when every firm able to certify does so, with uniform quality and a uniform standard.

5. Step (d): Labeling and Non-Labeling Equilibria

The equilibrium concept is perfect Bayesian equilibrium with one refinement, described below. In a *labeling equilibrium* every firm that meets the standard certifies, so the lack of a label reveals failure. The firm certifies if the gain is at least the cost,

$$E[Q | Q \geq S] - E[Q | Q < S] \geq c. \quad (11)$$

In a *non-labeling equilibrium* no firm certifies, so the absence of a label is no news and the firm earns μ . A label is then off the equilibrium path, and the paper assumes that consumers attribute an unexpected label to any type that meets the standard with equal likelihood, so a firm that deviates earns $E[Q | Q \geq S]$. No firm deviates if

$$E[Q | Q \geq S] - \mu \leq c. \quad (12)$$

Closed forms. Substituting (4) and (5) into (11),

$$\begin{aligned} \bar{c} &\equiv E[Q | Q \geq S] - E[Q | Q < S] = \frac{E[Q^2]}{\mu} - \frac{\mu - E[Q^2]}{1 - \mu} \\ &= \frac{E[Q^2](1 - \mu) - \mu(\mu - E[Q^2])}{\mu(1 - \mu)} = \frac{E[Q^2] - \mu^2}{\mu(1 - \mu)} = \frac{\sigma^2}{\mu(1 - \mu)}, \end{aligned} \quad (13)$$

and, from (12),

$$\underline{c} \equiv E[Q | Q \geq S] - \mu = \frac{E[Q^2] - \mu^2}{\mu} = \frac{\sigma^2}{\mu} = (1 - \mu) \bar{c}. \quad (14)$$

With a uniform standard, the labeling cutoff is the variance of quality relative to its largest possible value given the mean, $\mu(1 - \mu)$. With uniform F , $\underline{c} = \frac{1}{6}$ and $\bar{c} = \frac{1}{3}$.

Figure 9 depicts the two incentive conditions for uniform F , with the certification cost on the horizontal axis. In panel (a), the labeling equilibrium, a firm that meets the standard compares the payoff from certifying, $E[Q | Q \geq S] - c = \frac{2}{3} - c$ (blue line), with the payoff from staying silent, $E[Q | Q < S] = \frac{1}{3}$ (red line): certifying pays up to $c = \frac{1}{3} = \bar{c}$. In panel (b), the non-labeling equilibrium, a firm compares the payoff from deviating to a label, $\frac{2}{3} - c$, with the equilibrium payoff $\mu = \frac{1}{2}$: deviating does not pay from $c = \frac{1}{6} = \underline{c}$ on.

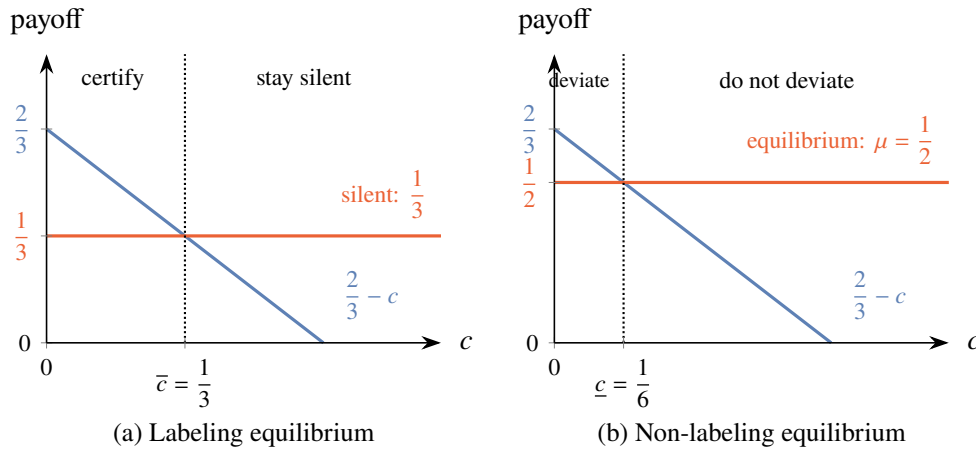


Figure 9: A firm's payoff from each action against the certification cost c , with uniform quality and a uniform standard.

Result 3 (Proposition 2 of the paper).

- A labeling equilibrium exists if and only if $c \leq \bar{c}$.
- A non-labeling equilibrium exists if and only if $c \geq \underline{c}$.
- Both exist if and only if $\underline{c} \leq c \leq \bar{c}$.

Since $\underline{c} = (1 - \mu)\bar{c} < \bar{c}$, the interval of multiple equilibria is never empty. The two cutoffs split the cost axis into three regions, which the figures below name by the equilibria they support:

- L: $c < \underline{c}$, only the labeling equilibrium exists;
- LN: $\underline{c} \leq c \leq \bar{c}$, both equilibria exist;
- N: $c > \bar{c}$, only the non-labeling equilibrium exists.

The known-standard benchmark. With a known standard s and uniform F , (2) gives

$$\bar{c}(s) = \bar{q}(s) - \underline{q}(s) = \frac{1}{2}, \quad \underline{c}(s) = \bar{q}(s) - \mu = \frac{s}{2}, \quad (15)$$

so a labeling equilibrium exists for $c \leq \frac{1}{2}$ whatever the standard, and a non-labeling equilibrium exists for $c \geq \frac{s}{2}$, a cutoff whose average over s is $\frac{1}{4}$. Figure 10 depicts the three regions for a known standard, with the standard s on the horizontal axis and the cost on the vertical axis. The labeling cutoff $\bar{c}(s) = \frac{1}{2}$ is flat, and the non-labeling cutoff $\underline{c}(s) = \frac{s}{2}$ rises with s , so the region of multiple equilibria LN is the wedge between them; the dotted line marks the average $\frac{1}{4}$ of the rising cutoff.

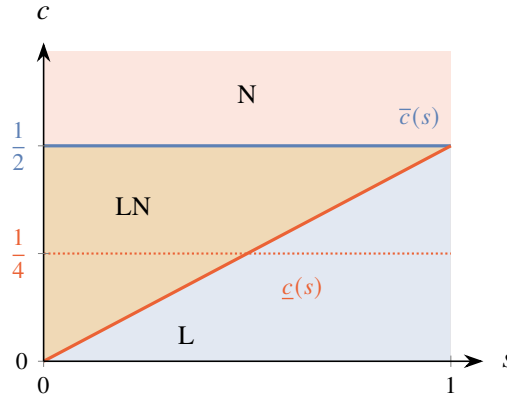


Figure 10: Labeling (L), multiple-equilibria (LN), and non-labeling (N) regions with a known standard s , for uniform quality.

Result 4 (Proposition 3 of the paper). Uncertainty about the standard:

- shrinks the range of costs that supports labeling;
- enlarges the range of costs that supports non-labeling.

Figure 11 depicts the two ranges on a cost axis for uniform quality: labeling survives up to $c = \frac{1}{3}$ instead of $\frac{1}{2}$, and non-labeling starts at $c = \frac{1}{6}$ instead of $\frac{1}{4}$. Intuitively, the Groucho effect makes a label less valuable, which hurts the labeling equilibrium, and the reverse Groucho effect makes the lack of a label less damaging, which helps the non-labeling equilibrium.

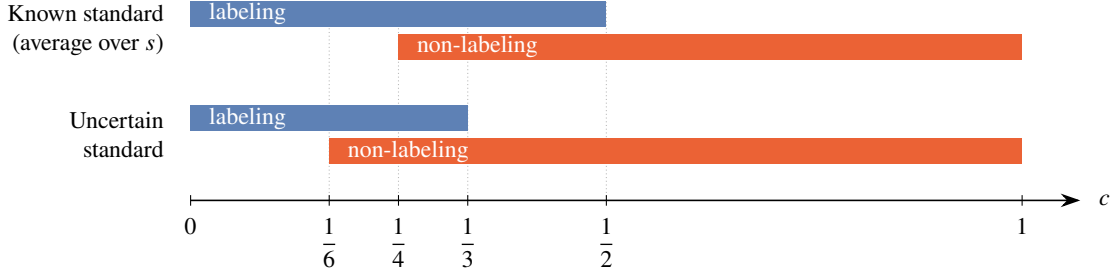


Figure 11: Costs that support a labeling equilibrium and a non-labeling equilibrium, with uniform quality and a uniform standard.

6. Step (e): Reputations

Let quality follow a Beta distribution with parameters (α, β) , so that $\mu = \frac{\alpha}{\alpha + \beta}$ and $\sigma^2 = \frac{\mu(1-\mu)}{\alpha + \beta + 1}$. Substituting into (13) and (14),

$$\bar{c} = \frac{1}{\alpha + \beta + 1}, \quad \underline{c} = \frac{\beta}{(\alpha + \beta)(\alpha + \beta + 1)}. \quad (16)$$

The paper models a firm with a *bad reputation* as $Q \sim \text{Beta}(1, b)$ with $b \geq 1$, so that $\mu = \frac{1}{1+b} \leq \frac{1}{2}$, and a firm with a *good reputation* as $Q \sim \text{Beta}(a, 1)$ with $a \geq 1$, so that $\mu \geq \frac{1}{2}$. Writing each family in terms of its mean, (16) becomes

$$\bar{c} = \begin{cases} \frac{\mu}{1 + \mu} & \text{if } \mu \leq \frac{1}{2} \text{ (bad reputation),} \\ \frac{1 - \mu}{2 - \mu} & \text{if } \mu \geq \frac{1}{2} \text{ (good reputation),} \end{cases} \quad \underline{c} = (1 - \mu) \bar{c}. \quad (17)$$

Figure 12 depicts the two families for $b = a = 3$, together with the uniform density. A bad reputation piles the density of quality near $q = 0$ (mean $\frac{1}{4}$), and a good reputation piles it near $q = 1$ (mean $\frac{3}{4}$).

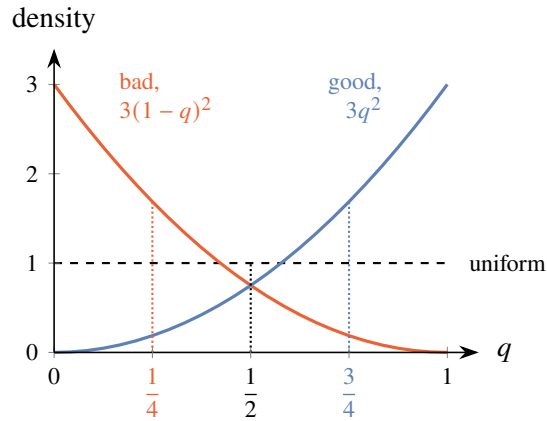


Figure 12: Density of quality for a bad reputation, Beta(1,3), a good reputation, Beta(3,1), and the uniform distribution.

Figure 13 depicts the regions L, LN, and N of Step (d) against the prior mean μ , with the certification cost on the vertical axis: panel (a) for an uncertain standard, from (17), and panel (b) for a known standard, where the cutoffs are averaged over s . In panel (a) the labeling cutoff $\bar{c}$ peaks at $\mu = \frac{1}{2}$, where $\bar{c} = \frac{1}{3}$, and falls to

zero as the reputation becomes very bad or very good; the non-labeling cutoff $\underline{c}$, below which labeling is the only equilibrium, peaks at $\mu = \sqrt{2} - 1$, where $\underline{c} = 3 - 2\sqrt{2}$. Intuitively, consumers who see a label on a product with a bad reputation conclude that the standard must be easy: in the bad-reputation family,

$$E[S | Q \geq S] = \frac{E[Q^2]}{2\mu} = \frac{1}{b+2}, \quad (18)$$

which falls to zero as b grows, so the label is worth little. For a product with a good reputation the reverse Groucho effect dominates: consumers who see no label conclude that the standard must be hard, since in the good-reputation family

$$E[S | Q < S] = \frac{1 - E[Q^2]}{2(1 - \mu)} = \frac{a+1}{a+2}, \quad (19)$$

which rises to one as a grows, so the firm loses little by not certifying. In panel (b), with a known standard, a firm with a bad reputation that is in fact of high quality can prove it, and the labeling cutoff stays above 0.45 for every reputation.

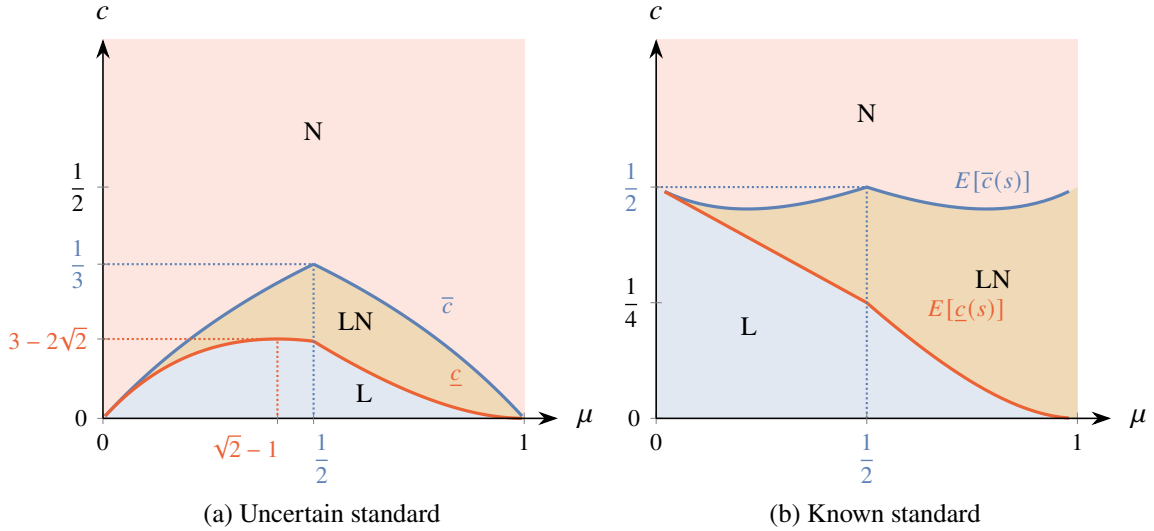


Figure 13: Labeling (L), non-labeling (N), and multiple-equilibria (LN) regions against the prior mean μ , with a uniform standard; $Q \sim \text{Beta}(1, b)$ for $\mu \leq \frac{1}{2}$ and $Q \sim \text{Beta}(a, 1)$ for $\mu \geq \frac{1}{2}$.

Figure 14 depicts the posterior estimates of the standard in (18) and (19) for every mean μ , from the same two families. In terms of μ ,

$$E[S | Q \geq S] = \begin{cases} \frac{\mu}{1+\mu} & \text{if } \mu \leq \frac{1}{2}, \\ \frac{1}{2(2-\mu)} & \text{if } \mu \geq \frac{1}{2}, \end{cases} \quad E[S | Q < S] = \begin{cases} \frac{1+2\mu}{2(1+\mu)} & \text{if } \mu \leq \frac{1}{2}, \\ \frac{1}{2-\mu} & \text{if } \mu \geq \frac{1}{2}. \end{cases} \quad (20)$$

A label lowers the estimate of the standard below the prior value $\frac{1}{2}$ (blue curve), most strongly when the reputation is bad; the lack of a label raises it above $\frac{1}{2}$ (red curve), most strongly when the reputation is good.

What drives the hump. Equation (16) shows that, with a uniform standard, the labeling cutoff $\bar{c} = \frac{1}{\alpha+\beta+1}$ depends only on how concentrated the prior is, measured by $\alpha + \beta$, and not on where it is centered. In the two families above, moving the mean toward 0 or 1 also raises $\alpha + \beta$ (to $1 + b$ or $a + 1$), and that is what shrinks the labeling region at both ends of Figure 13(a). The paper points to this in its footnote 6, where it

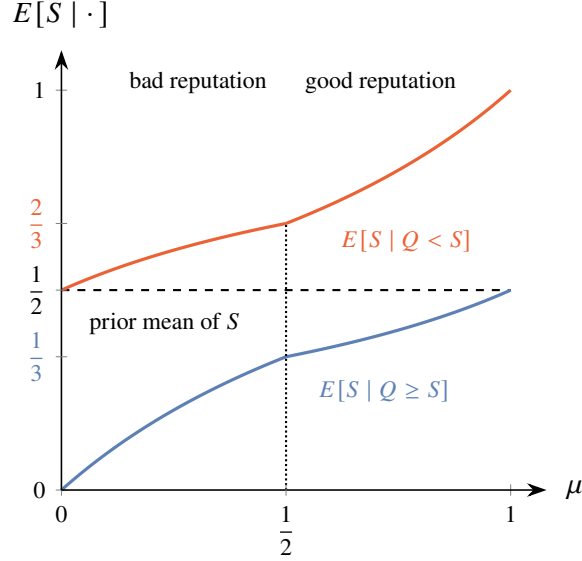


Figure 14: Expected standard after a label and after no label against the prior mean μ , with a uniform standard; $Q \sim \text{Beta}(1, b)$ for $\mu \leq \frac{1}{2}$ and $Q \sim \text{Beta}(a, 1)$ for $\mu \geq \frac{1}{2}$.

notes that these families are most diffuse at $\mu = \frac{1}{2}$. A firm with mean $\frac{1}{4}$ and $\text{Beta}(\frac{1}{2}, \frac{3}{2})$ quality, for instance, has the same $\bar{c} = \frac{1}{3}$ as a firm with uniform quality. The strength of the Groucho effect does depend on the mean: for any Beta prior, $E[S | Q \geq S] = \frac{\alpha+1}{2(\alpha+\beta+1)}$, which, for a given concentration $\alpha + \beta$, is lower the worse the reputation. A label is less valuable for a product with a bad reputation because it signals an easy standard, and also because a concentrated prior leaves little to learn.

7. Step (f): Label Proliferation and Focal Labels

Now let there be $n \geq 1$ labels with standards drawn independently from the uniform distribution and the same cost c , and let the firm display at most one label. Consider the *symmetric* strategy: the firm adopts the toughest label it meets. Consumers cannot tell which label is toughest, and the labels are interchangeable to them, so a label tells consumers only that the firm meets at least the *easiest* standard, $S_{1:n} \equiv \min\{S_1, \dots, S_n\}$. A firm of quality q meets it with probability

$$\Pr(S_{1:n} \leq q) = 1 - (1 - q)^n. \quad (21)$$

Using (21) in place of (1) in the steps behind (4) and (5), and taking uniform F ,

$$E[Q | Q \geq S_{1:n}] = \frac{n+3}{2(n+2)}, \quad E[Q | Q < S_{1:n}] = \frac{1}{n+2}. \quad (22)$$

Figure 15 depicts (21) in panel (a) and (22) in panel (b). Panel (a) shows that, for any given quality, the probability of meeting the easiest standard rises with n , so a label becomes easier to get; panel (b) shows that both posterior means fall with n , toward $\mu = \frac{1}{2}$ and toward 0, so failing every label becomes a sign of very low quality.

The two cutoffs are

$$\bar{c}_n = \frac{n+1}{2(n+2)}, \quad \underline{c}_n = \frac{1}{2(n+2)}. \quad (23)$$

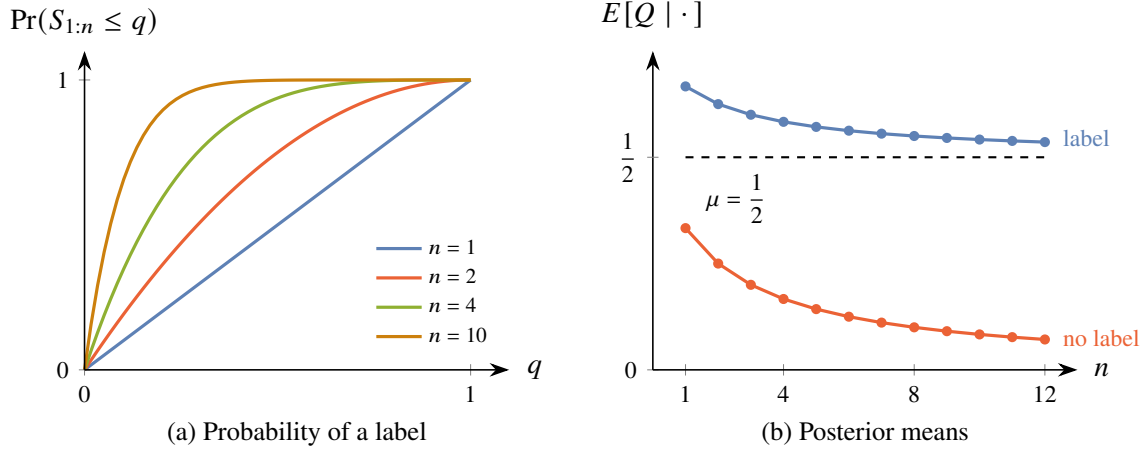


Figure 15: Probability of meeting the easiest of n uniform standards, and expected quality given a label and given no label, with uniform quality.

Result 5 (Proposition 4 of the paper). With uncertain standards, adding labels enlarges the range of costs that supports the non-labeling equilibrium. As n grows:

- non-labeling becomes an equilibrium for every $c > 0$;
- labeling remains an equilibrium if and only if $c \leq \mu$;
- the labeling equilibrium becomes uninformative.

In the uniform case the cutoffs in (23) are exact for every n : for a given cost c , a non-labeling equilibrium exists if and only if $n \geq \frac{1}{2c} - 2$, and a labeling equilibrium exists if and only if $n \geq \frac{4c-1}{1-2c}$ (for $c < \frac{1}{2}$). As n grows, $\underline{c}_n \rightarrow 0$ and $\bar{c}_n \rightarrow \frac{1}{2} = \mu$, so the interval of multiple equilibria widens toward $[0, \frac{1}{2}]$.

Informativeness. Under the symmetric strategy the label splits quality at the random cutoff $S_{1:n}$, and from (9) and (22),

$$R_{\text{sym}}^2(n) = \frac{3n}{(n+2)^2}. \quad (24)$$

With one label $R_{\text{sym}}^2 = \frac{1}{3}$, with two labels it rises to $\frac{3}{8}$, it returns to $\frac{1}{3}$ at $n = 4$, and it falls toward zero after that. Intuitively, a second label makes the lack of any label a sharper signal of low quality, and at first this outweighs the fact that a label becomes easier to get; with many labels, almost every firm meets the easiest standard, so a label proves almost nothing. The paper's Proposition 4 concerns the limit $n \rightarrow \infty$; the uniform example shows that the decline in informativeness need not start at $n = 1$.

Focal labels. Consider now that one label, say label X with standard S_X , is *focal*: consumers expect a firm to adopt it whenever it can, and to adopt another label only if it fails X . The market then sorts firms into three groups:

$$\{Q \geq S_X\}, \quad \{S_{1:n} \leq Q < S_X\}, \quad \{Q < S_{1:n}\}, \quad (25)$$

with posterior means $\frac{2}{3}$, $\frac{n+4}{3(n+2)}$, and $\frac{1}{n+2}$ in the uniform case. Figure 16 depicts the partitions of the quality interval for two labels, X and Y , with realized standards $s_Y < s_X$, so that $s_{1:2} = s_Y$. Under the symmetric strategy consumers see only whether a firm has some label, and they split quality at $s_{1:2}$; under the focal strategy they also see whether the label is X , and the labeled group splits at s_X into two groups.

A firm in the middle group must prefer some label to none, $\frac{n+4}{3(n+2)} - \frac{1}{n+2} \geq c$, and a firm in the top group must prefer label X to no label, $\frac{2}{3} - \frac{1}{n+2} \geq c$ (it also prefers X to any other label, since $\frac{2}{3}$ exceeds the middle-group

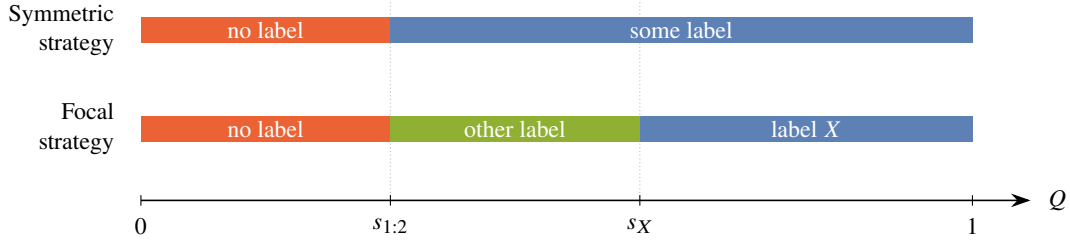


Figure 16: Groups of firms by quality under the symmetric and the focal strategies, for two labels with realized standards $s_{1:2} = 0.3$ and $s_X = 0.6$.

mean); the first condition binds, so for $n \geq 2$ a focal labeling equilibrium exists if and only if

$$c \leq \frac{n+1}{3(n+2)}. \quad (26)$$

The cutoff in (26) equals $\frac{2}{3}\bar{c}_n$, so a focal equilibrium requires a lower cost than the symmetric labeling equilibrium. The partition (25) is finer than the symmetric one, and

$$R_{\text{focal}}^2(n) = \frac{n(n+8)}{3(n+2)^2}, \quad R_{\text{focal}}^2(n) - R_{\text{sym}}^2(n) = \frac{n(n-1)}{3(n+2)^2} \geq 0. \quad (27)$$

Result 6 (Proposition 5 of the paper).

- With uncertain standards, a focal labeling equilibrium exists if the cost satisfies (26), and it is more informative than the symmetric one.
- With known standards, no focal labeling equilibrium exists.

Figure 17 depicts the cutoffs (23) and (26) in panel (a), with the regions L, LN, and N of the symmetric labeling strategy shaded, and the informativeness of the label in (24) and (27) in panel (b), with the number of labels on the horizontal axis. Panel (b) shows R_{sym}^2 peaking at $\frac{3}{8}$ with two labels, and R_{focal}^2 peaking at $\frac{4}{9}$ with four labels and converging to $\frac{1}{3}$, the value for a single uncertain label, because non-focal labels become uninformative as they multiply. With known standards a firm always prefers the toughest label it meets, and under the paper's refinement a deviation to a tougher label is rewarded, so focality cannot be sustained. This is the paper's case for “look for the label” campaigns and mandatory labels: they keep the information that a single label conveys, which proliferation destroys under the symmetric strategy.

Table 2 reports the uniform values.

Known standards. With known standards, more labels let a firm reveal its quality more finely, and in the limit the problem becomes the one Lizzeri (1999) studies, in which every quality level can be certified. The paper shows that as n grows, non-labeling is an equilibrium almost surely if and only if $\mu \geq 1 - c$, and, for log-concave F , labeling is an equilibrium almost surely if and only if $\mu \leq 1 - c$. With uniform quality both conditions turn on whether c is above or below $\frac{1}{2}$, and a label becomes highly informative, instead of uninformative as with uncertain standards.

8. Step (g): Several Firms and the Strategic Choice of Labels

One label, many firms. Let m firms with independent qualities from F face one label. In a labeling equilibrium the share of firms with a label converges to $1 - F(s)$, so consumers learn the standard, and

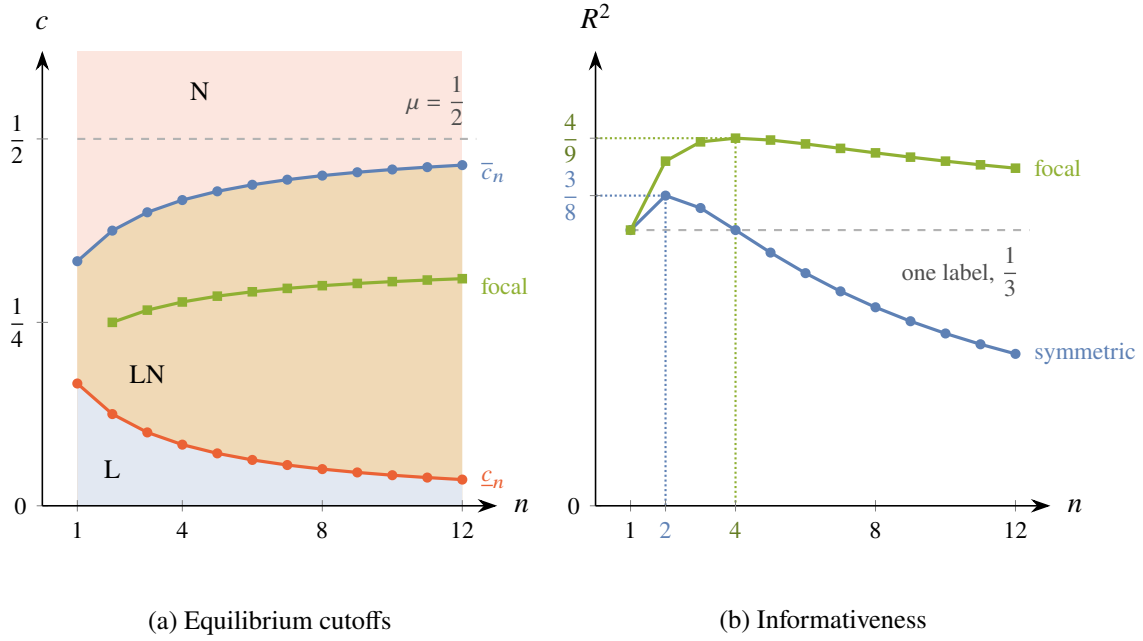


Figure 17: Equilibrium cutoffs and informativeness of a label against the number of labels n , with uniform quality and uniform standards.

Labels n	$\underline{c}_n$	$\bar{c}_n$	Focal cutoff	R^2_{sym}	R^2_{focal}
1	$\frac{1}{6}$	$\frac{1}{3}$	—	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{5}{12}$
3	$\frac{1}{10}$	$\frac{2}{5}$	$\frac{4}{15}$	$\frac{9}{25}$	$\frac{11}{25}$
4	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{18}$	$\frac{1}{3}$	$\frac{4}{9}$
6	$\frac{1}{16}$	$\frac{7}{16}$	$\frac{7}{24}$	$\frac{9}{32}$	$\frac{7}{16}$
10	$\frac{1}{24}$	$\frac{11}{24}$	$\frac{11}{36}$	$\frac{5}{24}$	$\frac{5}{12}$
$\rightarrow \infty$	0	$\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{1}{3}$

Table 2: Cutoffs and informativeness under the symmetric and the focal labeling strategies, with uniform quality and uniform standards.

each firm effectively faces a known standard: with uniform quality, labeling is supported for $c \leq \bar{c}(s) = \frac{1}{2}$ whatever the standard. In a non-labeling equilibrium no firm certifies, so consumers have nothing to learn from, and the condition stays $c \geq \underline{c} = \frac{1}{6}$ (Proposition 6 of the paper). As the number of firms grows, the interval of multiple equilibria therefore widens from $[\frac{1}{6}, \frac{1}{3}]$ to $[\frac{1}{6}, \frac{1}{2}]$. Figure 18 depicts the two ranges of costs, as in Figure 11: with many firms, labeling survives up to $c = \frac{1}{2}$ instead of $\frac{1}{3}$, and non-labeling still starts at $c = \frac{1}{6}$.

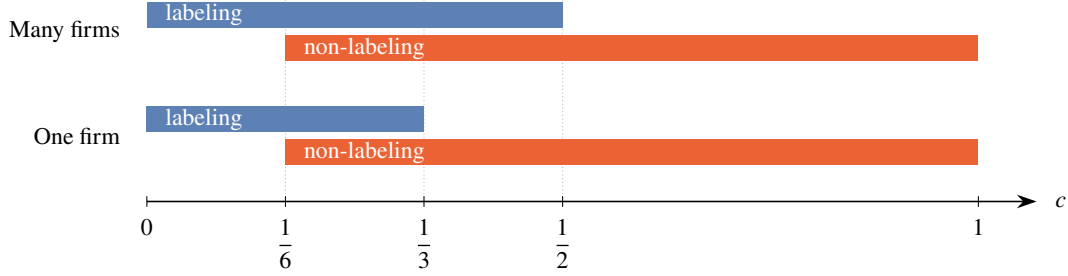


Figure 18: Costs that support a labeling equilibrium and a non-labeling equilibrium with one firm and with many firms, for uniform quality and a uniform standard.

Two firms, two labels. With two firms and two labels, the label one firm adopts carries news about the standards the other firm faces. Under the symmetric strategy, consumers see either both firms with the *same* label or the firms with *different* labels, and in the second case they cannot tell which firm met the tougher standard. Denoting by $S_{1:2} < S_{2:2}$ the easier and the tougher standard, the events are

$$\text{Same} = \{S_{1:2} \leq Q_1 < S_{2:2}, S_{1:2} \leq Q_2 < S_{2:2}\} \cup \{Q_1 \geq S_{2:2}, Q_2 \geq S_{2:2}\}, \quad (28)$$

$$\text{Different} = \{S_{1:2} \leq Q_1 < S_{2:2} \leq Q_2\} \cup \{S_{1:2} \leq Q_2 < S_{2:2} \leq Q_1\}. \quad (29)$$

Figure 19 depicts the two events for given standards, in the plane of the two firms' qualities. The events Same and Different are the shaded cells; in the remaining region at least one firm meets neither standard and has no label.

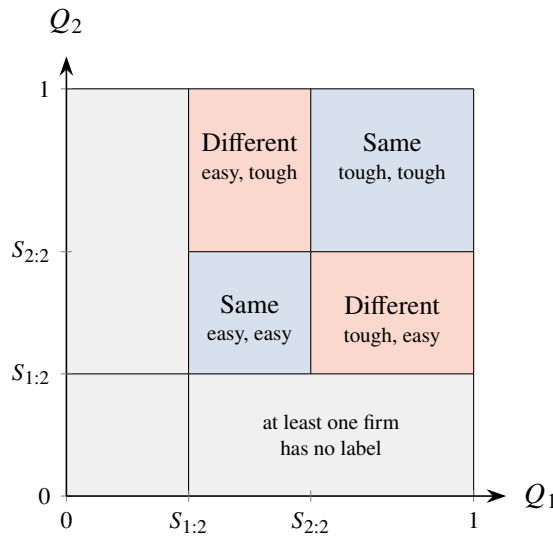


Figure 19: Events Same and Different in the plane of the two qualities, for two labels with given standards $S_{1:2} < S_{2:2}$; each cell lists the label of firm 1 and of firm 2.

A firm that meets both standards while its rival meets only the easier one can pool with the rival by adopting the easier label, and a firm that meets both standards while its rival does too can separate by adopting the easier label. The symmetric strategy survives only if neither deviation pays, which requires $E[Q_i | \text{Same}] = E[Q_i | \text{Different}]$. With uniform quality the two events occur with probabilities $\frac{1}{3}$ and $\frac{1}{6}$, and

$$E[Q_i | \text{Same}] = E[Q_i | \text{Different}] = \frac{3}{5}, \quad (30)$$

so the uniform case is exactly the knife-edge at which the symmetric strategy is an equilibrium.

Figure 20 depicts both conditional means against the prior mean μ , for the two reputation families of Step (e), with uniform standards. The curves cross at $\mu = \frac{1}{2}$, at the value $\frac{3}{5}$ in (30). To the right of the crossing, the same label is better news than different labels: when reputations are good, the label a rival adopts is likely to be the tougher one, so each firm wants to adopt it too. The paper says that a firm with a good reputation *legitimizes* the label it adopts. To the left of the crossing, different labels are better news: the label a rival with a bad reputation adopts is likely to be the easier one, so each firm wants to avoid it, and a firm with a bad reputation *spoils* the label it adopts. Away from $\mu = \frac{1}{2}$ the symmetric labeling equilibrium does not exist, while a focal labeling equilibrium exists under the same cost condition as with a single firm (Proposition 7 of the paper). When firms choose in sequence, these incentives produce labeling cascades: the second firm herds onto the first firm's label if reputations are good and moves away from it if they are bad. The paper offers this spillover as one reason why certifiers recruit reputable firms as early adopters of a new label.

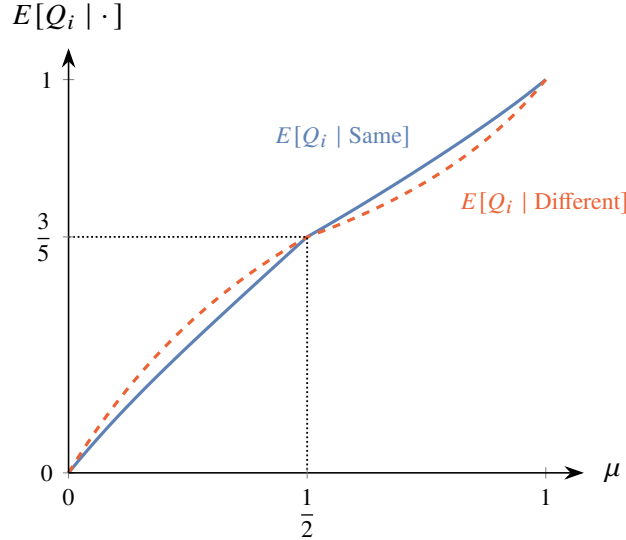


Figure 20: Expected quality of firm i when both firms display the same label and when they display different labels, against the prior mean μ ; two firms, two uniform standards, $Q \sim \text{Beta}(1, b)$ for $\mu \leq \frac{1}{2}$ and $Q \sim \text{Beta}(a, 1)$ for $\mu \geq \frac{1}{2}$.

9. Step (h): What the Toy Model Keeps and Drops

Every one of the paper's main conclusions survives the uniform standard:

1. The Groucho and reverse Groucho effects: a label lowers the estimate of the standard from $\frac{1}{2}$ to $\frac{1}{3}$ and its absence raises it to $\frac{2}{3}$ (Step b).
2. Uncertain standards reveal less: R^2 falls from $\frac{1}{2}$ to $\frac{1}{3}$ (Step c, Proposition 1).

3. Labeling is harder to sustain and non-labeling easier: the cutoffs move from $\frac{1}{2}$ and $\frac{1}{4}$ to $\frac{1}{3}$ and $\frac{1}{6}$, and for any F , $\bar{c} = \frac{\sigma^2}{\mu(1-\mu)}$ and $\underline{c} = \frac{\sigma^2}{\mu}$ (Step d, Propositions 2 and 3).
4. Proliferation makes non-labeling an equilibrium for every $c > 0$ and the labeling equilibrium uninformative, while a focal label restores information (Step f, Propositions 4 and 5).
5. Several firms widen the interval of multiple equilibria, and with several labels firms legitimize or spoil labels (Step g, Propositions 6 and 7).

With a uniform standard, the labeling cutoff depends on the prior only through $\frac{\sigma^2}{\mu(1-\mu)}$, which separates the role of a firm's reputation (its mean) from the precision of consumers' prior (its dispersion); the general analysis leaves this split implicit. In addition, the informativeness of the symmetric labeling equilibrium is not monotone in the number of labels: in the uniform case it peaks at two labels and declines after that.

Relative to Harbaugh et al. (2011), the toy model takes the standard to be uniform, so the pass probability (1) is linear in quality; the paper allows any distribution G , and its propositions hold for any F and G . As in the paper, the firm's payoff is linear in expected quality, without prices or product-market competition between firms, and quality and standards are fixed: firms do not invest in quality, and certifiers do not choose their standards or fees. Each of these features can be relaxed one at a time.

References

- Harbaugh, R., J. W. Maxwell, and B. Roussillon (2011). "Label Confusion: The Groucho Effect of Uncertain Standards." *Management Science*, 57(9), pp. 1512–1527. <https://doi.org/10.1287/mnsc.1110.1412>
- Lizzeri, A. (1999). "Information Revelation and Certification Intermediaries." *RAND Journal of Economics*, 30(2), pp. 214–231.

EconS 594 – Ph.D. Industrial Organization

Fall 2026

Eco-labeling with Noisy Certification Tests: a Toy Model

Overview. This note works through a stripped-down version of Mason’s (2011) eco-labeling market. Each seller offers a single unit of a credence good, consumers value a green unit at 1 and a brown unit at 0, and the certifier runs a noisy test indexed by a single *precision* $\tau \in [0, 1]$. With these normalizations the analysis is short enough to fit on a whiteboard: **posterior beliefs equal equilibrium prices**, and the equilibrium conditions are linear in those beliefs. Seven steps follow: (a) the no-information benchmark, (b) the noisy test, (c) Bayes’ rule and prices, (d) application payoffs and the ranking $W_G > W_B$, (e) the equilibrium classes and the mixing probability in detail, (f) a numerical example, and (g) a note on what the toy model keeps and drops relative to Mason.

1. The Toy Model

Consumers and demand. Consumers value a green unit more than a brown one but cannot tell them apart directly. As in Mason (2011), demand is perfectly elastic. **Normalize** the two willingness-to-pay levels to

$$v_G = 1, \quad v_B = 0,$$

so the green “premium” is the whole willingness to pay. A consumer who believes a unit is green with probability μ pays the expected value $P(\mu) = \mu v_G + (1 - \mu) v_B = \mu$: *prices are beliefs*.

Sellers. There is a unit mass of sellers, each offering one unit. A fraction $\lambda \in (0, 1)$ are green and the complementary fraction $1 - \lambda$ are brown; a seller privately knows its own type. Because each seller brings a single unit to market, a seller’s payoff is simply the price it receives, net of any certification fee it pays. This single-unit structure is what keeps the equilibrium conditions linear in the beliefs.

Market structure. Within each observable segment (labeled or unlabeled) the market is competitive: the price equals consumers’ expected willingness to pay given the belief that a random unit in that segment is green. This is Mason’s rational-expectations pricing rule.

2. Step (a): The No-Information Benchmark

Suppose no eco-label is available. The whole market pools, and consumers pay the expected willingness to pay for a random unit,

$$P_0 = \lambda v_G + (1 - \lambda) v_B = \lambda. \quad (1)$$

Every seller, green or brown, receives $P_0 = \lambda$. This is the reference price against which the label is measured, and the outcome the market reverts to whenever the fee is too high for anyone to seek certification.

3. Step (b): The Noisy Certification Test

A certifying agency offers a test. A seller that applies pays a fee $A \geq 0$ and receives a public verdict (pass or fail). We index the test by a single *precision* $\tau \in [0, 1]$: a green seller passes with probability ϕ_G and a brown seller with probability ϕ_B , where

$$\phi_G \equiv \Pr(\text{pass} \mid G) = \frac{1 + \tau}{2}, \quad \phi_B \equiv \Pr(\text{pass} \mid B) = \frac{1 - \tau}{2}. \quad (2)$$

At $\tau = 0$ the test is uninformative ($\phi_G = \phi_B = \frac{1}{2}$); at $\tau = 1$ it is perfect ($\phi_G = 1, \phi_B = 0$). A green seller fails with probability $1 - \phi_G = \frac{1 - \tau}{2}$ (a *type I* error, or false rejection), and a brown seller passes with probability $\phi_B = \frac{1 - \tau}{2}$ (a *type II* error, or false acceptance). The two error rates are equal, so τ is the single dial that sets test quality, with informativeness $\phi_G - \phi_B = \tau$. Table 1 collects the four outcomes conditional on applying, and Table 2 shows its two extremes. Figure 1 plots the two pass probabilities against τ , and Figure 2 shows the same four flows as a tree.

	Labeled (pass)	Unlabeled (fail)
Green seller	$\frac{1 + \tau}{2}$	$\frac{1 - \tau}{2}$ (type I error)
Brown seller	$\frac{1 - \tau}{2}$ (type II error)	$\frac{1 + \tau}{2}$

Table 1: Certification outcomes by seller type, conditional on applying. Under the symmetric precision test (2), the diagonal (correct classification) carries mass $\frac{1 + \tau}{2}$ and each off-diagonal error carries mass $\frac{1 - \tau}{2}$; a more precise test (higher τ) moves mass onto the diagonal.

$\tau = 0$	Labeled	Unlabeled
Green seller	$\frac{1}{2}$	$\frac{1}{2}$
Brown seller	$\frac{1}{2}$	$\frac{1}{2}$

(a) Uninformative test.

$\tau = 1$	Labeled	Unlabeled
Green seller	1	0
Brown seller	0	1

(b) Perfect test.

Table 2: The two extremes of the confusion matrix. At $\tau = 0$ (panel a) every cell equals $\frac{1}{2}$: the label is pure noise, so the two rows are identical and a pass says nothing about type. At $\tau = 1$ (panel b) the diagonal is 1 and both errors vanish: the label separates green from brown perfectly.

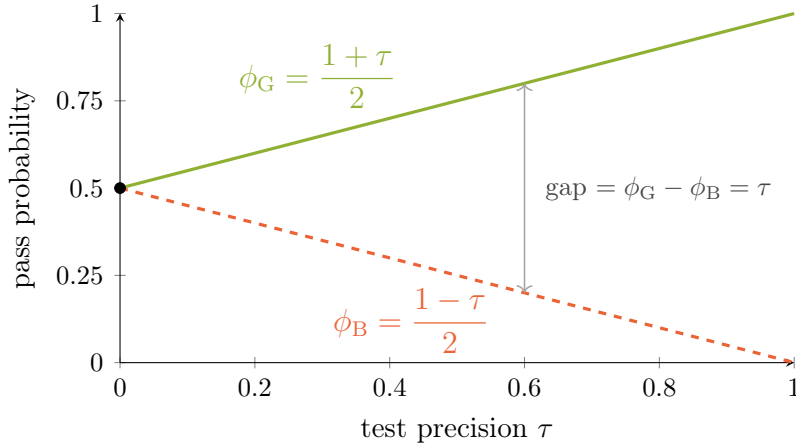


Figure 1: The two pass probabilities as functions of precision τ : a green seller passes with probability $\phi_G = \frac{1 + \tau}{2}$ (rising, green) and a brown seller with probability $\phi_B = \frac{1 - \tau}{2}$ (falling, red). They coincide at $\frac{1}{2}$ when $\tau = 0$ (uninformative) and separate to 1 and 0 when $\tau = 1$ (perfect). The vertical gap between them is the test's informativeness, $\phi_G - \phi_B = \tau$.

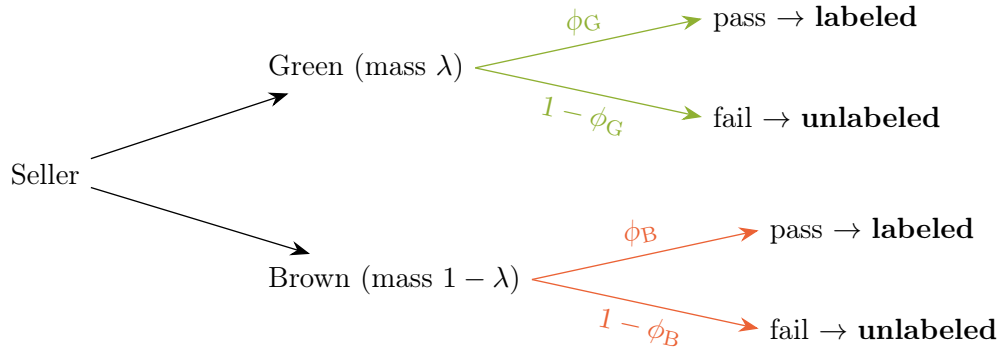


Figure 2: The four flows through the noisy test. Green sellers pass with the higher probability ϕ_G ; brown sellers pass with the lower probability ϕ_B . Failures and non-applicants are pooled together in the *unlabeled* segment.

4. Step (c): Posterior Beliefs and Prices

Suppose (for now) that every seller applies. Because each seller brings one unit, the mass of green units in a segment relative to the total is a plain Bayes ratio, and the two posteriors are

$$\mu \equiv \Pr(G \mid \text{labeled}) = \frac{\lambda\phi_G}{\lambda\phi_G + (1-\lambda)\phi_B}, \quad (3)$$

$$\nu \equiv \Pr(G \mid \text{unlabeled}) = \frac{\lambda(1-\phi_G)}{\lambda(1-\phi_G) + (1-\lambda)(1-\phi_B)}. \quad (4)$$

Substituting the precision test (2),

$$\mu = \frac{\lambda(1+\tau)}{\lambda(1+\tau) + (1-\lambda)(1-\tau)}, \quad \nu = \frac{\lambda(1-\tau)}{\lambda(1-\tau) + (1-\lambda)(1+\tau)}. \quad (5)$$

At $\lambda = \frac{1}{2}$ these collapse to $\mu = \frac{1+\tau}{2}$ and $\nu = \frac{1-\tau}{2}$, the pass probabilities themselves.

Prices equal posteriors. Within each segment competition sets the price to the expected willingness to pay, which under $v_G = 1$ and $v_B = 0$ is exactly the posterior:

$$P_c = \mu, \quad P_u = \nu, \quad P_0 = \lambda. \quad (6)$$

Sorting. Any precision $\tau > 0$ makes $\phi_G > \phi_B$, which translates into $\mu > \lambda > \nu$: an informative test raises the green share (and hence the price) in the labeled segment above the population share and lowers it in the unlabeled segment. Using $\phi_G - \phi_B = \tau$,

$$\mu - \lambda = \frac{\lambda\phi_G - \lambda[\lambda\phi_G + (1-\lambda)\phi_B]}{\lambda\phi_G + (1-\lambda)\phi_B} = \frac{\lambda(1-\lambda)\tau}{\lambda\phi_G + (1-\lambda)\phi_B} > 0,$$

and $\lambda - \nu > 0$ by an analogous computation.

Result 1 (sorting). The label puts the two segments on opposite sides of the prior, $\nu < \lambda < \mu$, with a gap proportional to the precision τ and to the mixing factor $\lambda(1-\lambda)$.

5. Step (d): Application Payoffs and Ordered Incentives

A seller of type $k = G, B$ that applies is labeled with probability ϕ_k , earning the certified price $P_c = \mu$, and unlabeled with probability $1 - \phi_k$, earning $P_u = \nu$; it pays the fee A either

way. If it does not apply it is unlabeled for sure and earns ν . The gain from applying is therefore

$$W_k \equiv [\phi_k \mu + (1 - \phi_k) \nu - A] - \nu = \phi_k (\mu - \nu) - A, \quad (7)$$

which is *linear* in the price gap $\mu - \nu$. Subtracting the two gains, and using $\phi_G - \phi_B = \tau$,

$$W_G - W_B = (\phi_G - \phi_B)(\mu - \nu) = \tau (\mu - \nu) > 0. \quad (8)$$

Ordered incentives. Green sellers have a strictly stronger incentive to seek certification than brown sellers, because their higher pass probability multiplies the same price gap into a larger gain. As a corollary, no equilibrium can have some brown sellers applying while no green sellers do; if brown sellers are in, greens are too.

6. Step (e): The Equilibrium Classes

Depending on the size of the fee A , the equilibrium falls into one of four regions, corresponding to three kinds of Perfect Bayesian equilibrium (PBE): a *separating* PBE, a *partially separating* PBE, and a *pooling* PBE that appears in two forms (both sellers apply, or no seller applies). The cutoffs come out in closed form.

Pooling PBE (both sellers apply). Both types apply, so the market cannot tell green from brown among applicants. This requires $W_B \geq 0$, i.e.,

$$A \leq \phi_B (\mu - \nu) \equiv \bar{A}_{\text{pool}}, \quad (9)$$

with μ, ν the pooling posteriors (3)–(4).

Separating PBE (green apply, brown do not). Since only greens apply, every labeled unit is green, so $\hat{\mu} = 1$ and $\hat{P}_c = 1$. The unlabeled posterior contains greens that failed plus all browns:

$$\hat{\nu} = \frac{\lambda(1 - \phi_G)}{\lambda(1 - \phi_G) + (1 - \lambda)}, \quad \hat{P}_u = \hat{\nu}. \quad (10)$$

Because the certified price is now 1, the price gap is $1 - \hat{\nu}$, and the separating range on the fee is

$$\underbrace{\phi_B (1 - \hat{\nu})}_{\text{brown IC: do not apply}} \leq A \leq \underbrace{\phi_G (1 - \hat{\nu})}_{\text{green IC: do apply}}. \quad (11)$$

Since $\phi_G > \phi_B$, this interval has positive length.

Partially separating PBE (green apply, brown mix). For fees in the gap

$$\bar{A}_{\text{pool}} < A < \phi_B (1 - \hat{\nu}), \quad (12)$$

brown sellers mix so as to be indifferent between applying and not applying. The mixing probability $\lambda_B \in (0, 1)$ solves the indifference condition

$$\phi_B (\mu^* - \nu^*) = A, \quad (13)$$

where the labeled and unlabeled posteriors, when the green seller applies for sure and the brown seller applies with probability λ_B , are

$$\mu^* = \frac{\lambda \phi_G}{\lambda \phi_G + (1 - \lambda) \lambda_B \phi_B}, \quad (14)$$

$$\nu^* = \frac{\lambda(1 - \phi_G)}{\lambda(1 - \phi_G) + (1 - \lambda)(1 - \lambda_B \phi_B)}. \quad (15)$$

These reduce to the pooling posteriors (3)–(4) when $\lambda_B = 1$ and to the separating posteriors ($\mu^* = 1$, $\nu^* = \hat{\nu}$) when $\lambda_B = 0$. This regime deserves a closer look, since the mixing probability is what makes the label only partly trustworthy.

The mixing probability lies strictly between 0 and 1. Write the brown seller's gain from applying as a function of the mixing probability,

$$g(\lambda_B) \equiv \phi_B (\mu^* - \nu^*), \quad (16)$$

so the indifference condition (13) is $g(\lambda_B) = A$. As more brown sellers apply, the labeled pool is diluted and the unlabeled pool is cleansed: in (14) the numerator of μ^* is fixed while its denominator grows, so μ^* *falls* in λ_B ; in (15) the term $(1 - \lambda_B \phi_B)$ shrinks, so ν^* *rises*. Both moves shrink the gap $\mu^* - \nu^*$, so g is strictly decreasing, from

$$g(0) = \phi_B (1 - \hat{\nu}) = \underline{A}_{\text{sep}} \quad \text{down to} \quad g(1) = \phi_B (\mu - \nu) = \bar{A}_{\text{pool}}.$$

A partially separating fee lies strictly between these two cutoffs, $\bar{A}_{\text{pool}} < A < \underline{A}_{\text{sep}}$, so by continuity and strict monotonicity the equation $g(\lambda_B) = A$ has a *unique* solution $\lambda_B \in (0, 1)$. Outside the gap there is no interior solution: for $A \leq \bar{A}_{\text{pool}}$ the brown seller strictly prefers to apply ($\lambda_B = 1$, pooling), and for $A \geq \underline{A}_{\text{sep}}$ it strictly prefers not to ($\lambda_B = 0$, separation).

Comparative static in the fee. Differentiating $g(\lambda_B) = A$ and using $g' < 0$,

$$\frac{\partial \lambda_B}{\partial A} = \frac{1}{g'(\lambda_B)} < 0, \quad (17)$$

so a higher fee lowers the fraction of brown sellers that apply. The intuition is a dilution feedback: a higher fee makes the label less attractive to a brown seller, so fewer apply; but fewer brown applicants make the label *cleaner* (a higher μ^* and a wider price gap), which pushes the gain from applying back up until the marginal brown seller is again indifferent. At $\lambda = \frac{1}{2}$ and $\tau = \frac{4}{5}$, where (14)–(15) simplify to $\mu^* = \frac{9}{9 + \lambda_B}$ and $\nu^* = \frac{1}{11 - \lambda_B}$, the solution runs monotonically from $\lambda_B = 1$ at $A = \frac{2}{25}$ to $\lambda_B = 0$ at $A = \frac{1}{11}$, as Table 3 reports.

Fee A	Brown application probability λ_B
$\frac{2}{25} = 0.080$	1 (pooling boundary)
0.082	0.80
0.084	0.61
0.086	0.43
0.088	0.25
$\frac{1}{11} \approx 0.091$	0 (separating boundary)

Table 3: The brown seller’s equilibrium application probability λ_B across the partially separating gap, at $\lambda = \frac{1}{2}$ and $\tau = \frac{4}{5}$. As the fee rises from $\frac{2}{25}$ to $\frac{1}{11}$, λ_B falls from 1 to 0.

Comparative static in precision. The effect of τ on λ_B is not monotone, because a more precise test pulls in two directions. It lowers the brown pass probability $\phi_B = \frac{1 - \tau}{2}$, which discourages brown application, but it widens the price gap $\mu^* - \nu^*$, which encourages it. Either force can dominate, so λ_B can rise or fall with τ , the same non-monotonicity Mason (2011) reports. The gap itself, $(\bar{A}_{\text{pool}}(\tau), \underline{A}_{\text{sep}}(\tau))$, also shifts and widens with τ , so a fixed fee can move out of the partially separating region altogether as the test becomes more precise.

What this means for the green seller. The green seller applies for sure. At the brown seller’s indifference, $\mu^* - \nu^* = \frac{A}{\phi_B}$, so the green seller’s net gain from applying is

$$\phi_G(\mu^* - \nu^*) - A = A \frac{\phi_G - \phi_B}{\phi_B} = \frac{2A\tau}{1 - \tau} > 0, \quad (18)$$

strictly positive and increasing in both the fee and the precision. The brown seller’s mixing

is exactly what pins the label's value: it leaves the brown seller indifferent while the green seller, passing the test more often, still strictly profits. A more precise test widens this wedge, so the green seller's advantage from certification grows with τ even though the brown seller is held to zero rents.

Pooling PBE (no seller applies). When the fee clears the upper separating cutoff, $A > \phi_G(1 - \hat{\nu})$, not even a green seller applies. Both types pool on not applying, and the market reverts to the no-information benchmark of Step (a).

The map from test fee to equilibrium. As A rises from zero, the equilibrium moves through four regions, in this order:

- the pooling PBE in which both sellers apply;
- the partially separating PBE;
- the separating PBE;
- the pooling PBE in which no seller applies.

The cutoffs (9), (11), and (12) are ordered strictly, because $\phi_G > \phi_B$ and the pooling posterior μ is strictly less than the separating posterior $\hat{\mu} = 1$.

The equilibrium map. At $\lambda = \frac{1}{2}$ the pooling posteriors reduce to $\mu = \frac{1 + \tau}{2}$ and $\nu = \frac{1 - \tau}{2}$, so $\mu - \nu = \tau$ and $\hat{\nu} = \frac{1 - \tau}{3 - \tau}$. The three cutoffs of (9), (11), and (12) then become

$$\bar{A}_{\text{pool}}(\tau) = \frac{\tau(1 - \tau)}{2}, \quad \underline{A}_{\text{sep}}(\tau) = \frac{1 - \tau}{3 - \tau}, \quad \bar{A}_{\text{sep}}(\tau) = \frac{1 + \tau}{3 - \tau}. \quad (19)$$

These are the same cutoffs that Mason (2011) obtains. Plotting the three curves in the (τ, A) plane carves it into the four regions of Figure 3: the shaded separating region between $\underline{A}_{\text{sep}}$ and $\bar{A}_{\text{sep}}$, the pooling region (both apply) below $\bar{A}_{\text{pool}}$, the partially separating wedge between them, and the pooling region (no seller applies) above $\bar{A}_{\text{sep}}$. The vertical slice at $\tau = \frac{4}{5}$ gives the numerical example of Step (f): $\bar{A}_{\text{pool}} = \frac{2}{25}$, $\underline{A}_{\text{sep}} = \frac{1}{11}$, and $\bar{A}_{\text{sep}} = \frac{9}{11}$.

7. Step (f): A Numerical Example

Take

$$\lambda = \frac{1}{2}, \quad \tau = \frac{4}{5}, \quad A = \frac{1}{20},$$

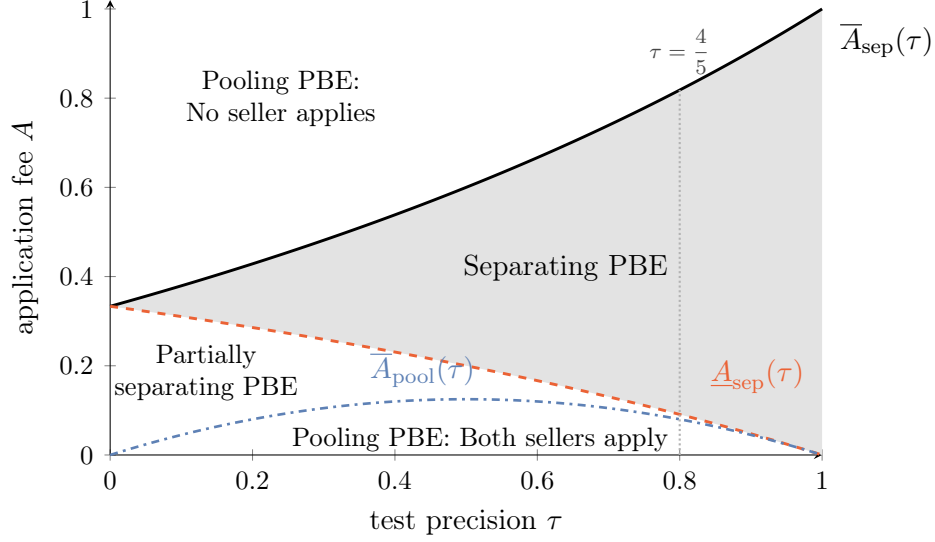


Figure 3: The equilibrium map in the (τ, A) plane, at $\lambda = \frac{1}{2}$. The three cutoffs of (19) split the plane into the four regions of Step (e): a separating PBE (only green apply, shaded), a partially separating PBE (brown randomizes), and the pooling PBE in its two forms (both sellers apply at low fees, no seller applies at high fees). The separating region widens with precision; the both-apply pooling region is a thin sliver that is largest at intermediate precision. The dotted vertical slice at $\tau = \frac{4}{5}$ is the numerical example of Step (f).

so from the precision test (2) the pass probabilities are $\phi_G = \frac{9}{10}$ and $\phi_B = \frac{1}{10}$ (a type I and a type II error rate of $\frac{1}{10}$ each).

Prices. The no-label price is $P_0 = \lambda = \frac{1}{2}$. At $\lambda = \frac{1}{2}$ the posteriors equal the pass probabilities, so

$$\mu = \frac{1 + \tau}{2} = \frac{9}{10}, \quad \nu = \frac{1 - \tau}{2} = \frac{1}{10}, \quad \mu - \nu = \frac{4}{5}.$$

Prices are the posteriors: $P_c = \frac{9}{10}$ and $P_u = \frac{1}{10}$, straddling $P_0 = \frac{1}{2}$.

Which equilibrium? The pooling cutoff (9) is

$$\bar{A}_{\text{pool}} = \phi_B(\mu - \nu) = \frac{1}{10} \cdot \frac{4}{5} = \frac{2}{25} = 0.08.$$

Since $A = \frac{1}{20} = 0.05 < 0.08 = \bar{A}_{\text{pool}}$, we are in the pooling PBE: both types apply.

Application gains. From (7), $W_k = \phi_k(\mu - \nu) - A = \phi_k \cdot \frac{4}{5} - \frac{1}{20}$:

$$W_G = \frac{9}{10} \cdot \frac{4}{5} - \frac{1}{20} = \frac{36}{50} - \frac{1}{20} = \frac{72 - 5}{100} = \frac{67}{100}, \quad W_B = \frac{1}{10} \cdot \frac{4}{5} - \frac{1}{20} = \frac{8 - 5}{100} = \frac{3}{100}.$$

Both are strictly positive, and $W_G > W_B$, as required. Raising the fee above $\frac{2}{25}$ would push the market into the partially separating region of Table 3, and above $\frac{1}{11}$ into a separating PBE.

8. Step (g): What the Toy Model Keeps and Drops

Every one of Mason's positive conclusions about the label survives the normalizations:

1. **Bayesian sorting**, $\nu < \lambda < \mu$, with a gap proportional to the precision τ (Result 1).
2. **Ordered incentives**, $W_G > W_B$, so brown sellers never apply unless green sellers do (Step d).
3. **Four equilibrium regions** indexed by the fee: a pooling PBE in which both apply (low A), a partially separating PBE (intermediate A), a separating PBE (high A), and a pooling PBE in which no seller applies (prohibitive A), with the same cutoffs $\underline{A}_{\text{sep}} = \frac{1 - \tau}{3 - \tau}$ and $\bar{A}_{\text{sep}} = \frac{1 + \tau}{3 - \tau}$ that Mason obtains, shown in Figure 3.
4. **The mixing probability is interior and well-behaved**: it is unique, decreasing in the fee (Table 3), and non-monotone in precision, and it leaves the green seller with a strictly positive gain $\frac{2A\tau}{1 - \tau}$ that grows with τ (Step e).

What the toy model drops, relative to Mason (2011) and to the certifier-design literature, is threefold: (i) sellers here bring a single unit, so there is no choice of output and no producer-surplus welfare accounting; (ii) the two error rates are tied together by the symmetric test (2) at $\frac{1 - \tau}{2}$, whereas Mason lets the two pass rates move independently; and (iii) the precision τ is fixed, rather than chosen by a certifier or a regulator who trades the value of information against the cost of a more accurate test. Each can be reinstated one at a time; the payoff of the toy model is that the sorting mechanism and the four equilibrium regions come through with almost no algebra.

Reference

Mason, C. F. (2011). “Eco-Labeling and Market Equilibria with Noisy Certification Tests.” *Environmental and Resource Economics*, 48(4), pp. 537–560. <https://doi.org/10.1007/s10640-010-9402-5>

Bayesian Persuasion: Four Papers

Félix Muñoz-García*

1 Introduction

Consider one person, whom we call the Sender, who wants to change the action of another person, the Receiver. The Sender cannot pay the Receiver, threaten her, or in any way alter her payoffs. He can only choose what information she will see before she acts. Examples of this problem are everywhere:

- a prosecutor deciding which evidence to collect before a judge rules;
- a lobbyist funding a study that a politician will read;
- a firm designing the free trial that a consumer will try before buying;
- a school issuing only coarse grades so the labor market places its graduates better;
- a company choosing how much to disclose about its environmental record to a green-minded consumer.

The question we study, following Kamenica and Gentzkow (2011), is whether a Sender can benefit from this kind of persuasion even when the Receiver is a rational Bayesian who understands that the Sender designed the information with his own goal in mind. The answer, perhaps surprisingly, is yes. This handout works through the simplest example in the paper, the prosecutor and the judge, and shows every step needed to arrive at the same number the authors report: a rational judge who knows that 70 percent of defendants are innocent can nonetheless be led to convict 60 percent of them.

We proceed slowly, and each paper builds on the one before it. Section 2 solves the baseline model of Kamenica and Gentzkow (2011): with a free investigation the prosecutor convicts a rational judge $2\mu_0$ of the time, and the answer is the value of the concave closure of his payoff evaluated at the prior. Section 3 makes the investigation costly (Gentzkow and Kamenica 2014)

*Professor. School of Economic Sciences, Washington State University, Pullman, WA. This material was prepared for the EconS 594 course, Industrial Organization.

and recovers the baseline when the cost is zero. Section 4 lets the prosecutor and the judge hold different priors (Alonso and Cámara 2016) and recovers the baseline when they agree. Section 5 sets a defense attorney beside the prosecutor (Gentzkow and Kamenica 2017) and recovers the baseline when one side stays silent. Sections 6 and 7 tour the rest of the literature and point to open questions. One thread runs through all four papers: write the Sender’s payoff as a function of the Receiver’s belief, and the value of the concave closure of that payoff, evaluated at the prior, is the best the Sender can do. Each extension only reshapes that payoff, and each contains the baseline as the value of a single parameter. We keep the prior μ_0 generic throughout, so that every result is a formula in μ_0 , and we set $\mu_0 = 0.3$ only for the tables and figures.

2 Costless Persuasion (Kamenica and Gentzkow, 2011)

2.1 The model

States. There are two states of the world, $\omega \in \{G, I\}$, where G means the defendant is guilty and I means innocent. The prosecutor (Sender) and the judge (Receiver) share a common prior that the defendant is guilty. We call this prior μ_0 and keep it as a free parameter, so that every result below is a formula in μ_0 ; our running numerical illustration sets $\mu_0 = 0.3$. Because there are only two states, any belief is a single number, the probability it places on guilt, and we write $\mu \equiv \Pr(\text{guilty})$, so the prior places

$$\mu_0(G) = \mu_0 \quad \text{and} \quad \mu_0(I) = 1 - \mu_0.$$

We assume $\mu_0 < \frac{1}{2}$: a judge with no information would acquit, which is the case in which persuasion has any work to do. The parameter μ_0 is the one free parameter we carry through the whole handout; each later part adds one more parameter of its own, but the prior stays generic throughout.

Actions. The judge chooses one of two actions, $a \in \{\text{convict}, \text{acquit}\}$.

Payoffs. The judge wants to match the verdict to the truth. She earns 1 from the just action, convicting the guilty or acquitting the innocent, and 0 from the unjust action:

$$u(\text{convict}, G) = 1, \quad u(\text{acquit}, I) = 1, \quad u(\text{convict}, I) = 0, \quad u(\text{acquit}, G) = 0.$$

The prosecutor wants a conviction regardless of the truth. He earns 1 whenever the judge convicts and 0 whenever she acquits:

$$v(\text{convict}, \omega) = 1 \quad \text{and} \quad v(\text{acquit}, \omega) = 0 \quad \text{for both } \omega.$$

Notice that the prosecutor’s payoff depends only on the judge’s *action*, not on the state. This will

simplify the geometric picture we draw later.

The instrument. The prosecutor chooses an *investigation*: a signal π that generates a message for the judge, with the message drawn from state-dependent distributions $\pi(\cdot | G)$ and $\pi(\cdot | I)$. He commits to the investigation before the evidence is in, and reports its outcome honestly; he cannot rewrite the rules once he sees what turns up. The judge observes the design of the investigation and its realized message, updates her belief by Bayes' rule, and then convicts or acquits.

When does the judge convict? At any belief μ , the judge convicts if the expected payoff from convicting is at least that of acquitting:

$$\underbrace{\mu \cdot 1 + (1 - \mu) \cdot 0}_{\text{convict}} \geq \underbrace{\mu \cdot 0 + (1 - \mu) \cdot 1}_{\text{acquit}} \iff \mu \geq 1 - \mu \iff \mu \geq \frac{1}{2}.$$

We record the judge's optimal action as a function of her belief,

$$\bar{a}(\mu) = \begin{cases} \text{convict} & \text{if } \mu \geq \frac{1}{2}, \\ \text{acquit} & \text{if } \mu < \frac{1}{2}, \end{cases}$$

where we assign a tie in the prosecutor's favor, consistent with the Sender-preferred equilibrium of the paper; without this convention the best conviction rate would be a value the prosecutor could approach but never quite reach.

2.2 Two benchmarks

Before designing anything clever, we compute what happens at the two extremes. These give the prosecutor a floor to improve upon.

Example 1 (No disclosure). Suppose the investigation is completely uninformative, so the judge learns nothing and keeps her prior μ_0 . Since $\mu_0 < \frac{1}{2}$, the judge acquits, and the prosecutor's payoff is

$$v_{\text{no info}} = 0.$$

We call *acquit* the *default action*, since it is what the judge does when she has no information beyond the prior. □

Example 2 (Full disclosure). Suppose the investigation reveals the state with certainty. Then the judge's posterior is either $\mu = 1$ (and she convicts) or $\mu = 0$ (and she acquits). She convicts exactly when the defendant is truly guilty, which happens with probability μ_0 . The prosecutor's expected

payoff is

$$v_{\text{full info}} = \mu_0 \cdot 1 + (1 - \mu_0) \cdot 0 = \mu_0 \quad (= 0.3 \text{ at } \mu_0 = 0.3).$$

Full disclosure beats silence, raising the conviction rate from 0 to μ_0 . The question is whether the prosecutor can do better still. $\square$

Intuitively, revealing everything is a crude instrument. It hands the judge a conviction only on the guilty, and it throws away the possibility of convicting on the strength of merely suggestive evidence. We now show that a partial investigation does better than both benchmarks.

2.3 The one idea that makes the problem tractable

The prosecutor has infinitely many investigations to choose among, differing in how many messages they send and how each message correlates with the state. Searching over all of them directly looks hopeless. The insight of Kamenica and Gentzkow is that we do not need to. What the prosecutor ultimately cares about is the judge's *belief* when she acts, because the belief is what determines the verdict. So we can search over the beliefs an investigation can produce, rather than over the investigations themselves.

Each message the judge might receive leaves her with some posterior belief μ . An investigation therefore induces a *distribution of posteriors*: a list of the beliefs that can arise, together with the probability of each. The question becomes: which distributions of posteriors are attainable?

The answer is that a distribution of posteriors is attainable if and only if it is **Bayes plausible**, meaning that the beliefs average back to the prior:

$$\mathbb{E}[\mu] = \mu_0.$$

The requirement is nothing more than the law of iterated expectations. Whatever the judge might come to believe after seeing a message, her beliefs, weighted by how likely each message is, must return to where she started. No investigation can, on average, make the judge more convinced of guilt than the prior warrants; it can only redistribute her conviction across messages.

This is the whole content of Proposition 1 in the paper. It lets us restate the prosecutor's problem in a clean form:

$$\max_{\tau \text{ Bayes plausible}} \mathbb{E}_{\tau}[\hat{v}(\mu)],$$

where τ is the distribution of posteriors and $\hat{v}(\mu)$ denotes the prosecutor's payoff when the judge holds belief μ and acts optimally. In our example, since the prosecutor earns 1 whenever the judge

convicts,

$$\hat{v}(\mu) = v(\bar{a}(\mu)) = \begin{cases} 1 & \text{if } \mu \geq \frac{1}{2}, \\ 0 & \text{if } \mu < \frac{1}{2}. \end{cases}$$

2.4 Solving the prosecutor's problem

We now solve the problem step by step. It is convenient to work with *straightforward* investigations: those whose messages are recommendations, “convict” or “acquit,” which the judge is willing to obey. The paper shows this is without loss of generality, exactly as the revelation principle lets us restrict attention to truthful mechanisms. There are only two messages to design, so the problem reduces to two numbers. Two messages are all the prosecutor ever needs: because the judge takes only one of two actions, a signal with more realizations cannot raise his payoff.

Step 1. Name the choice variables. Let

$$\sigma_G = \pi(g \mid G) \quad \text{and} \quad \sigma_I = \pi(g \mid I)$$

be the probabilities that the investigation sends the message g (“recommend convict”) in each state. The complementary probabilities, $1 - \sigma_G$ and $1 - \sigma_I$, are the chances of the message i (“recommend acquit”).

Step 2. Write the objective. If the judge obeys, the prosecutor wins a conviction exactly when the message is g . That message arrives with total probability

$$\Pr(g) = \mu_0 \sigma_G + (1 - \mu_0) \sigma_I,$$

so the prosecutor maximizes $\mu_0 \sigma_G + (1 - \mu_0) \sigma_I$.

Step 3. Impose obedience on the conviction recommendation. For the judge to actually convict when she sees g , her posterior on guilt given g must be at least $\frac{1}{2}$. By Bayes' rule,

$$\mu_g \equiv \Pr(G \mid g) = \frac{\mu_0 \sigma_G}{\mu_0 \sigma_G + (1 - \mu_0) \sigma_I} \geq \frac{1}{2}.$$

Cross-multiplying, this obedience condition is

$$\mu_0 \sigma_G \geq (1 - \mu_0) \sigma_I.$$

Step 4. Impose obedience on the acquittal recommendation. For the judge to acquit when

she sees i , her posterior on guilt given i must be at most $\frac{1}{2}$:

$$\mu_i \equiv \Pr(G \mid i) = \frac{\mu_0(1 - \sigma_G)}{\mu_0(1 - \sigma_G) + (1 - \mu_0)(1 - \sigma_I)} \leq \frac{1}{2},$$

which simplifies to $\mu_0(1 - \sigma_G) \leq (1 - \mu_0)(1 - \sigma_I)$.

Step 5. Argue that a guilty defendant is always recommended for conviction. The objective $\mu_0 \sigma_G + (1 - \mu_0) \sigma_I$ rises in σ_G , and raising σ_G also loosens the conviction-obedience condition in Step 3, since it puts more guilt mass behind the message g . There is thus no reason to hold back: the prosecutor sets

$$\sigma_G = 1.$$

Intuitively, evidence pointing at a guilty defendant can only help the case, so the optimal investigation never recommends acquittal for someone who is in fact guilty. This is the example's version of the general rule that when the judge takes the prosecutor's least-preferred action, she does so only when she is certain of the state.

Step 6. Convict as many innocents as obedience allows. With $\sigma_G = 1$, the objective becomes $\mu_0 + (1 - \mu_0) \sigma_I$, which rises in σ_I . The prosecutor would like σ_I as large as possible, but Step 3 caps it at $\mu_0 \geq (1 - \mu_0) \sigma_I$. The constraint binds, so

$$\sigma_I = \frac{\mu_0}{1 - \mu_0}.$$

We check Step 4 is satisfied: $\mu_0(1 - 1) = 0 \leq (1 - \mu_0)(1 - \frac{\mu_0}{1 - \mu_0}) = 1 - 2\mu_0$, which is positive because $\mu_0 < \frac{1}{2}$, so the judge is indeed willing to acquit on the message i .

Step 7. Read off the optimal investigation. Translating (σ_G, σ_I) back into the language of the two messages,

$$\begin{aligned} \pi(g \mid G) &= 1, & \pi(g \mid I) &= \frac{\mu_0}{1 - \mu_0}, \\ \pi(i \mid G) &= 0, & \pi(i \mid I) &= \frac{1 - 2\mu_0}{1 - \mu_0}. \end{aligned}$$

At the running prior $\mu_0 = 0.3$ this is $(1, \frac{3}{7}, 0, \frac{4}{7})$, precisely the investigation reported in equation (1) of Kamenica and Gentzkow (2011).

Step 8. Compute the conviction rate. The prosecutor's payoff equals the probability of the message g :

$$v^* = \Pr(g) = \mu_0(1) + (1 - \mu_0) \frac{\mu_0}{1 - \mu_0} = \mu_0 + \mu_0 = 2\mu_0.$$

The optimal investigation therefore convicts with probability $2\mu_0$, exactly twice the full-disclosure

rate μ_0 , whenever $\mu_0 < \frac{1}{2}$. At $\mu_0 = 0.3$ this is 0.6: a judge who knows that 70 percent of defendants are innocent convicts 60 percent of them, while fully aware that the investigation was built to maximize convictions.

Why the judge goes along

The prosecutor and the judge are not pure adversaries: both want to convict the guilty, so some messages serve them at once. And because the prosecutor commits to the investigation in advance and the judge knows its design, she can trust what a message implies rather than second-guess his motive. With that in place, it is worth reading the two posteriors the investigation produces, because they show there is no trick. On the message i , the posterior is

$$\mu_i = \frac{\mu_0 \cdot 0}{\mu_0 \cdot 0 + (1 - \mu_0) \cdot \frac{1-2\mu_0}{1-\mu_0}} = 0,$$

so the judge is *certain* the defendant is innocent and acquits without hesitation. On the message g , the posterior is

$$\mu_g = \frac{\mu_0 \cdot 1}{\mu_0 \cdot 1 + (1 - \mu_0) \cdot \frac{\mu_0}{1-\mu_0}} = \frac{\mu_0}{\mu_0 + \mu_0} = \frac{1}{2},$$

so the judge is exactly indifferent between convicting and acquitting, and, breaking the tie in the prosecutor's favor, she convicts. The prosecutor has engineered the weakest belief consistent with a conviction: just enough to tip the judge over, and not a shade more.

We can also confirm Bayes plausibility directly. The message g arrives with probability $2\mu_0$ and leaves belief $\frac{1}{2}$; the message i arrives with probability $1 - 2\mu_0$ and leaves belief 0. Averaging,

$$2\mu_0 \cdot \frac{1}{2} + (1 - 2\mu_0) \cdot 0 = \mu_0,$$

so the two posteriors do return to the prior, as they must.

2.5 Graphical representation

The problem has a picture that gives the answer at once, and it generalizes far beyond this example. Plot the prosecutor's payoff $\hat{v}(\mu)$ against the judge's belief μ . As computed above, $\hat{v}$ is a step function: it is 0 for every belief below $\frac{1}{2}$, where the judge acquits, and jumps to 1 once the belief reaches $\frac{1}{2}$, where she convicts.

Now recall the prosecutor's problem: choose beliefs that average to the prior μ_0 , so as to maximize the average height of $\hat{v}$. Mixing two beliefs traces the straight line joining the two points

on the graph of $\hat{v}$. The best the prosecutor can do at any prior is therefore given by the **concave closure** of $\hat{v}$, written V : the lowest concave curve that lies weakly above $\hat{v}$ everywhere. Figure 1 plots $\hat{v}$, its concave closure V , and the construction at the prior.

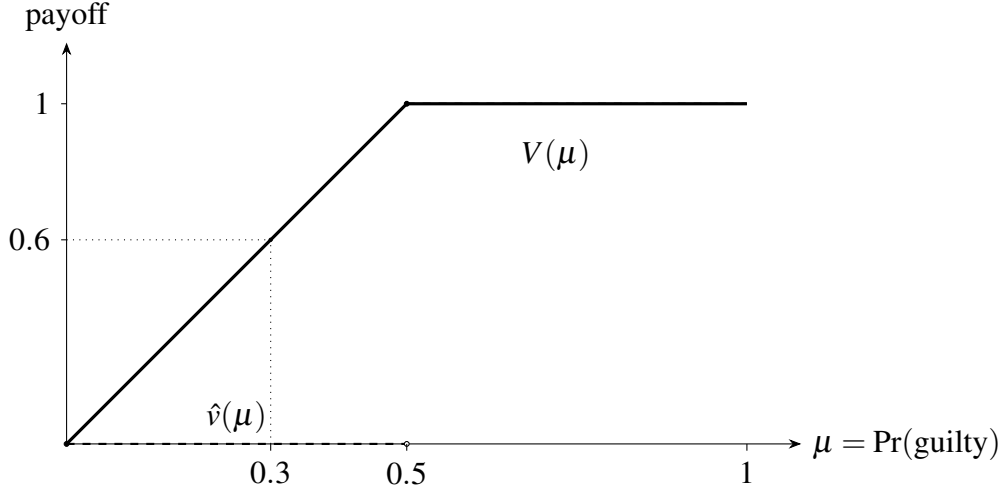


Figure 1: The step payoff $\hat{v}$ (dashed), its concave closure V (solid), and the value $V(0.3) = 0.6$ at the prior.

Reading the figure recovers everything from the algebra above, and faster. The concave closure is the line $V(\mu) = 2\mu$ for $\mu \leq \frac{1}{2}$ and $V(\mu) = 1$ for $\mu > \frac{1}{2}$. Its value at the prior is

$$V(\mu_0) = 2\mu_0 \quad (= 0.6 \text{ at } \mu_0 = 0.3),$$

which is the optimal conviction rate. The two points the closure touches, $\mu = 0$ and $\mu = \frac{1}{2}$, are exactly the two posteriors the optimal investigation induces. And the weights that place their average at μ_0 are $1 - 2\mu_0$ on $\mu = 0$ and $2\mu_0$ on $\mu = \frac{1}{2}$, the probabilities of the two messages. The gap between the curves at the prior,

$$V(\mu_0) - \hat{v}(\mu_0) = 2\mu_0 - 0 = 2\mu_0,$$

is the prosecutor's gain from persuasion. Wherever this gap is positive, the Sender benefits; wherever $\hat{v}$ already equals its own concave closure, no investigation helps.

2.6 Main points from the above example

The example teaches lessons that recur throughout what follows.

- Persuasion works when the payoff is not concave. The prosecutor gains precisely because $\hat{v}$ lies below its concave closure at the prior. If the prosecutor's payoff were concave in the judge's

belief, no spreading of beliefs could raise its average, and silence would be optimal. If it were convex, full disclosure would be optimal. The interesting case, and the realistic one, is the step in between.

- The Sender loses nothing on the downside. Providing information sometimes moves the judge toward guilt and sometimes toward innocence. The move toward guilt earns a conviction; the move toward innocence costs the prosecutor nothing, because the judge was already acquitting by default. Persuasion benefits the Sender whenever the Receiver does not already take the Sender's preferred action, and her action stays fixed over a range of beliefs around the prior.
- The Sender induces the worst belief consistent with each action. When the judge acquits, she is made certain of innocence; when she convicts, she is made just indifferent. The prosecutor never leaves the judge more convinced than the action requires, because any leftover conviction could be spent buying additional convictions elsewhere. This is the content of Propositions 4 and 5 in the paper, and the reader can see both at work in the two posteriors computed above.

2.7 An environmental application: green signaling and greenwashing

The courtroom is one dressing of a structure that recurs, and a second reading will run alongside us through the rest of the handout. Read the Sender as a firm and the Receiver as a green-minded consumer, or a regulator. The state is whether the firm's product is genuinely clean or dirty; the prior μ_0 is now the share of clean products, and the consumer buys, or the regulator approves, only when she is convinced enough that the product is clean, say when her belief that it is clean is at least $\frac{1}{2}$. The firm, like the prosecutor, wants the favorable action whatever the truth, and it chooses what to disclose about its environmental performance: which tests to run, which metrics to report, how to design a sustainability label.

The baseline result then reads as a theory of greenwashing. A firm sure to be exposed as dirty says nothing; a firm with a genuinely clean case reveals it; and in between the firm designs a label that raises a skeptical consumer's belief to exactly the buying threshold and not a shade higher. The optimal green signal is therefore a partial, noisy label that clears the consumer at rate $2\mu_0$, twice the share of products that are actually clean, precisely the "convict 60 percent while only 30 percent are guilty" logic of the courtroom. The three extensions carry over intact: costly testing (Section 3) shrinks how much the firm discloses; a firm more optimistic about its own product than its audience is (Section 4) values the same label more without changing it; and a watchdog or a rival that can also disclose (Section 5) pushes the market toward full revelation of the firm's true type. This environmental reading is the subject of a large literature on certification and greenwashing (Lyon and Maxwell 2011).

2.8 A numerical summary

Table 1 collects the three investigations, no disclosure, full disclosure, and the optimal partial investigation, at the running prior $\mu_0 = 0.3$. Reading down the conviction column shows the prosecutor's problem in one glance: silence convicts no one, full disclosure convicts only the truly guilty, and the optimal partial investigation doubles the full-disclosure rate. The judge's own welfare, in the last column, is highest under full disclosure, where she always matches the verdict to the truth, and falls back to her uninformed level of 0.7 under the optimal investigation, since the prosecutor leaves her exactly indifferent whenever she convicts.

Investigation	Conviction rate	Prosecutor payoff	Judge welfare
No disclosure	0.00	0.00	0.70
Full disclosure	0.30	0.30	1.00
Optimal (partial)	0.60	0.60	0.70

Table 1: The three investigations at the prior $\mu_0 = 0.3$.

How much the prosecutor gains depends on the prior. Since the value of persuasion is $V(\mu_0) = \min(2\mu_0, 1)$ while full disclosure convicts at rate μ_0 , the gain is μ_0 for every prior below $\frac{1}{2}$ and vanishes once $\mu_0 \geq \frac{1}{2}$, where the judge convicts by default. Figure 2 plots the two rates.

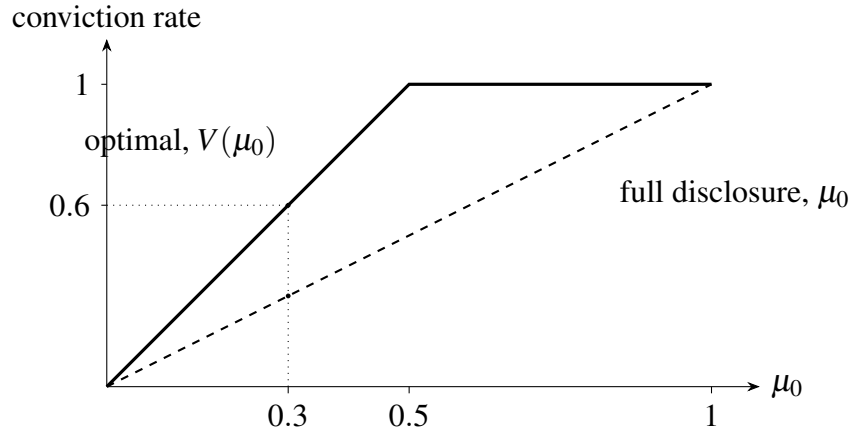


Figure 2: Conviction rate under the optimal investigation, $V(\mu_0) = \min(2\mu_0, 1)$, and under full disclosure, μ_0 .

3 Costly Persuasion (Gentzkow and Kamenica, 2014)

3.1 From free information to costly information

The investigation in Section 2 was free. A prosecutor, though, does not gather evidence at no charge: forensic tests, expert witnesses, depositions, and subpoenas all consume time and money. It is natural to ask whether the elegant concavification of Section 2 survives once the Sender must pay for the information he generates.

At first sight the answer looks like no. The whole method rested on writing the Sender's payoff as a function of the Receiver's posterior alone, $\hat{v}(\mu)$, so that the problem reduced to averaging $\hat{v}$ over Bayes-plausible beliefs. Once signals are costly, this breaks: two investigations can leave the judge at the very same posterior μ while costing the prosecutor different amounts, one being a cheap, coarse investigation and the other an expensive, finely tuned one. The Sender's payoff at a belief then depends on *how* the belief was reached, not on the belief alone, and the concavification picture no longer applies.

Gentzkow and Kamenica (2014), henceforth GK, identify a family of cost functions for which the difficulty disappears. For this family the cost can be absorbed into the value function, producing a *cost-adjusted* payoff that again depends on the posterior alone, so the same picture goes through. And when the cost vanishes, the model returns to Kamenica and Gentzkow (2011), henceforth KG. The costless model of Section 2 is thus the special case of a one-parameter family, recovered when that parameter equals zero.

3.2 The model with costs

The states, actions, payoffs, and prior are exactly those of Section 2: two states $\omega \in \{G, I\}$ with the same generic prior $\mu_0 = \Pr(\text{guilty}) < \frac{1}{2}$ (running value 0.3); the judge convicts if and only if her belief satisfies $\mu \geq \frac{1}{2}$; and the prosecutor earns 1 from a conviction and 0 from an acquittal. The only change is that running an investigation π now costs the prosecutor $c(\pi)$, so his payoff becomes

$$v(a, \omega) - c(\pi).$$

GK's central assumption is that the cost of an investigation is proportional to the *expected reduction in uncertainty* it produces. Let $H(\mu)$ be a measure of the uncertainty in a belief μ : a concave function that is largest when the belief is most spread out and zero when the belief is degenerate. Then the assumption is

$$c(\pi) = k \mathbb{E}[H(\mu_{\text{ref}}) - H(\mu_s)],$$

where μ_{ref} is a fixed reference belief, μ_s the posterior after message s , and $k \geq 0$ a cost parameter. Concavity of H guarantees, by Jensen's inequality, that information reduces uncertainty on average, so the cost is never negative, and it is zero for an uninformative investigation.

Two measures of uncertainty are standard. The first is Shannon entropy, $H(\mu) = -\sum_{\omega} \mu(\omega) \ln \mu(\omega)$, which underlies the rational-inattention literature. The second is residual variance, $H(\mu) = \sum_{\omega} \mu(\omega) (1 - \mu(\omega))$, which in our two-state model is

$$H(\mu) = 2\mu(1 - \mu).$$

We adopt residual variance, and we place the reference belief at the prior, $\mu_{\text{ref}} = \mu_0$, so that the cost measures the uncertainty an investigation removes relative to what the judge knew at the outset. Both are choices GK explicitly allow. This pair keeps the algebra in closed form and lets us see every step; GK's own headline example instead uses entropy with a uniform reference belief, and we reproduce its numbers in the numerical illustration below. With the reference at the prior the cost simplifies to

$$c(\pi) = k(H(\mu_0) - \mathbb{E}_{\tau}[H(\mu)]),$$

where τ is, as before, the distribution of posteriors the investigation induces.

3.3 The unifying result: concavify a cost-adjusted payoff

The Sender's problem is again to choose a Bayes-plausible distribution of posteriors, but now net of cost:

$$\begin{aligned} \max_{\tau \text{ Bayes plausible}} \mathbb{E}_{\tau}[\hat{v}(\mu)] - c(\pi) &= \max_{\tau} \mathbb{E}_{\tau}[\hat{v}(\mu)] - k(H(\mu_0) - \mathbb{E}_{\tau}[H(\mu)]) \\ &= \max_{\tau} \mathbb{E}_{\tau}[\hat{v}(\mu) + kH(\mu)] - kH(\mu_0). \end{aligned}$$

The last line is the key. The cost has been absorbed into a single *cost-adjusted value function*

$$\hat{v}_k(\mu) \equiv \hat{v}(\mu) + kH(\mu),$$

which depends on the posterior alone, so the concavification method of Section 2 applies without change. Writing V_k for the concave closure of $\hat{v}_k$, the Sender's value is

$$V_k(\mu_0) - kH(\mu_0),$$

and the Sender benefits from persuasion if and only if $V_k(\mu_0) > \hat{v}_k(\mu_0)$. This is GK's Proposition 1, specialized to our cost.

The nesting. Set $k = 0$. Then $\hat{v}_k = \hat{v}$, the concave closure V_k coincides with V from Section 2, and the constant $kH(\mu_0)$ vanishes. The Sender's value is $V(\mu_0) = 2\mu_0$, exactly the costless result of Section 2. Kamenica and Gentzkow (2011) is the $k = 0$ member of the family.

Intuitively, adding the concave term $kH(\mu)$ to $\hat{v}$ makes the payoff more concave. Since persuasion feeds on the gap between a function and its concave closure, and a more concave function leaves a smaller gap, information costs shrink the room for persuasion. The subtracted constant $kH(\mu_0)$ is what the prosecutor forgoes simply by moving the judge's belief away from the prior at all.

3.4 The toy model, worked

Step 1. Write the cost-adjusted payoff. With $\hat{v}(\mu)$ the step function of Section 2 and $H(\mu) = 2\mu(1 - \mu)$,

$$\hat{v}_k(\mu) = \begin{cases} 2k\mu(1 - \mu) & \text{if } \mu < \frac{1}{2}, \\ 1 + 2k\mu(1 - \mu) & \text{if } \mu \geq \frac{1}{2}. \end{cases}$$

Each piece is a downward parabola, and the two pieces are separated by an upward jump of size 1 at the judge's indifference belief $\mu = \frac{1}{2}$.

Step 2. The conviction belief stays at the threshold. To win a conviction the prosecutor must leave the judge with a posterior $\mu_h \geq \frac{1}{2}$. Raising μ_h above $\frac{1}{2}$ buys no extra conviction, since the judge already convicts, and it is costly, since H is falling above $\frac{1}{2}$ and a sharper belief removes more uncertainty. The prosecutor therefore leaves the judge exactly indifferent, $\mu_h = \frac{1}{2}$, just as in Section 2.

Step 3. The acquittal belief now involves a trade-off. Let $\mu_l < \frac{1}{2}$ be the belief that leads the judge to acquit. In Section 2 the prosecutor drove μ_l all the way to 0, making the judge certain of innocence. That is no longer clearly best. Lowering μ_l toward 0 still raises the conviction probability, because Bayes plausibility then allows more weight on the conviction message; but it also raises the cost, because a belief nearer certainty carries less uncertainty and the cost rewards uncertainty left in place. The prosecutor must weigh the two.

Step 4. Solve for the acquittal belief. With $\mu_h = \frac{1}{2}$ and weight p on the conviction message, Bayes plausibility reads $p \cdot \frac{1}{2} + (1 - p)\mu_l = \mu_0$, so

$$p = \frac{\mu_0 - \mu_l}{\frac{1}{2} - \mu_l}.$$

The prosecutor's net payoff, gross conviction probability p minus cost, is

$$N(\mu_l) = p - k \left(H(\mu_0) - [p \cdot H(\frac{1}{2}) + (1-p)H(\mu_l)] \right),$$

with $H(\frac{1}{2}) = \frac{1}{2}$ and $H(\mu_0) = 2\mu_0(1 - \mu_0)$. Differentiating and evaluating at $\mu_l = 0$ gives the clean expression

$$\left. \frac{dN}{d\mu_l} \right|_{\mu_l=0} = 2(1 - 2\mu_0) \left(\frac{k}{2} - 1 \right),$$

which at $\mu_0 = 0.3$ is $-0.8 + 0.4k$.

Step 5. Read off the three regimes. Because $1 - 2\mu_0 > 0$, the sign of the derivative is the sign of $\frac{k}{2} - 1$, so the threshold sits at $k = 2$ for every prior $\mu_0 < \frac{1}{2}$. This sorts the problem into three cases.

- If $k \leq 2$, the derivative is not positive, so the prosecutor keeps $\mu_l = 0$. The optimal investigation is *identical* to the costless one of Section 2: the judge is made certain of innocence on the acquittal message and indifferent on the conviction message, and the conviction rate stays at $2\mu_0$. Only the prosecutor's net payoff falls, to

$$N = 2\mu_0 - k\mu_0(1 - 2\mu_0) = \mu_0[2 - k(1 - 2\mu_0)],$$

which at $\mu_0 = 0.3$ is $0.6 - 0.12k$.

- If $2 < k < k^*(\mu_0)$, the derivative at $\mu_l = 0$ is positive, so the prosecutor raises μ_l above 0. He no longer fully clears the innocent, the acquittal message stops implying certain innocence, and the conviction rate falls below $2\mu_0$.
- If $k \geq k^*(\mu_0)$, no investigation is worthwhile, and the prosecutor gathers no evidence.

The knife-edge at $k = 2$ is where a binary verdict first makes it worth economizing on information; the upper cutoff $k^*(\mu_0)$, where persuasion stops altogether, depends on the prior and equals about 12.5 at $\mu_0 = 0.3$ (see the numerical illustration below).

Step 6. Recover the investigation at an interior cost. Return to the running prior $\mu_0 = 0.3$ and take $k = 3$, in the middle regime. Solving $N'(\mu_l) = 0$ yields $\mu_l = 0.0918$, hence $p = 0.510$. The two induced posteriors are $\mu_l = 0.0918$ with weight $1 - p = 0.490$ and $\mu_h = \frac{1}{2}$ with weight $p = 0.510$. Recovering the investigation from $\pi(s | \omega) = \mu_s(\omega) \tau(\mu_s) / \mu_0(\omega)$,

$$\begin{aligned} \pi(g | G) &= \frac{0.5 \times 0.510}{0.3} = 0.850, & \pi(g | I) &= \frac{0.5 \times 0.510}{0.7} = 0.364, \\ \pi(i | G) &= \frac{0.0918 \times 0.490}{0.3} = 0.150, & \pi(i | I) &= \frac{0.9082 \times 0.490}{0.7} = 0.636. \end{aligned}$$

The conviction rate is $\Pr(g) = 0.3(0.850) + 0.7(0.364) = 0.510$, down from 0.6. Notice that $\pi(i | G) = 0.150 > 0$: the prosecutor now recommends acquittal for some guilty defendants. This is the sharpest contrast with Section 2. When information was free, the acquittal message meant certain innocence (Proposition 4 of KG); once information is costly, buying that certainty is too expensive, so the prosecutor accepts an acquittal message that leaves the judge with lingering doubt.

3.5 What cost adds to the picture

Costly persuasion reshapes the baseline in ways worth drawing out.

- The costless model is a boundary, not a different model. Everything in Section 3 collapses to Section 2 at $k = 0$. The extension does not overturn Bayesian persuasion; it embeds it in a family and asks what happens as one moves away from the free-information boundary. This is the shape of a good extension: the earlier result sits inside the newer one at a single parameter value, and the new results describe the neighborhood around it.
- Costs shrink disclosure, and do so in stages. As k rises, the optimal investigation is first unchanged, then Blackwell less informative, then silent. The prosecutor first absorbs the cost while running the same investigation, then economizes by tolerating a noisier acquittal signal, and finally abandons persuasion. Lower costs always yield a weakly more informative investigation in this example, matching GK's comparative static.
- The Receiver captures none of the value of information. A short calculation shows the judge's expected payoff under the optimal investigation is

$$p \cdot \frac{1}{2} + (1 - p)(1 - \mu_l) = 1 - \mu_0 \quad \text{for every } k,$$

using $p \cdot \frac{1}{2} + (1 - p)\mu_l = \mu_0$. This is exactly the judge's payoff with no information at all, $\Pr(\text{innocent}) = 1 - \mu_0$ (namely 0.7 at $\mu_0 = 0.3$). Whatever the cost of investigation, the prosecutor leaves the judge no better off than her uninformed benchmark and keeps the entire value of information for himself.

3.6 A numerical summary

Table 2 traces the optimal investigation as the cost parameter k rises, under the residual-variance cost worked out above. The three regimes are visible in the columns. For $k \leq 2$ the acquittal belief stays at $\mu_l = 0$ and the conviction rate holds at 0.6, exactly the costless investigation; only the prosecutor's net payoff falls, one-for-one at rate 0.12 per unit of cost. For $2 < k < k^*$ the

acquittal belief lifts off 0 and the conviction rate slides down. At $k^* \approx 12.5$ the prosecutor stops investigating. The judge's welfare, not shown, equals 0.7 in every row.

k	μ_l	Conviction rate	Prosecutor net	Regime
0	0.000	0.600	0.600	identical to KG
1	0.000	0.600	0.480	signal = KG
2	0.000	0.600	0.360	threshold
3	0.092	0.510	0.260	interior
4	0.146	0.434	0.189	interior
6	0.211	0.307	0.094	interior
10	0.276	0.106	0.011	interior
12.5	0.300	0.000	0.000	no disclosure

Table 2: Optimal investigation as the cost k rises, residual-variance cost with reference belief $\mu_0 = 0.3$.

Figure 3 draws the case $k = 3$ the way Figure 1 drew the costless problem. The cost bends each branch of the payoff into a parabola and lifts the whole picture, but the method is unchanged: the concave closure V_3 of the cost-adjusted payoff $\hat{v}_3$ still touches the graph at the two induced posteriors, now $\mu_l = 0.092$ and $\mu_h = \frac{1}{2}$. Reading the closure at the prior gives $V_3(0.3) = 1.52$, and the prosecutor's payoff is $V_3(0.3) - kH(\mu_0) = 1.52 - 1.26 = 0.26$, matching Table 2.

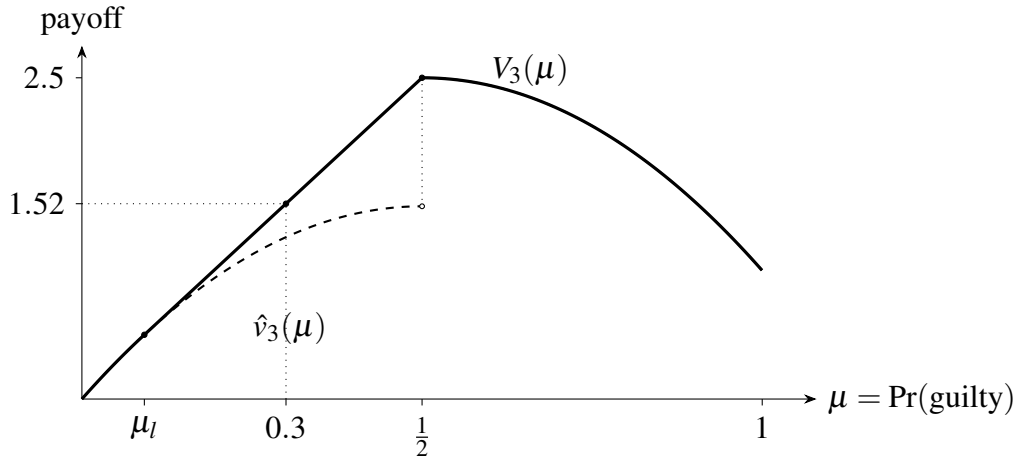


Figure 3: The cost-adjusted payoff $\hat{v}_3$ (dashed) and its concave closure V_3 (solid) at $k = 3$, with induced posteriors $\mu_l = 0.092$ and $\mu_h = \frac{1}{2}$.

Finally, Table 3 reports GK's own version of the example, which uses Shannon entropy as the measure of uncertainty and the uniform belief as the reference, rather than the residual variance and prior reference used above. The pattern is the same, and $k = 0$ again returns the costless investigation of Section 2.

k	Conviction rate	μ_i	Prosecutor net
0	0.600	0.000	0.600
1	0.543	0.062	0.384
2	0.457	0.131	0.259
3	0.382	0.176	0.177
4	0.315	0.208	0.119
5	0.255	0.232	0.077
7	0.147	0.265	0.025
10	0.009	0.298	0.000

Table 3: GK’s entropy version of the example, reference belief uniform.¹

4 Persuasion under Disagreement (Alonso and Cámara, 2016)

4.1 From a shared prior to disagreement

Sections 2 and 3 both assume that the prosecutor and the judge start from the same prior, μ_0 . Often they do not. A prosecutor who has lived with a case for months may be more convinced of guilt than the judge who sees it for the first time; an advocate is typically more optimistic than the audience he addresses. Alonso and Cámara (2016) allow the Sender and the Receiver to hold different priors, both common knowledge, so that the two agree to disagree.

We keep the judge’s prior at the same generic value μ_0 (running value 0.3), so her behavior is exactly as in Section 2, and we let the prosecutor’s prior be a number $q \in (0, 1)$. When $q = \mu_0$ the two agree and we are back in Section 2. The question is what disagreement changes when $q \neq \mu_0$.

4.2 The model with two priors

The states, actions, and payoffs are those of Section 2. The judge holds the prior μ_0 and convicts if and only if her posterior is at least $\frac{1}{2}$, unchanged. The prosecutor holds a possibly different prior q . He designs a signal; the judge updates from her prior μ_0 ; but the prosecutor evaluates outcomes using his prior q , that is, using his own assessment of how likely each message, and each resulting conviction, is. Both priors are common knowledge.

¹The published example reports a conviction rate of 0.43 at $k = 2$. The signal listed there induces a conviction belief $\mu_g = 0.4988$, just short of the judge’s indifference point $\frac{1}{2}$, so it narrowly fails her obedience constraint; the obedience-respecting optimum, shown here, convicts at rate 0.457. The qualitative comparative statics are identical.

4.3 The unifying result: reweight the value function

The prosecutor and the judge read the same evidence but weigh the possible verdicts differently, because they disagree about how likely guilt is.

First, the judge's behavior still depends only on her posterior μ : she convicts if and only if $\mu \geq \frac{1}{2}$, and the prosecutor's gross payoff at that posterior is the same step function $\hat{v}(\mu)$ as in Section 2.

Second, the prosecutor weights each message not by the judge's probability but by his own. For a message that leaves the judge at posterior μ , the ratio of the prosecutor's probability of that message to the judge's probability of it is

$$\ell(\mu) = \mu \frac{q}{\mu_0} + (1 - \mu) \frac{1 - q}{1 - \mu_0},$$

which depends only on the belief the message induces, not on the signal that produced it. The reason is that, once the judge's posterior μ is known, the message's likelihood in each state relative to its average likelihood is fixed, and the two priors simply reweight those likelihoods.

Putting the two facts together, the prosecutor's expected payoff is $\mathbb{E}_\tau[\ell(\mu) \hat{v}(\mu)]$, where τ is the judge's Bayes-plausible distribution of posteriors. Define the reweighted value $w(\mu) \equiv \ell(\mu) \hat{v}(\mu)$. The prosecutor's problem is

$$\max_{\tau \text{ Bayes plausible}} \mathbb{E}_\tau[w(\mu)] \quad \text{subject to } \mathbb{E}_\tau[\mu] = \mu_0,$$

solved once again by concavifying w and reading it at the judge's prior μ_0 .

The nesting. When $q = \mu_0$, the reweighting is $\ell(\mu) = \mu + (1 - \mu) = 1$, so $w = \hat{v}$ and we recover Section 2 exactly. Heterogeneous priors is the family; the common-prior model of Kamenica and Gentzkow (2011) is its $q = \mu_0$ member.

This reweighting is close kin to the device behind costly persuasion. When the cost in Section 3 is measured from a reference belief other than the prior, Gentzkow and Kamenica (2014) translate between that reference belief and the prior in exactly the way we here translate between the prosecutor's and the judge's priors, and they credit the device to Alonso and Cámara's work on heterogeneous priors. A single belief-map underlies both extensions, which is why the two share a shape.

4.4 The toy model, worked

Step 1. Reweight the payoff. With judge prior μ_0 ,

$$\ell(\mu) = \mu \frac{q}{\mu_0} + (1 - \mu) \frac{1 - q}{1 - \mu_0}, \quad w(\mu) = \begin{cases} 0 & \text{if } \mu < \frac{1}{2}, \\ \ell(\mu) & \text{if } \mu \geq \frac{1}{2}. \end{cases}$$

Since ℓ is positive and rises at most linearly in μ , the reweighted payoff w is still a single upward step, now with a sloped top.

Step 2. The optimal signal is unchanged. Because $\hat{v}$ is a step and ℓ is a positive affine function of μ , the concave closure of w touches the graph at the same two beliefs as in Section 2: $\mu_l = 0$ and $\mu_h = \frac{1}{2}$. Raising μ_h above $\frac{1}{2}$ lowers the ratio $\ell(\mu_h)/\mu_h$, and lowering μ_l to 0 frees weight for the conviction message, exactly as before. So the judge's optimal investigation is the one derived in Section 2, the same two messages and the same conviction rate of $2\mu_0$, for every prosecutor prior q . Intuitively, disagreement about the prior does not change which beliefs are cheapest to induce, because the judge's Bayes plausibility and her indifference point are hers alone, and neither depends on q .

Step 3. What the disagreement changes. It changes how much the prosecutor values the unchanged investigation. His expected payoff is

$$\text{value} = 2\mu_0 \ell\left(\frac{1}{2}\right) = q + \frac{\mu_0(1 - q)}{1 - \mu_0},$$

which rises in his own conviction q . At $q = \mu_0$ it equals $2\mu_0$, the Section 2 value; at the running $\mu_0 = 0.3$ the value is $(3 + 4q)/7$. A prosecutor more sure of guilt than the judge, $q > \mu_0$, values the same investigation more, because he believes convictions are more likely than the judge does.

Step 4. The prosecutor always benefits. With no investigation the judge keeps her prior $\mu_0 < \frac{1}{2}$ and acquits, so the prosecutor's payoff is 0 whatever his own prior. Since the value $q + \mu_0(1 - q)/(1 - \mu_0)$ is positive for every q , the prosecutor benefits from persuasion at every prior, even when he is nearly certain of guilt. This is Alonso and Cámara's striking observation: a Sender gains from generating information even when the audience is already inclined his way.

4.5 A numerical summary

Table 4 reports the prosecutor's value as his prior q varies while the judge's prior stays at 0.3. The judge's conviction rate and her welfare do not move; only the prosecutor's valuation does, climbing along the line $(3 + 4q)/7$.

q	$\ell(\frac{1}{2})$	Prosecutor value	Judge conviction	Judge welfare
0.1	0.810	0.486	0.60	0.70
0.2	0.905	0.543	0.60	0.70
0.3	1.000	0.600	0.60	0.70
0.5	1.190	0.714	0.60	0.70
0.7	1.381	0.829	0.60	0.70
0.9	1.571	0.943	0.60	0.70

Table 4: Prosecutor value as the prosecutor’s prior q varies, judge’s prior fixed at $\mu_0 = 0.3$. The row $q = \mu_0 = 0.3$ is Section 2.

Figure 4 draws the case $q = 0.5$ in the same way as Figures 1 and 3. The reweighted payoff w is again a single step, but its upper branch now slopes upward, because the prosecutor values a firmer conviction belief more than a marginal one. The concave closure still touches the graph at $\mu_l = 0$ and $\mu_h = \frac{1}{2}$, the very posteriors of Section 2, and reading it at the prior gives the prosecutor’s value 0.71.

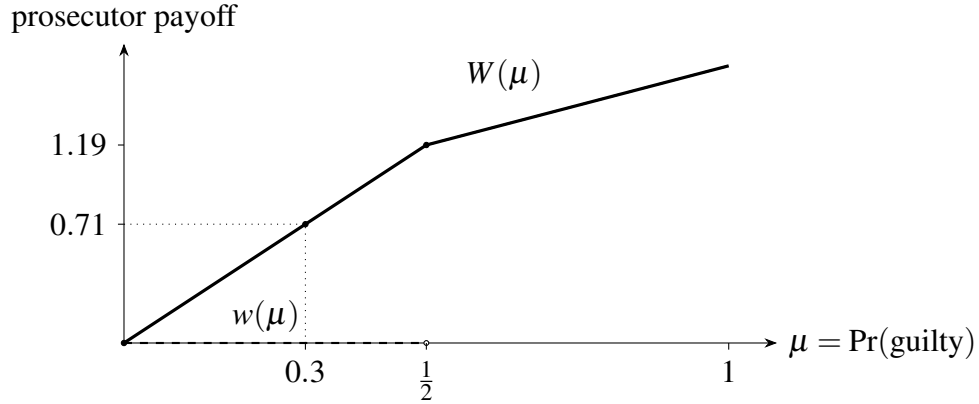


Figure 4: The reweighted payoff w (dashed) and its concave closure W (solid) at $q = 0.5$, with the same induced posteriors $\mu_l = 0$ and $\mu_h = \frac{1}{2}$ as in Figure 1.

5 Competition in Persuasion (Gentzkow and Kamenica, 2017)

5.1 Adding a second persuader

So far a single Sender controlled all the evidence. A courtroom has two. The prosecutor wants a conviction; the defense attorney wants an acquittal; each is free to commission investigations, and the judge sees the results of both. The old faith in the adversarial system holds that setting two interested persuaders against each other protects the audience. Gentzkow and Kamenica (2017) make

the idea precise, and in our binary courtroom the answer is stark: two fully opposed persuaders reveal the whole truth.

5.2 The model with two senders

The states, actions, judge, and prior μ_0 are those of Section 2. Now two senders design evidence. The prosecutor earns $\hat{v}_P(\mu) = 1$ when the judge convicts ($\mu \geq \frac{1}{2}$) and 0 otherwise; the defense attorney earns the mirror image, $\hat{v}_D(\mu) = 1$ when the judge acquits ($\mu < \frac{1}{2}$) and 0 otherwise. Each sender chooses a signal; the judge pools everything she sees into a single posterior and acts; a profile of signals is an equilibrium when neither sender can gain by changing his own signal alone. With only the prosecutor present, this is exactly Section 2.

5.3 The unifying result: both senders sit on their concave closure

The concavification of Section 2 told a lone sender where he would place the judge's belief. With two senders the same object characterizes equilibrium, through a short argument. Suppose an equilibrium leaves the judge, with positive probability, at a belief μ where the prosecutor's payoff is strictly below its concave closure, $\hat{v}_P(\mu) < V_P(\mu)$. The prosecutor could then add information at that belief, splitting it into two whose average is μ , and raise his expected payoff, since the concave closure is the most that splitting can reach. He would deviate, so this is not an equilibrium. The same argument applies to the defense attorney. Hence, in equilibrium, every belief the judge might hold lies where *both* senders sit exactly on their own concave closure.

Write $M_i = \{\mu : \hat{v}_i(\mu) = V_i(\mu)\}$ for the beliefs at which sender i cannot improve by splitting, and $M = M_P \cap M_D$. Gentzkow and Kamenica (2017) show that a Bayes-plausible distribution of posteriors is an equilibrium outcome if and only if its support lies in M .

The nesting. With a single sender, $M = M_P$, and the requirement is simply that the judge's beliefs sit on the prosecutor's concave closure, which is the concavification of Section 2. Kamenica and Gentzkow (2011) is the one-sender case; competition is what a second, opposed sender adds.

5.4 The toy model, worked

Step 1. The prosecutor's coincidence set. The prosecutor's payoff $\hat{v}_P$ is 0 below $\frac{1}{2}$ and 1 above, and its concave closure is $V_P(\mu) = \min(2\mu, 1)$. The two agree wherever $\hat{v}_P$ is already at its ceiling, the interval $[\frac{1}{2}, 1]$, together with the single belief $\mu = 0$ where both equal 0. Hence

$$M_P = \{0\} \cup [\frac{1}{2}, 1].$$

Step 2. The defense attorney's coincidence set. His payoff $\hat{v}_D$ is 1 below $\frac{1}{2}$ and 0 above, with concave closure $V_D(\mu) = \min(1, 2(1 - \mu))$. The two agree on $[0, \frac{1}{2})$, where $\hat{v}_D$ is already at its ceiling, together with $\mu = 1$ where both equal 0. Hence

$$M_D = [0, \frac{1}{2}) \cup \{1\}.$$

Step 3. Intersect. The only beliefs in both sets are the two certainties,

$$M = M_P \cap M_D = \{0, 1\}.$$

At every other belief one of the two senders would add information.

Step 4. The equilibrium. The judge's beliefs must therefore be 0 or 1, so the only Bayes-plausible outcome is full disclosure: the evidence reveals the state, leaving belief 1 with probability μ_0 and belief 0 with probability $1 - \mu_0$. The judge convicts exactly the guilty, at rate μ_0 , half the lone prosecutor's $2\mu_0$.

Intuitively, at any partial belief short of certainty, one side always holds evidence that pulls the judge its way: if the prosecutor tries to stop at the indifference belief, the defense attorney can push her toward innocence, and conversely. The only beliefs no one wishes to disturb are the two at which nothing more can be said. Competition unravels all the way to the truth.

5.5 What competition adds

Competition reveals more, and the judge gains. A lone prosecutor convicts $2\mu_0$ of defendants; two opposed persuaders convict only the guilty, μ_0 , and the judge, now fully informed, reaches her first-best. This is the sharpest form of the general comparative static: competition never reduces the information revealed, and here it forces the maximum.

Full opposition is doing the work. The unraveling is this clean because the senders' interests are exactly reversed, so their coincidence sets meet only at certainty. Gentzkow and Kamenica (2017) show, well beyond the binary case, that a competitive equilibrium is never strictly less informative than what the senders would choose by colluding, whatever their preferences; full disclosure is the extreme our courtroom lands on.

5.6 A numerical summary

Table 5 sets the one-sender and two-sender outcomes side by side at the running prior $\mu_0 = 0.3$. Adding the defense attorney halves the conviction rate and lifts the judge to full information.

Setting	Conviction rate	Prosecutor payoff	Judge welfare	Defense payoff
One sender (Section 2)	0.60	0.60	0.70	—
Two opposed senders	0.30	0.30	1.00	0.70

Table 5: One persuader versus two opposed persuaders, at $\mu_0 = 0.3$. In general the conviction rate falls from $2\mu_0$ to μ_0 and the judge’s welfare rises from $1 - \mu_0$ to 1.

Figure 5 draws the two senders’ payoffs. The prosecutor’s step rises at $\mu = \frac{1}{2}$; the defense attorney’s falls there. Each payoff meets its concave closure only on the half where it is already at its ceiling, plus the far certainty, so the two coincidence sets overlap only at $\mu = 0$ and $\mu = 1$: the beliefs of full disclosure.

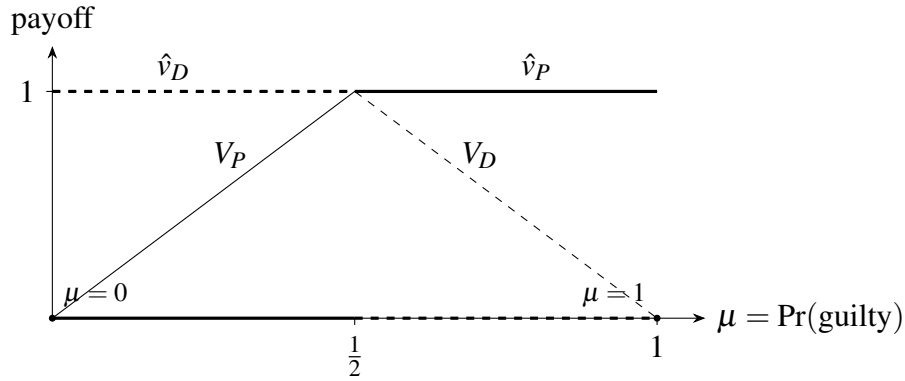


Figure 5: The prosecutor’s payoff $\hat{v}_P$ (thick solid) and the defense attorney’s $\hat{v}_D$ (thick dashed), with their concave closures V_P, V_D (thin). The closures are met by their own payoffs only at $\mu = 0$ and $\mu = 1$, the full-disclosure beliefs.

6 A Guided Tour of the Broader Literature

The four sections above share one move: write the Sender’s payoff as a function of the Receiver’s belief, and the best the Sender can do is the value of the concave closure of that payoff at the prior. Costs, disagreement, and competition enter only by reshaping that function or by intersecting two of them. The method works cleanly under three conditions: the state space is small enough to plot the value function, the Sender commits fully to the experiment, and there is a single Receiver whose action depends on her posterior. Each active branch of the literature relaxes one of these conditions. This section takes the branches in turn, keeping the prosecutor and judge in view. For each, we say what it changes relative to our baseline, describe concretely how the model would be modified to capture it, and point to representative papers, so that a student can move from our courtroom to the frontier without changing examples. The literature is large; we refer the reader to

Kamenica (2019) and Bergemann and Morris (2019) for the complete bibliography.

6.1 Many receivers

What changes. There is now a panel rather than a single judge, and the Sender may address the panel with one public signal or with private, receiver-specific signals; the receivers may also care about one another’s verdicts. *How the model changes.* Replace the judge with n jurors and a collective rule, say conviction by majority. If the prosecutor is restricted to a single *public* signal, nothing else changes: reinterpret $\bar{a}(\mu)$ as the panel’s collective verdict when all jurors share the posterior μ , and concavify $\hat{v}$ exactly as in Section 2. If instead he sends *private* signals and jurors care about the collective outcome, a juror’s vote depends not only on her own belief but on what she thinks the others learned, and the belief-based value function no longer captures the problem. One then works in the language of Bayes correlated equilibria (Bergemann and Morris 2013, 2016) and information design (Taneva 2019; Mathevet, Perego, and Taneva 2020): characterize the outcomes attainable under *some* information structure, a set carved out by each juror’s obedience constraints, and pick the best. Persuading voters is the workhorse case (Alonso and Cámara 2016b; Wang 2015; Bardhi and Guo 2018), and the same tools yield the map of feasible profit–surplus pairs under third-degree price discrimination in Bergemann, Brooks, and Morris (2015).

6.2 Many senders: competition in persuasion

What changes. The single persuader becomes several, with possibly opposed preferences, each free to design evidence that the Receiver pools. *How the model changes.* Give sender i a value function $\hat{v}_i$ and its concave closure V_i , exactly as we did for the prosecutor and the defense attorney; an equilibrium is a Bayes-plausible belief distribution supported only where *every* sender sits on his own concave closure, that is, on $\bigcap_i M_i$. Gentzkow and Kamenica (2017) show, well beyond the binary case, that a competitive equilibrium is never strictly less informative than what the senders would choose by colluding; our full-disclosure result is the extreme this comparative static reaches when the two sides are exactly opposed. The conclusion is not automatic once senders cannot freely correlate their signals: Board and Lu (2018) show that in a search market where sellers can coordinate their disclosures, competition may reveal no more than a monopolist would, while the same sellers reveal everything once coordination is impossible. Whether senders who move sequentially rather than simultaneously reveal more is studied by Li and Norman (2018).

6.3 Limited commitment

What changes. The prosecutor no longer commits fully; with some probability he can quietly alter the message he reports, which is the assumption that had separated persuasion from cheap talk (Crawford and Sobel 1982) and from verifiable disclosure (Grossman 1981; Milgrom 1981). *How the model changes.* Add one parameter, the probability ρ that the prosecutor can secretly replace an unfavorable message with a favorable one. At $\rho = 0$ we are back in Section 2; at $\rho = 1$ the prosecutor is a pure cheap talker, and the judge trusts nothing that is not self-serving to reveal. In between, the value is governed by an object that combines the *concavification* of full commitment with the *quasi-concavification* that solves the cheap-talk problem for a state-independent Sender (Lipnowski and Ravid 2020; Lipnowski, Ravid, and Shishkin 2022; Min 2017). A striking lesson is that the judge can be better off when the prosecutor is *less* credible, since a prosecutor who cannot be believed cannot mislead. A related relaxation lets the Sender observe something about the state before he designs the experiment, so that his very choice of investigation leaks information (Perez-Richet 2014; Hedlund 2017), and laboratory tests of how closely subjects follow these predictions are beginning to appear (Fréchette, Lizzeri, and Perego 2022).

6.4 A better-informed receiver

What changes. The judge is no longer a blank slate: she holds private information, about her own threshold for conviction or about the state, or she can gather her own evidence after seeing the prosecutor's. *How the model changes.* Give the judge a private type θ , say a privately known conviction threshold drawn from a known distribution, and redefine the prosecutor's value $\hat{v}$ by averaging over θ ; the problem stays concavification-friendly, but the optimal investigation acquires structure, becoming a monotone rule that reveals middling cases and pools extreme ones (Kolotilin, Mylovanov, Zapechelnyuk, and Li 2017; Kolotilin 2018), or an interval rule when the private information is about the state (Guo and Shmaya 2019). Rayo and Segal (2010), contemporaneously with Kamenica and Gentzkow (2011), already give the receiver a private outside option. Alternatively, let the judge pay a cost to run her own test after the prosecutor's message: the mere threat disciplines him into disclosing more, and it is without loss to look at equilibria in which she never actually investigates (Matysková 2018).

6.5 Dynamics

What changes. The problem is no longer static: the state evolves over time, or the receiver's past choices shape her future options, so the Sender chooses not only how much to reveal but when. *How the model changes.* Let the defendant's guilt evolve as a Markov process and let the judge

act myopically each period on her current belief; the prosecutor now chooses a disclosure policy over time rather than a single experiment. Ely (2017) solves this case and finds that the optimal policy discloses good news with a deterministic delay. When the receiver’s sunk actions change her opportunities, dynamic disclosure can sustain more of the action the Sender wants than any one-shot experiment (Ely and Szydlowski 2020). A separate dynamic branch puts beliefs directly into the utility function, modelling entertainment as the management of suspense and surprise (Ely, Frankel, and Kamenica 2015).

6.6 Large state spaces and solution methods

What changes. The state is a continuum and the Sender’s payoff depends only on the *mean* of the receiver’s posterior, so the value function can still be drawn but its concave closure no longer solves the problem, because not every distribution of posterior means with the right average is attainable.

How the model changes. Let guilt be a continuous severity in $[0, 1]$ rather than a yes-or-no, with the judge’s action depending on expected severity. Two reformulations restore tractability: represent each signal by a convex function of the posterior mean and search over that space (Gentzkow and Kamenica 2016), or cast the design as a linear program, equivalently a price-theoretic problem in which the Sender “buys” posterior means using the prior as his budget, with a simple dual test for optimality (Kolotilin 2018; Dworczak and Martini 2019). These methods say when the optimum is a monotone partition (Ivanov 2015; Mensch 2018) and now carry much of the applied work. When even they fail, computation is the fallback: the general problem is hard, but good approximation algorithms exist (Dughmi and Xu 2016; Dughmi 2017).

6.7 Applications

What changes. Nothing in the mechanism; only the names of the Sender, the Receiver, and the state. *How the model changes.* Rename the players. A regulator designing a bank stress test reveals just enough about solvency to avert a run (Goldstein and Leitner 2018; Inostroza and Pavan 2023); a school choosing a grading policy reveals just enough about ability to place its graduates (Ostrovsky and Schwarz 2010; Boleslavsky and Cotton 2015); a routing app gives drivers noisy advice to unclog roads (Das, Kamenica, and Mirka 2017); a police force randomizes speed traps to deter speeding (Lazear 2006); and, in the environmental reading we return to throughout, a firm designs a sustainability label that raises a green-minded consumer’s belief to the buying threshold and no further, which is precisely greenwashing (Lyon and Maxwell 2011). In each case the prosecutor–judge diagram is a faithful first picture, and the comparative statics of Sections 3–5 carry over: costlier testing means coarser disclosure, a more confident designer values the same

disclosure more, and a watchdog forces the truth out.

6.8 A neighboring market: selling information

What changes. Information is no longer given away to move an action but sold as a good, and the designer, a data broker, a credit bureau, a rating agency, is paid in fees rather than in the Receiver's action (Admati and Pfleiderer 1990; Bergemann and Bonatti 2019). *How the model changes.* Replace the prosecutor's verdict payoff with a price: instead of valuing a conviction, the Sender posts a menu of experiments and prices and the Receiver buys one. The design question shifts from what to reveal to what to sell and at what price, and the difference between selling an information structure up front and selling its realization after the fact maps onto the products that brokers and publishers actually trade. Our courtroom becomes a pricing problem, and the two literatures meet in a shared object.

7 Directions for Future Work

The prosecutor and the judge are not only a teaching device; several of the field's open questions can be posed inside them, which is a useful discipline on speculation. We describe four directions where the model of this handout runs into the research frontier.

Combining contracts with experiments. Incentives and information are usually studied in separate literatures, yet real designers set both at once: a firm chooses a wage schedule and an evaluation system, a court sets both penalties and rules of evidence, a platform sets both prices and what buyers are shown. Inside our model this means giving the prosecutor a second instrument, a transfer to the judge, alongside the investigation, and asking how the optimal experiment changes when the Sender can also move the Receiver's payoffs directly. Li (2017) takes a first step by adding transfers to the two-action, two-state model, but a general theory of the joint design of a contract and an experiment is missing, and it is the natural home for the question of when disclosure and payment are substitutes and when they are complements.

A theory, not a catalogue, of many receivers. The papers on persuading voters, juries, and bidders each solve a particular case, but there is no organizing statement of when a designer prefers a single public signal to receiver-specific private ones, or of how that preference turns on whether the receivers' actions are strategic complements or substitutes. The machinery to pose the question exists, in the form of Bayes correlated equilibria (Bergemann and Morris 2016); what is missing is the comparative-statics theory built on top of it. In our setting this is the question of how the optimal investigation changes as the lone judge becomes a voting panel, and as the voting rule moves from unanimity to simple majority.

Designing information and options together, over time. A court controls both the evidence a jury sees and the order in which questions are put to it; a firm controls both an employee's feedback and the tasks that remain open to her after she acts. Treating the design of information and the design of the receiver's evolving choice set as one problem, rather than two, is young and promising (Ely and Szydlowski 2020), and the promotion and trial narratives used in this handout are natural laboratories for it: let the prosecutor choose not only what the judge learns but when the verdict must be cast, and the timing of disclosure becomes a second lever.

Solving harder problems, and testing the benchmark. Most models past the smallest examples resist a closed form. The convex-analysis and linear-programming methods of Section 6 have widened the solvable class, but many applied designs, especially with several interacting receivers, remain out of analytic reach, and the algorithmic methods developed in computer science are only beginning to be imported into economics (Dughmi 2017). Running alongside is an empirical question the theory cannot answer on its own: do real receivers update as Bayesian persuasion assumes? Laboratory work has just begun to test this (Fréchette, Lizzeri, and Perego 2022), and systematic evidence would tell how far to trust the rational benchmark that makes the concavification picture exact.

References

- Admati, A. R., and P. Pfleiderer (1990). Direct and Indirect Sale of Information. *Econometrica*, 58(4), 901–928.
- Alonso, R., and O. Cámara (2016a). Bayesian Persuasion with Heterogeneous Priors. *Journal of Economic Theory*, 165, 672–706.
- Alonso, R., and O. Cámara (2016b). Persuading Voters. *American Economic Review*, 106(11), 3590–3605.
- Aumann, R. J., and M. B. Maschler (1995). *Repeated Games with Incomplete Information*. Cambridge, MA: MIT Press.
- Bardhi, A., and Y. Guo (2018). Modes of Persuasion toward Unanimous Consent. *Theoretical Economics*, 13(3), 1111–1150.
- Bergemann, D., and A. Bonatti (2019). Markets for Information: An Introduction. *Annual Review of Economics*, 11, 85–107.
- Bergemann, D., B. Brooks, and S. Morris (2015). The Limits of Price Discrimination. *American Economic Review*, 105(3), 921–957.

- Bergemann, D., and S. Morris (2013). Robust Predictions in Games with Incomplete Information. *Econometrica*, 81(4), 1251–1308.
- Bergemann, D., and S. Morris (2016). Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium. *American Economic Review: Papers & Proceedings*, 106(5), 586–591.
- Bergemann, D., and S. Morris (2019). Information Design: A Unified Perspective. *Journal of Economic Literature*, 57(1), 44–95.
- Board, S., and J. Lu (2018). Competitive Information Disclosure in Search Markets. *Journal of Political Economy*, 126(5), 1965–2010.
- Boleslavsky, R., and C. Cotton (2015). Grading Standards and Education Quality. *American Economic Journal: Microeconomics*, 7(2), 248–279.
- Crawford, V. P., and J. Sobel (1982). Strategic Information Transmission. *Econometrica*, 50(6), 1431–1451.
- Das, S., E. Kamenica, and R. Mirka (2017). Reducing Congestion through Information Design. In *55th Annual Allerton Conference on Communication, Control, and Computing*, 1279–1284.
- Dughmi, S. (2017). Algorithmic Information Structure Design: A Survey. *SIGecom Exchanges*, 15(2), 2–24.
- Dughmi, S., and H. Xu (2016). Algorithmic Bayesian Persuasion. *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, 412–425.
- Dworczak, P., and G. Martini (2019). The Simple Economics of Optimal Persuasion. *Journal of Political Economy*, 127(5), 1993–2048.
- Ely, J. C. (2017). Beeps. *American Economic Review*, 107(1), 31–53.
- Ely, J. C., A. Frankel, and E. Kamenica (2015). Suspense and Surprise. *Journal of Political Economy*, 123(1), 215–260.
- Ely, J. C., and M. Szydlowski (2020). Moving the Goalposts. *Journal of Political Economy*, 128(2), 468–506.
- Fréchette, G. R., A. Lizzeri, and J. Perego (2022). Rules and Commitment in Communication: An Experimental Analysis. *Econometrica*, 90(5), 2283–2318.
- Gentzkow, M., and E. Kamenica (2014). Costly Persuasion. *American Economic Review: Papers & Proceedings*, 104(5), 457–462.

- Gentzkow, M., and E. Kamenica (2016). A Rothschild–Stiglitz Approach to Bayesian Persuasion. *American Economic Review: Papers & Proceedings*, 106(5), 597–601.
- Gentzkow, M., and E. Kamenica (2017). Competition in Persuasion. *Review of Economic Studies*, 84(1), 300–322.
- Goldstein, I., and Y. Leitner (2018). Stress Tests and Information Disclosure. *Journal of Economic Theory*, 177, 34–69.
- Grossman, S. J. (1981). The Informational Role of Warranties and Private Disclosure about Product Quality. *Journal of Law and Economics*, 24(3), 461–483.
- Guo, Y., and E. Shmaya (2019). The Interval Structure of Optimal Disclosure. *Econometrica*, 87(2), 653–675.
- Hedlund, J. (2017). Bayesian Persuasion by a Privately Informed Sender. *Journal of Economic Theory*, 167, 229–268.
- Inostroza, N., and A. Pavan (2023). Adversarial Coordination and Public Information Design. *Northwestern working paper*.
- Ivanov, M. (2015). Optimal Signals in Bayesian Persuasion Mechanisms. *McMaster working paper*.
- Kamenica, E. (2019). Bayesian Persuasion and Information Design. *Annual Review of Economics*, 11, 249–272.
- Kamenica, E., and M. Gentzkow (2011). Bayesian Persuasion. *American Economic Review*, 101(6), 2590–2615.
- Kolotilin, A. (2018). Optimal Information Disclosure: A Linear Programming Approach. *Theoretical Economics*, 13(2), 607–635.
- Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017). Persuasion of a Privately Informed Receiver. *Econometrica*, 85(6), 1949–1964.
- Lazear, E. P. (2006). Speeding, Terrorism, and Teaching to the Test. *Quarterly Journal of Economics*, 121(3), 1029–1061.
- Li, F. (2017). A Note on the Effect of Payments on Persuasion. *Economics Letters*, 156, 155–157.
- Li, F., and P. Norman (2018). On Bayesian Persuasion with Multiple Senders. *Economics Letters*, 170, 66–70.

- Lipnowski, E., and D. Ravid (2020). Cheap Talk with Transparent Motives. *Econometrica*, 88(4), 1631–1660.
- Lipnowski, E., D. Ravid, and D. Shishkin (2022). Persuasion via Weak Institutions. *Journal of Political Economy*, 130(10), 2705–2730.
- Lyon, T. P., and J. W. Maxwell (2011). Greenwash: Corporate Environmental Disclosure under Threat of Audit. *Journal of Economics & Management Strategy*, 20(1), 3–41.
- Mathevet, L., J. Perego, and I. Taneva (2020). On Information Design in Games. *Journal of Political Economy*, 128(4), 1370–1404.
- Matysková, L. (2018). Bayesian Persuasion with Costly Information Acquisition. *CERGE-EI working paper*.
- Mensch, J. (2018). Monotone Persuasion. *Games and Economic Behavior*, 122, 220–237.
- Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications. *Bell Journal of Economics*, 12(2), 380–391.
- Min, D. (2017). Bayesian Persuasion under Partial Commitment. *Bank of Korea working paper*.
- Ostrovsky, M., and M. Schwarz (2010). Information Disclosure and Unraveling in Matching Markets. *American Economic Journal: Microeconomics*, 2(2), 34–63.
- Perez-Richet, E. (2014). Interim Bayesian Persuasion: First Steps. *American Economic Review: Papers & Proceedings*, 104(5), 469–474.
- Rayo, L., and I. Segal (2010). Optimal Information Disclosure. *Journal of Political Economy*, 118(5), 949–987.
- Taneva, I. (2019). Information Design. *American Economic Journal: Microeconomics*, 11(4), 151–185.
- Wang, Y. (2015). Bayesian Persuasion with Multiple Receivers. *Available at SSRN 2625399*.

Funding a Review Agency under Obfuscation

A step-by-step exercise in the spirit of Figueroa and Lemus (2024)

Companion note to Figueroa, N., and J. Lemus (2024), “The Benefit of the Doubt: Patent Examination under Strategic Obfuscation,” *Management Science* 70(9): 5853–5870. Full reference at the end.

Overview. Competition authorities examine merger requests, patent offices examine patent applications, and drug regulators examine drug applications. In each case the agency decides how many resources to devote to examination, and the applicants who know that their projects are bad respond by making the application hard to evaluate: they add claims, documents, and information that are costly to check. Figueroa and Lemus (2024) show that these two choices interact. More examination resources raise the effort that applicants with bad projects spend on obfuscation, so the agency can be worse off with more resources. This note builds the simplest version of their model and then asks the question that a funding agency faces: how much should it spend? Under a quadratic cost of resources, the answer is not a smooth function of the cost. When good projects are valuable enough, the optimal budget is either full or modest, and it jumps from one to the other at a single unit cost. Step (a) sets up the benchmark without obfuscation; Step (b) derives the effort of an applicant with a bad project; Step (c) solves the agency’s approval decision and its payoff; Step (d) shows when more resources lower that payoff; Step (e) solves the budget problem and explains why the argument extends to any strictly increasing cost; Step (f) solves it in closed form; Step (g) draws two policy implications; and Step (h) varies the parameters and lists what the toy model keeps and drops. Section 1 fixes one set of numbers, and every step uses it, so that each expression below is a number rather than a formula.

1. The Toy Model

Applicant and agency. An applicant has a good project with probability $p = \frac{1}{2}$ and a bad project otherwise, and it privately observes which. The agency maximizes a welfare measure, such as consumer surplus. Approving a good project yields the agency a value $v = 6$, approving a bad project costs it $c = 2$, and rejection yields zero. Approval yields the applicant a gain that we normalize to one. In merger review the applicant is the pair of merging firms, who know whether the merger is good or bad for consumers; in patent examination the applicant knows whether granting a patent on its idea raises or lowers welfare. Two numbers summarize the stakes: the expected value of a good project is $pv = 3$ and the expected cost of a bad project is $(1 - p)c = 1$. An agency that cannot tell the projects apart gains $3 - 1 = 2 > 0$ from approving an application, so it approves when it is uncertain, unless obfuscation changes the pool of uncertain applications.

Obfuscation and examination. The agency invests resources $\rho \in [0, 1]$, which the applicant observes. The applicant then chooses an obfuscation effort $e \geq 0$ at cost e , which the agency does not observe. With probability $1 - \varphi(\rho, e)$ the agency learns the type of the project and approves if and only if it is good. With probability $\varphi(\rho, e)$ it remains uncertain, and it then approves with probability $\lambda \in [0, 1]$, which it chooses. We use the specification of Figueroa and Lemus (2024),

$$\varphi(\rho, e) = 1 - \rho(1 - \sqrt{2e}), \quad e \in [0, \tfrac{1}{2}]. \quad (1)$$

More resources raise the probability of learning the type, and more obfuscation lowers it. With no obfuscation ($e = 0$) the agency stays uncertain with probability $1 - \rho$, and with full resources and no obfuscation it learns the type for sure, $\varphi(1, 0) = 0$.

Budget. Resources cost $K(\rho) = k\rho^2$, where $k \geq 0$ is the *unit cost* of resources. The agency's payoff is $U(\rho) - k\rho^2$, where $U(\rho)$ is its equilibrium payoff from examination, derived below.

Timing. (1) The agency chooses ρ , which the applicant observes. (2) The applicant learns its type and chooses e . (3) The agency learns the type with probability $1 - \varphi(\rho, e)$; otherwise it approves with probability λ . We solve for the perfect Bayesian equilibrium by backward induction.

Parameter values. We fix $p = \frac{1}{2}$, $v = 6$, and $c = 2$ in every step below. The values keep the arithmetic round, and they have the two properties that the analysis needs: the payoff from examination has a dip (Step d), and the optimal budget jumps from full to modest (Step f). Step (h) states the conditions on p , v , and c under which these properties hold and shows that other values give the same qualitative results.

2. Step (a): The Benchmark without Obfuscation

Consider first the case in which applicants cannot obfuscate, so $e = 0$ and the agency stays uncertain with probability $1 - \rho$ about every project. An uncertain agency faces good and bad projects in equal proportions, and approving yields $\frac{1}{2} \cdot 6 - \frac{1}{2} \cdot 2 = 2 > 0$ in expectation, so it approves every uncertain application, $\lambda = 1$. It collects the value 3 of the good projects and bears the cost 1 only when a bad project goes unexamined:

$$U_0(\rho) = 3 - (1 - \rho) = 2 + \rho. \quad (2)$$

This payoff is strictly increasing in ρ at the constant rate 1: a unit of resources lowers by one the probability that a bad project goes unexamined, which avoids the expected cost 1 of approving it. The agency maximizes $2 + \rho - k\rho^2$ over $[0, 1]$, so

$$\rho_0^* = \min \left\{ \frac{1}{2k}, 1 \right\}. \quad (3)$$

Result 1 (benchmark). Without obfuscation the optimal budget is continuous and weakly decreasing in the unit cost: it equals one up to $k = \frac{1}{2}$ and falls gradually afterwards.

The budget is one for $k \leq \frac{1}{2}$, $\frac{1}{2}$ at $k = 1$, and $\frac{1}{4}$ at $k = 2$. There is no jump. Steps (b) to (f) show that obfuscation changes this picture.

3. Step (b): The Applicant with a Bad Project

Consider the applicant's choice of effort in stage 2 for given resources ρ and for an anticipated approval probability λ . An applicant with a good project is approved with probability $1 - \varphi(1 - \lambda)$, which is weakly decreasing in φ , so it never spends effort that raises φ at a positive cost. An applicant with a bad project is approved only when the agency remains uncertain and then approves, which has probability $\lambda\varphi$, so it maximizes

$$\lambda\varphi(\rho, e) - e = \lambda(1 - \rho + \rho\sqrt{2e}) - e. \quad (4)$$

This function is concave in e . For $\lambda\rho > 0$ the first-order condition is $\frac{\lambda\rho}{\sqrt{2e}} = 1$, and it gives

$$e = \frac{(\lambda\rho)^2}{2} \leq \frac{1}{2}. \quad (5)$$

Substituting (5) into (1), the probability that a bad project remains uncertain is

$$\varphi^*(\lambda) = 1 - \rho + \lambda\rho^2. \quad (6)$$

Equation (6) drives the rest of the analysis. Without obfuscation the bad project stays uncertain with probability $1 - \rho$; obfuscation adds the term $\lambda\rho^2$, which rises with the resources ρ that the agency invests and with the probability λ that an uncertain application is approved. Obfuscation is worthwhile only if approval is possible and only if there is examination to evade: if $\lambda = 0$ or $\rho = 0$, the effort is zero.

4. Step (c): The Agency's Approval Decision and Payoff

Only applicants with bad projects obfuscate, so, when $\lambda\rho > 0$, uncertain applications contain a larger share of bad projects than the prior. An uncertain agency puts weight $\frac{1}{2}(1 - \rho)$ on a good project (a good applicant stays uncertain with probability $1 - \rho$) and weight $\frac{1}{2}\varphi^*(\lambda)$ on a bad project, so approving yields $\frac{1}{2}(1 - \rho) \cdot 6 - \frac{1}{2}\varphi^*(\lambda) \cdot 2$ in expectation. It is willing to approve if and only if

$$\Delta(\lambda) \equiv 3(1 - \rho) - (1 - \rho + \lambda\rho^2) = 2(1 - \rho) - \lambda\rho^2 \geq 0. \quad (7)$$

The function Δ is strictly decreasing in λ for $\rho > 0$: the more the agency approves, the more the bad applicants obfuscate, and the worse the uncertain pool becomes. Three cases follow.

1. If $\Delta(1) \geq 0$, then $\Delta(\lambda) > 0$ for every $\lambda < 1$, so the agency strictly prefers to approve unless $\lambda = 1$, and the only equilibrium is $\lambda = 1$. Rearranging, $\Delta(1) \geq 0$ is equivalent to $\rho^2 + 2\rho - 2 \leq 0$, which holds if and only if $\rho \leq \hat{\rho}$, where

$$\hat{\rho} \equiv \sqrt{3} - 1 = 0.732 \quad (8)$$

is the positive root of that quadratic.

2. If $\Delta(1) < 0$ and $\rho < 1$, then $\lambda = 1$ is not a best reply, and $\Delta(0) = 2(1 - \rho) > 0$, so $\lambda = 0$ is not either. The agency mixes, which requires $\Delta(\lambda) = 0$, and the unique solution is

$$\lambda = \frac{2(1 - \rho)}{\rho^2} \in (0, 1). \quad (9)$$

3. At $\rho = 1$ a good project is never uncertain, $\Delta(\lambda) = -\lambda$, and the agency rejects, $\lambda = 0$: if it approved with positive probability, bad applicants would obfuscate, uncertain applications would come only from bad projects, and approving would be strictly unprofitable.

The agency's payoff from examination is $U = 3[\rho + (1 - \rho)\lambda] - \lambda\varphi^*(\lambda) = 3\rho + \lambda\Delta(\lambda)$. For $\lambda = 1$ this equals $3 - (1 - \rho + \rho^2)$, and when $\lambda < 1$ either $\lambda = 0$ or $\Delta(\lambda) = 0$, so $U = 3\rho$:

$$U(\rho) = \begin{cases} 2 + \rho - \rho^2 & \text{if } \rho \leq \hat{\rho}, \\ 3\rho & \text{if } \rho > \hat{\rho}. \end{cases} \quad (10)$$

Result 2 (equilibrium for given resources; Figueroa and Lemus, 2024, specialized). The equilibrium outcome is unique, and only bad applicants obfuscate, with effort $e = \frac{(\lambda\rho)^2}{2}$. The agency

- approves every uncertain application if $\rho \leq \hat{\rho}$;
- approves uncertain applications with the probability (9) if $\hat{\rho} < \rho < 1$;
- rejects every uncertain application if $\rho = 1$.

Compared with (2), $U = U_0 - \rho^2$ whenever $\rho \leq \hat{\rho}$, so ρ^2 is the loss that obfuscation causes. Above $\hat{\rho}$ the agency is indifferent when it is uncertain, so obfuscation no longer affects its payoff, and $U = 3\rho$ is the value of the good projects that the agency identifies, because uncertain applications yield zero in expectation.

5. Step (d): When More Resources Lower the Payoff

Differentiating (10),

$$U'(\rho) = 1 - 2\rho \text{ for } \rho \leq \widehat{\rho}, \quad U'(\rho) = 3 \text{ for } \rho > \widehat{\rho}, \quad (11)$$

and U is continuous at $\widehat{\rho}$ because $\Delta(1) = 0$ there. The slope below $\widehat{\rho}$ has a simple decomposition. With $\lambda = 1$, the probability that a bad project stays uncertain is $\varphi^* = 1 - \rho + \rho^2$, so $\frac{d\varphi^*}{d\rho} = -1 + 2\rho$: a unit of resources lowers the probability by one, which gains 1, but it raises obfuscation, so the probability rises by 2ρ , which costs 2ρ . Since every good project is approved anyway when $\rho \leq \widehat{\rho}$, extra resources screen out only bad projects, and the net effect $1 - 2\rho$ is negative once ρ exceeds one half.

The dip exists because $\widehat{\rho} = 0.732 > \frac{1}{2}$. Evaluating $\Delta(1)$ at $\rho = \frac{1}{2}$ gives $2 \cdot \frac{1}{2} - \frac{1}{4} = \frac{3}{4} > 0$, so the agency still approves every uncertain application at one half, and since $\Delta(1)$ is strictly decreasing in ρ , $\widehat{\rho}$ lies above one half. At $\rho = \frac{1}{2}$ the payoff equals $U(\frac{1}{2}) = 2 + \frac{1}{2} - \frac{1}{4} = 2.25$, which can be written as $3\widetilde{\rho}$ with

$$\widetilde{\rho} \equiv \frac{3}{4} \in (\widehat{\rho}, 1). \quad (12)$$

Result 3 (non-monotone payoff; Figueroa and Lemus, 2024, specialized). The payoff U is

- strictly increasing on $[0, \frac{1}{2}]$,
- strictly decreasing on $[\frac{1}{2}, \widehat{\rho}]$, and
- strictly increasing on $[\widehat{\rho}, 1]$.

Moreover, $U(\rho) < U(\frac{1}{2})$ for every $\rho \in (\frac{1}{2}, \widetilde{\rho})$, and $U(1) = 3 > U(\frac{1}{2})$.

Figure 1 depicts $U(\rho)$, with the benchmark (2) as a dashed line. The payoff rises to a first peak $U(\frac{1}{2}) = 2.25$, falls to $U(\widehat{\rho}) = 2.196$ at $\widehat{\rho} = 0.732$, where the agency starts to mix, and then rises at the rate 3. It exceeds its first peak only for $\rho > \widetilde{\rho} = \frac{3}{4}$. The two peaks, the interval between them, and the kink at $\widehat{\rho}$ organize the rest of the note.

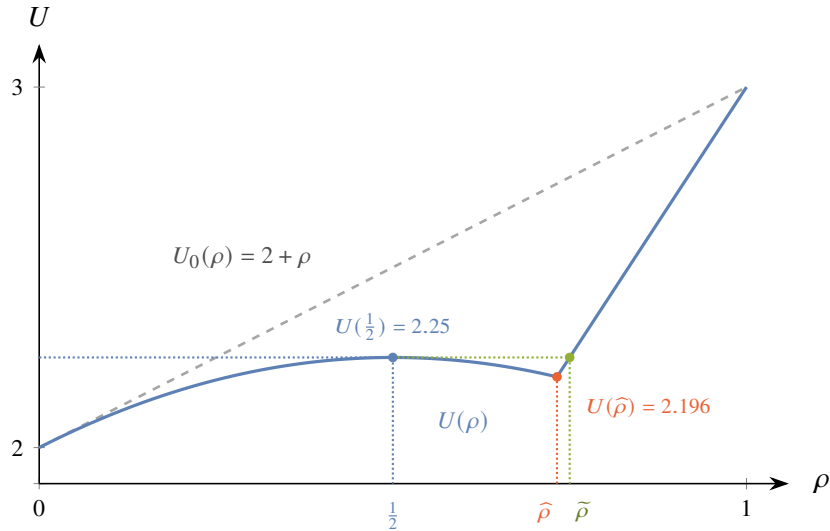


Figure 1: Payoff from examination $U(\rho)$ and the benchmark without obfuscation $U_0(\rho) = 2 + \rho$.

6. Step (e): The Optimal Budget

Throughout this step resources cost $k\rho^2$, and only the unit cost $k \geq 0$ and the budget ρ are free. By Result 3, the payoff from examination in Figure 1 rises to a first peak $U(\frac{1}{2}) = 2.25$, falls until $\widehat{\rho} = 0.732$, and rises again to $U(1) = 3$. The agency chooses $\rho \in [0, 1]$ to maximize its net payoff

$$\Pi_k(\rho) \equiv U(\rho) - k\rho^2, \quad (13)$$

which is continuous, so a maximizer exists. Figure 2 depicts Π_k for two unit costs. Each curve has two peaks: a modest one, below one half, and a high one, at full resources.

- At $k = \frac{1}{2}$ the high peak is higher: $\Pi_k = 2.500$ at $\rho = 1$ against 2.167 at $\rho = \frac{1}{3}$.
- At $k = 1$ the ranking reverses: $\Pi_k = 2.000$ at $\rho = 1$ against 2.125 at $\rho = \frac{1}{4}$.

The following three parts show that this picture holds for every unit cost.

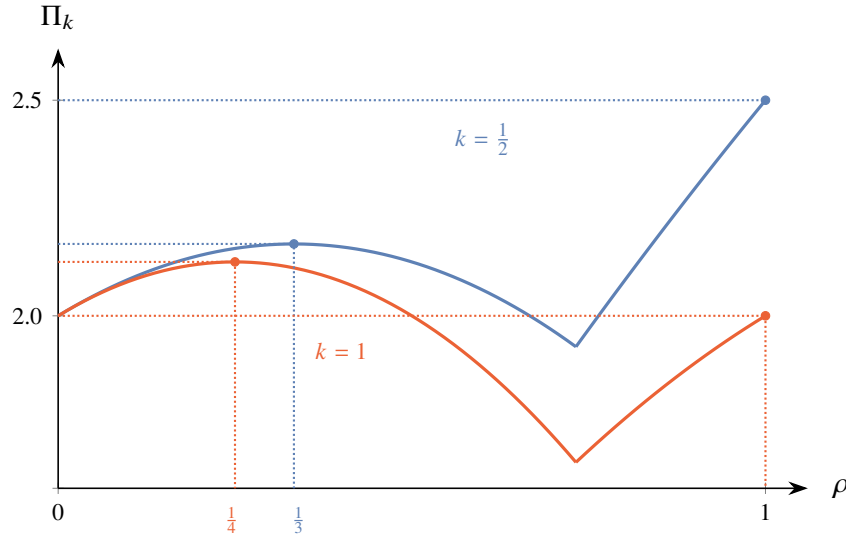


Figure 2: Net payoff $\Pi_k(\rho) = U(\rho) - k\rho^2$ for $k = \frac{1}{2}$ and $k = 1$.

(i) Budgets where more resources hurt are never chosen. On $[\frac{1}{2}, \widehat{\rho}]$, U is strictly decreasing and $k\rho^2$ is weakly increasing, so Π_k is strictly decreasing there, and $\Pi_k(\frac{1}{2}) > \Pi_k(\rho)$ for every $\rho \in (\frac{1}{2}, \widehat{\rho}]$. Cutting such a budget to one half raises the payoff from examination and lowers the cost of resources. The optimal budget therefore lies on the *modest* branch $[0, \frac{1}{2}]$ or on the *high* branch $(\widehat{\rho}, 1]$, whose values are

$$L(k) \equiv \max_{\rho \in [0, \frac{1}{2}]} \Pi_k(\rho), \quad H(k) \equiv \max_{\rho \in [\widehat{\rho}, 1]} \Pi_k(\rho). \quad (14)$$

(ii) The budget falls weakly as the unit cost rises. Let ρ be optimal at k and ρ' be optimal at $k' > k$. Optimality gives $U(\rho) - k\rho^2 \geq U(\rho') - k\rho'^2$ and $U(\rho') - k'\rho'^2 \geq U(\rho) - k'\rho^2$. Adding the two inequalities, the terms in U cancel and $(k' - k)(\rho^2 - \rho'^2) \geq 0$, so $\rho' \leq \rho$.

(iii) The budget jumps once. Let $D(k) \equiv H(k) - L(k)$ be the advantage of the high branch, which is continuous in k . At $k = 0$ each branch attains its peak, $H(0) = U(1) = 3$ and $L(0) = U(\frac{1}{2}) = 2.25$, so $D(0) = 0.75 > 0$. When the unit cost rises from k to k' , the value of a branch falls by $k' - k$ times the squared budget that it uses, and the high branch uses at least $\widehat{\rho}^2 = 0.536$ while the modest branch uses at most $\frac{1}{4}$.

Formally, take a maximizer ρ_ℓ of Π_k on the modest branch and a maximizer ρ'_u of $\Pi_{k'}$ on the high branch. Then $H(k') = \Pi_k(\rho'_u) - (k' - k)\rho_u'^2 \leq H(k) - (k' - k)\rho_u'^2$ and $L(k') \geq \Pi_{k'}(\rho_\ell) = L(k) - (k' - k)\rho_\ell^2$, so

$$D(k') \leq D(k) - (k' - k)(\rho_u'^2 - \rho_\ell^2) < D(k), \quad (15)$$

because $\rho_u'^2 \geq \widehat{\rho}^2 > \frac{1}{4} \geq \rho_\ell^2$. Hence D is strictly decreasing. Finally, $H(k) \leq U(1) - k\widehat{\rho}^2 = 3 - k\widehat{\rho}^2$ and $L(k) \geq \Pi_k(0) = U(0) = 2$, so $D(k) \leq 1 - k\widehat{\rho}^2 < 0$ for every $k > \frac{1}{\widehat{\rho}^2} = 1.866$. Therefore D has a unique zero $\widehat{k}$ between 0 and 1.866: the high branch is better below $\widehat{k}$ and the modest branch above it. Figure 2 places $\widehat{k}$ between $\frac{1}{2}$ and 1, and Step (f) shows that $\widehat{k} = 0.866$.

Result 4 (the optimal budget). Let the cost of resources be $k\rho^2$.

- (i) For every $k \geq 0$, no budget in $(\frac{1}{2}, \widehat{\rho}]$ is optimal.
- (ii) The optimal budget is weakly decreasing in k .
- (iii) There is a unique $\widehat{k} > 0$ such that every optimal budget lies in $(\widehat{\rho}, 1]$ if $k < \widehat{k}$ and in $[0, \frac{1}{2}]$ if $k > \widehat{k}$, and both branches contain an optimal budget at $k = \widehat{k}$.

Intuitively, the payoff has a first peak at one half and a second peak at full resources, and the cost of resources penalizes the second peak more because it is farther out. The value of the high branch therefore falls faster in k than the value of the modest branch, and the two values cross once. At $\widehat{k}$ the budget jumps from above $\widehat{\rho} = 0.732$ to at most one half, a fall of more than 0.232. Compare with Result 1: the benchmark payoff U_0 is linear in ρ , so the net payoff $U_0(\rho) - k\rho^2$ is strictly concave and its maximizer moves continuously with k .

Beyond the example. Parts (i) to (iii) used only two facts:

- the payoff U rises to a first peak a , falls until some $b > a$, and rises again to $U(1) > U(a)$;
- the cost of resources is k times a strictly increasing function $\psi(\rho)$ with $\psi(0) = 0$.

Replacing $\frac{1}{2}$ and $\widehat{\rho}$ by a and b , and ρ^2 by $\psi(\rho)$, every step goes through unchanged. Result 4 therefore holds for any strictly increasing cost, such as the power costs $k\rho^m$ of Step (h), and for any values of p , v , and c for which the payoff has the dip of Result 3 (Step h).

7. Step (f): Closed Form with a Quadratic Cost

By Result 4, the optimal budget lies on the modest branch $[0, \frac{1}{2}]$ or on the high branch $(\widehat{\rho}, 1]$. On the modest branch the objective is $2 + \rho - (1 + k)\rho^2$, a strictly concave quadratic with the vertex $\rho_1 \equiv \frac{1}{2(1+k)} \leq \frac{1}{2}$, so

$$L(k) = 2 + \frac{1}{4(1+k)}. \quad (16)$$

On the high branch the objective is $3\rho - k\rho^2$, whose derivative $3 - 2k\rho$ is positive for $\rho < 1$ whenever $k \leq \frac{3}{2}$, so the maximizer is $\rho = 1$ and

$$H(k) = 3 - k \quad (k \leq \frac{3}{2}). \quad (17)$$

The difference is

$$D(k) = 1 - k - \frac{1}{4(1+k)} = \frac{3 - 4k^2}{4(1+k)}, \quad (18)$$

which has the unique positive zero

$$\bar{k} \equiv \frac{\sqrt{3}}{2} = 0.866. \quad (19)$$

The zero is a valid jump point because $\bar{k} \leq \frac{3}{2}$.

Result 5 (closed form). The jump point is $\hat{k} = \bar{k}$, and

- if $k < \bar{k}$, the optimal budget is $\rho^* = 1$;
- if $k > \bar{k}$, the optimal budget is $\rho^* = \frac{1}{2(1+k)} < 2 - \sqrt{3}$;
- if $k = \bar{k}$, both $\rho = 1$ and $\rho = 2 - \sqrt{3}$ are optimal.

The set of optimal budgets over all $k \geq 0$ is therefore $(0, 2 - \sqrt{3}] \cup \{1\}$.

Here $2 - \sqrt{3} = 0.268$. Figure 3 depicts the two values in panel (a) and the optimal budget in panel (b). In panel (a), the value $H(k) = 3 - k$ of full resources falls faster than the value $L(k)$ of the modest budget, and the lines cross at $\bar{k}$. In panel (b), the budget equals one up to $\bar{k}$, falls discontinuously to 0.268, and then falls gradually, to 0.222 at $k = 1.25$ and to 0.167 at $k = 2$.

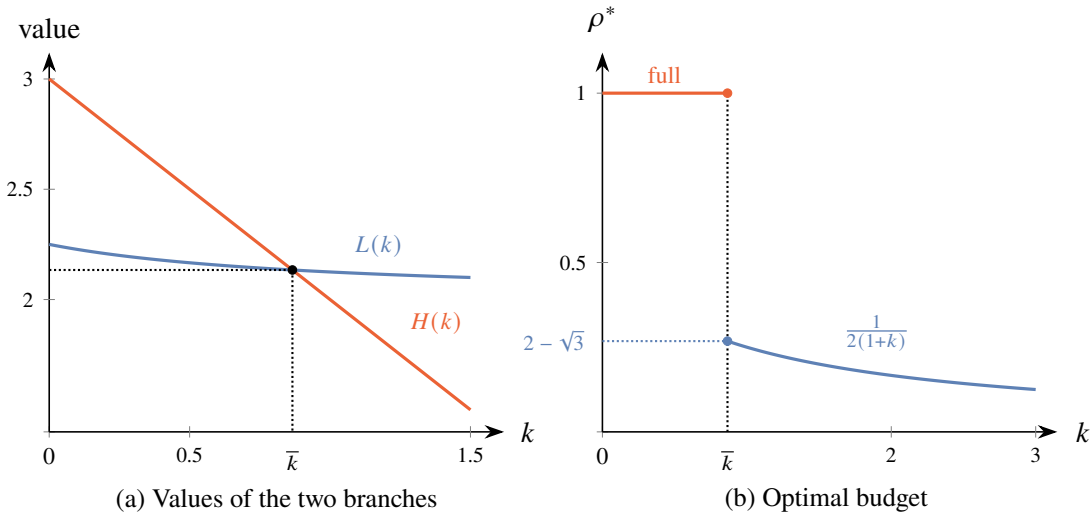


Figure 3: Values of the two branches and optimal budget against the unit cost k .

Why the value of good projects does not matter for the cutoff. The cutoff (19) comes from (18), which contains the cost 1 of a bad project and no trace of the value 3 of good projects. At full resources the agency learns every type and approves every good project, and at the modest budget it approves every uncertain application and therefore every good project too. Both branches collect the same 3, so it cancels from $D(k)$; the choice between them turns on the cost of the bad projects that the modest budget lets through. Step (h) shows how the cutoff changes with that cost.

The equilibrium behind the budget. Table 1 reports the equilibrium outcome for k from 0.25 to 3 in steps of 0.25. The columns show the optimal budget ρ^* ; the effort e^* of a bad applicant; the probability $1 - \varphi(\rho^*, e^*)$ that the agency learns the type of that applicant; the approval probability λ^* of an uncertain application; the expected value $V^* \equiv 3[\rho^* + (1 - \rho^*)\lambda^*]$ obtained from good projects; the expected cost $C^* \equiv \lambda^*\varphi(\rho^*, e^*)$ of the bad projects that it approves; the payoff $U = V^* - C^*$; and the payoff $U - k(\rho^*)^2$ net of the cost of resources.

k	ρ^*	e^*	$1 - \varphi(\rho^*, e^*)$	λ^*	V^*	C^*	U	$U - k(\rho^*)^2$
0.25	1.000	0.000	1.000	0.000	3.000	0.000	3.000	2.750
0.50	1.000	0.000	1.000	0.000	3.000	0.000	3.000	2.500
0.75	1.000	0.000	1.000	0.000	3.000	0.000	3.000	2.250
1.00	0.250	0.031	0.188	1.000	3.000	0.812	2.188	2.125
1.25	0.222	0.025	0.173	1.000	3.000	0.827	2.173	2.111
1.50	0.200	0.020	0.160	1.000	3.000	0.840	2.160	2.100
1.75	0.182	0.017	0.149	1.000	3.000	0.851	2.149	2.091
2.00	0.167	0.014	0.139	1.000	3.000	0.861	2.139	2.083
2.25	0.154	0.012	0.130	1.000	3.000	0.870	2.130	2.077
2.50	0.143	0.010	0.122	1.000	3.000	0.878	2.122	2.071
2.75	0.133	0.009	0.116	1.000	3.000	0.884	2.116	2.067
3.00	0.125	0.008	0.109	1.000	3.000	0.891	2.109	2.062

Table 1: Equilibrium outcome as the unit cost k rises.

Table 1 splits into two regimes:

- For $k \leq 0.75$, below $\bar{k} = 0.866$, the agency uses full resources, learns every type, and obtains $U = 3$ without any obfuscation.
- From $k = 1$ on, the budget is modest and falls with k , from 0.250 to 0.125. Bad applicants now obfuscate, but their effort falls with the budget, and the agency learns their type with a lower probability, from 0.188 to 0.109. The agency approves every uncertain application, so $V^* = 3$ in every row, and the cost C^* of the bad projects that it approves rises from 0.812 to 0.891, which lowers U from 2.188 to 2.109.

The jump lowers the payoff from examination from 3 to 2.188, because the agency now approves bad projects that it would have rejected, but the saving in the cost of resources makes the net payoff fall only from 2.250 (at $k = 0.75$) to 2.125 (at $k = 1$), and the net payoff falls continuously in k afterwards.

8. Step (g): Two Policy Implications

A budget imposed from outside can lower the agency's payoff. By Result 4(i), the agency never chooses a budget in $(\frac{1}{2}, \widehat{\rho}]$. An outside budget can, and it need not help. At $k = 1$ the agency chooses $\rho^* = 0.25$ and obtains the net payoff 2.125. A floor of $\rho \geq \frac{1}{2}$, such as a hiring target, leaves the agency with at most 2.000: on $[\frac{1}{2}, \widehat{\rho}]$ the net payoff falls from 2.000, and on $[\widehat{\rho}, 1]$ it rises back to 2.000 at $\rho = 1$. In contrast, a lower unit cost leads the agency to choose a high budget on its own: at $k = \frac{1}{2}$ it chooses $\rho^* = 1$ without any floor. A fee that is independent of the resources the agency spends is a lump sum and leaves k , and hence the budget, unchanged.

A small change in how the agency is funded can move its budget from full to modest. At $k = 0.85$ the agency chooses $\rho^* = 1$ and obtains the net payoff 2.150. At $k = 0.90$, an increase of 6 percent in the unit cost, it chooses $\rho^* = 0.263$, which is 74 percent lower, and obtains 2.132. Differences in budgets across agencies or years therefore need not reflect different valuations of examination, and a small difference in the cost at which agencies obtain resources can produce a large difference in how much they examine.

9. Step (h): Other Parameter Values, and What the Toy Model Keeps and Drops

Other values of p , v , and c . Write $G \equiv pv$ for the expected value of a good project and $B \equiv (1-p)c$ for the expected cost of a bad project, so that the numbers of Section 1 are $G = 3$ and $B = 1$. Repeating Steps (a) to (f) with general $G > B$ gives $U_0(\rho) = G - B(1 - \rho)$, $\hat{\rho} = \frac{-(G-B) + \sqrt{(G-B)^2 + 4B(G-B)}}{2B}$, and $U(\rho) = G - B(1 - \rho + \rho^2)$ for $\rho \leq \hat{\rho}$ and $G\rho$ above. Two conditions reproduce the results:

- the payoff has a dip if and only if $G > \frac{3B}{2}$: $\Delta(1)$ at $\rho = \frac{1}{2}$ equals $\frac{G}{2} - \frac{3B}{4}$, and $\tilde{\rho} = 1 - \frac{3B}{4G}$. If $G \leq \frac{3B}{2}$, U is strictly increasing on $[0, 1]$ and Result 3 does not hold;
- the budget jumps from full to modest if $G \geq \sqrt{3}B$, with

$$L(k) = G - B + \frac{B^2}{4(B+k)}, \quad H(k) = G - k, \quad \bar{k} = \frac{\sqrt{3}}{2} B, \quad (20)$$

and $\rho^* = \frac{B}{2(B+k)}$ above $\bar{k}$.

The cutoff $\bar{k}$ is proportional to the expected cost B of a bad project and does not depend on G . Any values that satisfy the two conditions, such as $G = 2$ and $B = 1$, or the values $G = 1.4$ and $B = 0.3$ of Figueroa and Lemus (2024), give the same qualitative results.

The condition $G \geq \sqrt{3}B$ cannot be relaxed. Evaluating $D(k) = G - k - L(k)$ at $k = \frac{G}{2}$ gives $D(\frac{G}{2}) = \frac{3B^2 - G^2}{2(G+2B)}$, which is positive if $G < \sqrt{3}B$. Then $\hat{k} > \frac{G}{2}$, and for $k \in (\frac{G}{2}, \hat{k})$ the high branch is still better but its maximizer is interior, $\rho = \frac{G}{2k} < 1$, so the agency uses neither full resources nor a modest budget. For $G = 1.6$ and $B = 1$, $\hat{k} = 0.873$ exceeds $\frac{G}{2} = 0.8$, and the optimal budget just below $\hat{k}$ is 0.917.

A common unit cost can lead to full examination of high-stakes cases and modest examination of low-stakes cases. Because $\bar{k} = \frac{\sqrt{3}}{2}B$ rises with the expected cost of a bad project and does not depend on G , applications that differ in B have different cutoffs. Consider two kinds of applications, both with $G = 3$, and a common unit cost $k = 1$:

- low-stakes applications, with $B = \frac{1}{2}$ (for example $c = 1$), have the cutoff 0.433, so at $k = 1$ the agency examines them with the modest budget $\frac{1/2}{2(1/2+1)} = 0.167$;
- high-stakes applications, with $B = \frac{3}{2}$ (for example $c = 3$), have the cutoff 1.299, so at $k = 1$ the agency examines them with full resources.

A rise in the unit cost therefore first cuts examination where bad approvals are less costly.

Four further variations show how much of the jump survives.

Obfuscation as waste. Let the agency's objective also subtract the effort of bad applicants, at the expected cost $(1-p)e$. This lowers the payoff by $w\rho^2$, with $w \equiv \frac{1-p}{2}$, when $\rho \leq \hat{\rho}$, by less between $\hat{\rho}$ and one, and by nothing at full resources, where nobody obfuscates. The payoff keeps the two-peak shape used in Step (e), with the first peak at $\frac{B}{2(B+w)}$, and the cutoff and the modest budget become

$$\bar{k}_w = \frac{\sqrt{(w+B)(w+3B)} - w}{2}, \quad \rho^* = \frac{B}{2(B+w+k)} \quad (k > \bar{k}_w), \quad (21)$$

provided $\bar{k}_w \leq \frac{G}{2}$. The cutoff satisfies $\bar{k} < \bar{k}_w < B$, so waste raises it by less than the factor $\frac{2}{\sqrt{3}} = 1.155$. With $p = \frac{1}{2}$, $w = \frac{1}{4}$: the cutoff rises from 0.866 to 0.883, the budget just above it is 0.234 instead of 0.268,

and at $k = 1.25$ the budget falls from 0.222 to 0.200. The waste falls only on the modest budget, which induces obfuscation, whereas full resources induce none.

Other cost functions. Result 4 holds for every strictly increasing cost (see the end of Step (e)), but the size of the jump depends on the cost. Table 2 reports, for $K(\rho) = k\rho^m$, the unit cost $\hat{k}$ at which the optimal budget moves from the high branch to the modest branch, and the budgets just below and just above $\hat{k}$:

- with a linear cost ($m = 1$), the agency jumps from full resources to no resources at $\hat{k} = B = 1$: the modest branch is $\rho = 0$ when $k \geq B$, so $L = G - B$, $H = G - k$, and $D = B - k$;
- with $m = 2$ and $m = 3$, the agency jumps from full resources to a modest budget;
- with steeper costs ($m \geq 4$), the budget just below $\hat{k}$ is already below one, because the last units of resources are very costly, but the budget still falls discontinuously at $\hat{k}$.

The budget just above the jump rises with the convexity of the cost.

m	1	2	3	4	5	6
Jump point $\hat{k}$	1.000	0.866	0.807	0.781	0.800	0.865
Budget just below $\hat{k}$	1	1	1	0.987	0.931	0.896
Budget just above $\hat{k}$	0	0.268	0.351	0.400	0.431	0.451

Table 2: Jump point and budgets on either side of it for the cost $K(\rho) = k\rho^m$, with $G = 3$ and $B = 1$.

A dip is sufficient for the jump but not necessary. Let the applicant's gain from approval be $g \in (\frac{1}{2}, 1]$ instead of one, so that the effort becomes $e = \frac{(g\lambda\rho)^2}{2}$. The payoff has a dip if and only if $G > \frac{B(4g-1)}{2(2g-1)}$. With $G = 3$ and $B = 1$:

- for $g = 0.75$ the threshold is $2 < 3$, so the payoff has a dip; the cutoff is 0.843 and the budget after the jump is 0.314;
- for $g = 0.55$ the threshold is $6 > 3$, so the payoff has no dip. Yet the optimal budget still jumps, from 1 to 0.366 at $k \approx 0.817$, because the payoff has a convex kink at $\hat{\rho}$, where the agency switches from approving every uncertain application to mixing.

The jump therefore does not require the dip: the convex kink at which the agency's behavior changes is enough.

Capacity. If the feasible resources are capped at $\bar{\rho} < 1$, so that examination is never perfect, the cutoff depends on G : for $B = 1$ and $\bar{\rho} = 0.95$ it is 0.787 at $G = 3$ and 0.603 at $G = 6$, and the jump disappears for $G \geq \frac{3B}{4(1-\bar{\rho})} = 15$. Independence of the cutoff from G is therefore a property of full capacity.

What the toy model drops. Figueroa and Lemus (2024) derive the equilibrium for a general probability φ and write down the budget problem for a convex cost of resources, but they do not characterize its solution. The toy model uses the single specification (1) and a quadratic (or power) cost, which yield closed forms. It has one applicant, a one-shot decision, and resources that the applicant observes before choosing its effort. The agency pays for resources through a cost that does not depend on who applies, and no fee changes the pool of applicants. Each of these features can be relaxed one at a time. A natural extension is a self-funded office, in which a fee sets who applies and the unit cost of resources at the same time.

References

Figuerola, N., and J. Lemus (2024). “The Benefit of the Doubt: Patent Examination under Strategic Obfuscation.” *Management Science*, 70(9), pp. 5853–5870.

The Attention Economy

A step-by-step exercise in the spirit of Chen (2026)

Companion note to Chen, D. (2026), “A Model of the Attention Economy: Data, Platforms, and Discovery,” *American Economic Review* 116(10): 3802–3836. Full references at the end.

Overview. Social media, streaming services, and podcasts are free to use. They earn their revenue by selling targeted ads, and the ads determine which products consumers discover. Chen (2026) builds a model with two linked markets. In the product market, consumers buy only from the firms they are aware of. In the attention market, platforms invest in the quality of their free services to attract attention, and sell ads through auctions in which firms bid using the platforms’ data on consumer preferences. The paper shows that the two markets cannot be studied separately: more informative data raises product consumption but tends to lower ad revenue and platform investment, interoperability mandates lower product consumption but raise investment, the effects of both policies can change sign over time, and platform investment can be too high or too low. This note builds these results in a version of the model with closed-form solutions. Step (a) solves the product market; Step (b) solves the attention market and the evolution of consideration sets; Step (c) solves the ad auctions and platform investment; Step (d) studies data policy; Step (e) studies interoperability; Step (f) compares the equilibrium with the planner’s allocation; and Step (g) lists what the toy model keeps and drops.

1. The Toy Model

Consumers and products. Time is continuous, $t \geq 0$. A unit mass of consumers has income $I = 1$ per unit of time. There is a large pool of products, each sold by a different firm at marginal cost 1. Consumer i values product j at v_{ij} , drawn independently from the uniform distribution on $[0, 1]$. Consumer i is aware of only a subset of products, her consideration set Ω_{it} , and buys only from it. Over the products in this set she has CES preferences with elasticity of substitution $\sigma = 2$,

$$C_{it} = \left[\int_{\Omega_{it}} v_{ij}^{\frac{1}{2}} c_{ijt}^{\frac{1}{2}} dj \right]^2. \quad (1)$$

Only two features of Ω_{it} matter: its mass M_t and the average value μ_t of the products in it. Both are common across consumers.

Platforms and attention. A unit mass of platforms, $K = 1$, offers free services. Each consumer has one unit of attention per unit of time and divides it across platforms to maximize

$$X_{it} = \left[\int_0^1 (v(a_k) q_k x_{ik})^{\frac{\epsilon-1}{\epsilon}} dk \right]^{\frac{\epsilon}{\epsilon-1}}, \quad (2)$$

where x_{ik} is the attention she gives to platform k , q_k is the platform’s quality, a_k is its ad rate, and $v(a) = 1 - a$ captures that ads are a nuisance. The elasticity $\epsilon \in (1, 3)$ measures how sensitive attention is to effective quality $v(a_k)q_k$; interoperability mandates raise ϵ . Quality depreciates at rate $\delta = 1$ and grows with investment ℓ_k ,

$$\dot{q}_k = \sqrt{\ell_k} - q_k, \quad (3)$$

so investment has decreasing returns, with the paper’s parameter $\varphi = \frac{1}{2}$.

Ads and discovery. A consumer who gives attention x_{ik} to platform k sees ads at rate $a_k x_{ik}$. Each ad is sold the instant it arises, in a second-price auction among $N = 5$ firms drawn at random from the products outside her consideration set. The winner enters the set, and every product in the set is forgotten at rate $\lambda_f = 1$. All firms and platforms discount the future at rate $\rho = 1$.

Data. Firms do not observe v_{ij} . The platforms' data give each bidder an expected value $\hat{v}_{ij}$, drawn from the uniform distribution G on $[\frac{1}{2} - s, \frac{1}{2} + s]$ with $s \in [0, \frac{1}{2}]$. With $s = 0$, data is uninformative and every firm expects the mean value $\frac{1}{2}$; with $s = \frac{1}{2}$, data is perfectly informative and $\hat{v}_{ij} = v_{ij}$. A higher s is a mean-preserving spread of G , that is, more informative data.

A large pool of products. The paper has a finite mass J of products. We let $J \rightarrow \infty$, which has two implications. First, the products outside a consideration set are a representative sample, so the five bidders in each auction have expected values drawn independently from G . Second, a product that a consumer forgets is invited to bid for her again only at rate $\frac{NA}{J} \approx 0$, so it treats being forgotten as permanent. Step (g) discusses what this simplification drops.

2. Step (a): The Product Market

Maximizing (1) subject to the budget constraint gives consumer i 's demand for product $j \in \Omega_{it}$,

$$c_{ij} = \frac{v_{ij} p_j^{-2}}{\int_{\Omega_{it}} v_{iz} p_z^{-1} dz}. \quad (4)$$

The value v_{ij} and the consideration set only scale demand, so firm j maximizes $(p_j - 1)p_j^{-2}$ whatever it knows about the consumer. The first-order condition, $p_j^{-2} - 2(p_j - 1)p_j^{-3} = 0$, gives $p_j = 2 = \frac{\sigma}{\sigma-1}$. Substituting $p = 2$ in (4), $c_{ij} = \frac{v_{ij}}{2M_t\mu_t}$, so the firm's flow profit from consumer i and the consumer's aggregate consumption are

$$(p - 1) c_{ij} = \frac{v_{ij}}{2M_t\mu_t}, \quad C_{it} = \frac{1}{2} M_t\mu_t. \quad (5)$$

Result 1 (product market, Proposition 1 of the paper). Every firm charges a price of 2, whatever it knows about consumers. A firm in a consideration set earns $\frac{v}{2M_t\mu_t}$ per unit of time from that consumer, and the firms in the set earn a total of $\frac{1}{2} = \frac{I}{\sigma}$, whatever the size and composition of the set. Aggregate consumption $\frac{1}{2} M_t\mu_t$ increases in both the mass and the average value of the products the consumer knows.

Intuitively, a consumer who knows more products, or better ones, consumes more, but firms' total profit stays the same: each firm faces more, or better matched, rivals for the consumer's budget. This fixed profit pie is what the attention market competes for.

3. Step (b): Attention, Ad Load, and Consideration Sets

Attention. Maximizing (2) subject to $\int_0^1 x_{ik} dk \leq 1$ gives the attention share

$$x_k = \frac{[v(a_k)q_k]^{\epsilon-1}}{\int_0^1 [v(a_z)q_z]^{\epsilon-1} dz}. \quad (6)$$

Ad rate. Let π_K denote the expected price of an ad. Platform k earns $\pi_K a_k x_k$ per unit of time, and since it is small, it takes the denominator of (6) as given and chooses a_k to maximize $a(1-a)^{\epsilon-1}$. The first-order condition, $(1-a) - (\epsilon-1)a = 0$, gives the ad rate

$$A = \frac{1}{\epsilon}. \quad (7)$$

All platforms choose the same ad rate and, in a symmetric equilibrium, the same quality, so each receives attention $x_k = 1$, and consumers see ads at rate A .

Consideration sets. Products enter a consideration set at rate A and leave at rate $\lambda_f M_t = M_t$, so $\dot{M}_t = A - M_t$ and M_t converges to $M = A$. The winner of each auction is the bidder with the highest expected value, the maximum of five draws from G , and products leave at random, so in steady state the average value in the set is the expected maximum of five draws. For the uniform distribution on $[\frac{1}{2} - s, \frac{1}{2} + s]$, the expected highest and second-highest of five draws are

$$\mu_H = E[\hat{v}_{(1)}] = \frac{1}{2} + \frac{2s}{3}, \quad E[\hat{v}_{(2)}] = \frac{1}{2} + \frac{s}{3}. \quad (8)$$

Result 2 (ad load and discovery, Theorem 1 of the paper). In steady state:

- every platform shows ads at rate $A = \frac{1}{\epsilon}$, which decreases in the elasticity of attention;
- consideration sets have mass $M = \frac{1}{\epsilon}$ and average value $\mu_H = \frac{1}{2} + \frac{2s}{3}$, which increases in data informativeness.

Interoperability therefore changes how many products consumers discover, and data changes which ones. By (5), steady-state consumption is $C = \frac{\mu_H}{2\epsilon}$.

4. Step (c): Ad Auctions and Platform Investment

Bids. In a second-price auction, a firm bids the value of joining the consideration set. A member with expected value $\hat{v}$ earns $\frac{\hat{v}}{2M\mu_H}$ per unit of time until it is forgotten, and a firm that loses earns nothing from this consumer, so, discounting at $\rho + \lambda_f = 2$,

$$B(\hat{v}) = \frac{\hat{v}}{4M\mu_H}. \quad (9)$$

Ad revenue. The winner pays the second-highest bid, so the expected ad price is $\pi_K = \frac{E[\hat{v}_{(2)}]}{4M\mu_H}$. Total ad revenue is $R = \pi_K A$ and, since $M = A$,

$$R = \underbrace{\frac{I}{\sigma}}_{=\frac{1}{2}} \cdot \underbrace{\frac{\lambda_f}{\rho + \lambda_f}}_{=\frac{1}{2}} \cdot \frac{E[\hat{v}_{(2)}]}{E[\hat{v}_{(1)}]} = \frac{1}{4} \cdot \frac{3 + 2s}{3 + 4s}. \quad (10)$$

This is the toy version of equation (17) of the paper. Ad revenue is a share of the fixed profit pie $\frac{I}{\sigma}$, and the share falls with the ratio of the second-highest to the highest expected value in an auction, which measures the winner's information rent.

Investment. Platform k chooses ℓ_k to maximize $\int_0^\infty e^{-t} (\pi_K A x_k - \ell_k) dt$ subject to (3). By (6), a higher q_k raises the platform's revenue at rate $\frac{(\epsilon-1)R}{q_k}$ in a symmetric equilibrium, so in steady state the shadow value of quality is $\lambda = \frac{(\epsilon-1)R}{(\rho+\delta)q} = \frac{(\epsilon-1)R}{2q}$. The first-order condition for investment, $1 = \frac{\lambda}{2\sqrt{\ell}}$, together with $q = \sqrt{\ell}$ from (3), gives

$$\ell = \frac{(\epsilon-1)R}{4}, \quad q = \sqrt{\ell}. \quad (11)$$

Result 3 (steady state, Theorem 1 of the paper). In steady state:

- ad revenue is $R = \frac{1}{4} \cdot \frac{3+2s}{3+4s}$, which decreases in data informativeness and does not depend on the elasticity of attention;
- platform investment is $\ell = \frac{(\epsilon-1)R}{4}$, and quality is $q = \sqrt{\ell}$;
- product consumption is $C = \frac{1}{2\epsilon} \left(\frac{1}{2} + \frac{2s}{3} \right)$.

The two markets are linked in both directions. Platform investment depends on ad revenue, which depends on how firms compete in the product market, and product consumption depends on the ad rate and the data that platforms supply. The paper requires $\epsilon - 1 < \frac{1}{\varphi} = 2$ for the platform's problem to be concave, which explains the restriction $\epsilon < 3$.

5. Step (d): Data Policy

Steady state. Set $\epsilon = 2$, so $A = M = \frac{1}{2}$, and compare three levels of data informativeness. From (8)–(11):

- uninformative data, $s = 0$: $\mu_H = E[\hat{v}_{(2)}] = \frac{1}{2}$, $R = \frac{1}{4}$, $C = \frac{1}{8}$, and $\ell = \frac{1}{16}$;
- intermediate data, $s = \frac{3}{10}$: $\mu_H = \frac{7}{10}$, $E[\hat{v}_{(2)}] = \frac{3}{5}$, $R = \frac{3}{14}$, $C = \frac{7}{40}$, and $\ell = \frac{3}{56}$;
- perfectly informative data, $s = \frac{1}{2}$: $\mu_H = \frac{5}{6}$, $E[\hat{v}_{(2)}] = \frac{2}{3}$, $R = \frac{1}{5}$, $C = \frac{5}{24}$, and $\ell = \frac{1}{20}$.

Result 4 (data, Proposition 2 of the paper). In steady state, more informative data:

- raises product consumption, $C = \frac{1}{8} + \frac{s}{6}$;
- lowers ad revenue, $R = \frac{1}{4} \cdot \frac{3+2s}{3+4s}$, which reaches its maximum, $\frac{I}{\sigma} \cdot \frac{\lambda_f}{\rho+\lambda_f} = \frac{1}{4}$, with uninformative data;
- lowers platform investment and quality.

Figure 1 depicts steady-state ad revenue and product consumption as functions of data informativeness s . Consumption increases from $\frac{1}{8}$ to $\frac{5}{24}$, while ad revenue decreases from $\frac{1}{4}$ to $\frac{1}{5}$. The dots mark the policy experiment below, from $s = \frac{3}{10}$ to $s = \frac{1}{2}$.

Intuitively, three forces drive the decline in ad revenue, as in the paper. First, by Result 1, firms' total profits are fixed at $\frac{1}{2}$ whatever the data. Second, with uninformative data all five bidders expect the same value and bid the same, so competition in the auction extracts the full value of joining a consideration set. Third, informative data creates information rents: the winner pays the second-highest bid, and the gap between the highest and second-highest expected values, $\frac{s}{3}$, increases in s . Better targeting does not raise the winner's profit either: the winner's expected value rises, but so does the average value μ_H of its rivals in the consideration set, and in steady state the two coincide.

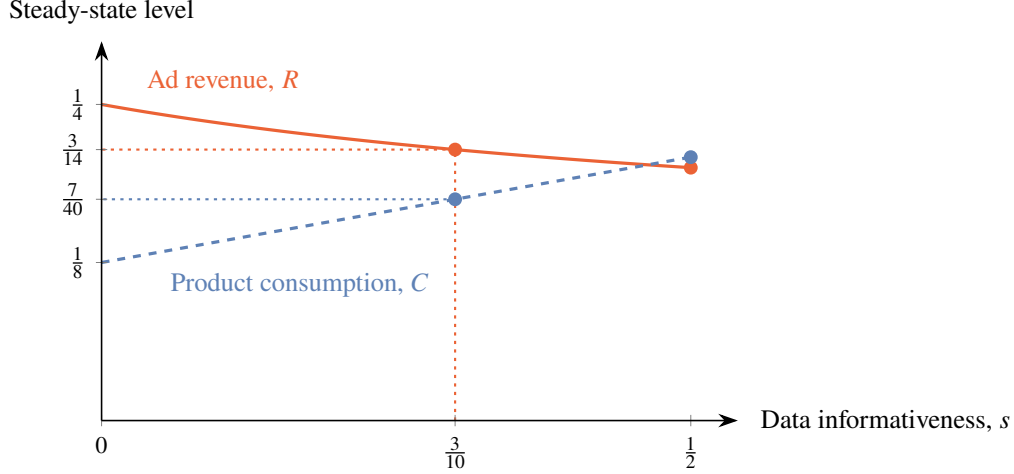


Figure 1: Steady-state ad revenue and product consumption as functions of data informativeness ($\epsilon = 2$).

Transition. Consider an unanticipated improvement in data at $t = 0$, from $s = \frac{3}{10}$ to $s = \frac{1}{2}$, as in Figure 2 of the paper. The auctions use the new data at once, so the second-highest expected value of the bidders jumps from $\frac{3}{5}$ to $\frac{2}{3}$. The composition of consideration sets, in contrast, changes only as products are forgotten and replaced. The products already in a set receive new expected values, but by the law of iterated expectations their average stays at $\frac{7}{10}$, and from then on $\mu_t = \frac{5}{6} - \frac{2}{15}e^{-t}$. A firm that joins at time t earns $\frac{\hat{v}}{2M\mu_\tau}$ at every later date τ , so its bid is $\hat{v} \int_0^\infty \frac{e^{-2u}}{\mu_{t+u}} du$, and ad revenue on impact is

$$R_0 = \frac{1}{2} \cdot \frac{2}{3} \int_0^\infty \frac{e^{-2u}}{\frac{5}{6} - \frac{2}{15}e^{-u}} du = \frac{75}{4} \left[\frac{5}{6} \ln \frac{25}{21} - \frac{2}{15} \right] \approx 0.224, \quad (12)$$

which exceeds the initial level $\frac{3}{14} \approx 0.214$ by about 5%. Ad revenue then declines toward $\frac{1}{5}$, about 7% below its initial level, and product consumption rises gradually from $\frac{7}{40}$ to $\frac{5}{24}$.

Figure 3 depicts this path (red curve), relative to the initial level. Ad revenue rises on impact, crosses its initial level at about $t = 0.48$, and settles below it. Firms bid more at first because they value the better targeting before their rivals in consideration sets become better matched. As in the paper, the rise on impact is larger when firms are less patient: it is about 2% with $\rho = 0.1$ and about 9% with $\rho = 10$. Steady-state investment falls from $\frac{3}{56}$ to $\frac{1}{20}$. With $\rho = 1$, platforms are impatient enough that investment first rises slightly, to about 0.055, before falling; in the paper, with more patient platforms, investment falls from the start.

6. Step (e): Interoperability

Steady state. Set $s = \frac{1}{2}$ and raise the elasticity of attention from $\epsilon = \frac{3}{2}$ to $\epsilon = 2$. By Results 2 and 3:

- the ad rate and the mass of consideration sets fall from $\frac{2}{3}$ to $\frac{1}{2}$;
- product consumption, $C = \frac{5}{12\epsilon}$, falls from $\frac{5}{18}$ to $\frac{5}{24}$;
- ad revenue stays at $R = \frac{1}{5}$;
- investment, $\ell = \frac{\epsilon-1}{20}$, doubles from $\frac{1}{40}$ to $\frac{1}{20}$.

Result 5 (interoperability, Proposition 3 of the paper). In steady state, a higher elasticity of attention:

- lowers the ad rate and the mass of consideration sets, and hence product consumption;
- leaves ad revenue unchanged in the toy model (in the paper it raises ad revenue, under a sufficient condition on G that the uniform distribution satisfies);
- raises platform investment and quality.

Figure 2 depicts steady-state product consumption, ad revenue, and investment as functions of ϵ . Consumption decreases and investment increases with ϵ , while ad revenue is flat. Intuitively, when attention is more elastic, ads drive users away faster, so platforms show fewer ads and compete for attention through quality instead. Fewer ads mean that consumers discover fewer products. Each ad, however, becomes more valuable: by (9), a smaller consideration set means fewer rivals for each consumer, which raises bids. With $M = A$, the two effects offset exactly. In the paper, fewer ads also reduce the rate at which forgotten firms are rediscovered, which raises bids further, so ad revenue increases; the large pool of products in the toy model shuts down this channel.

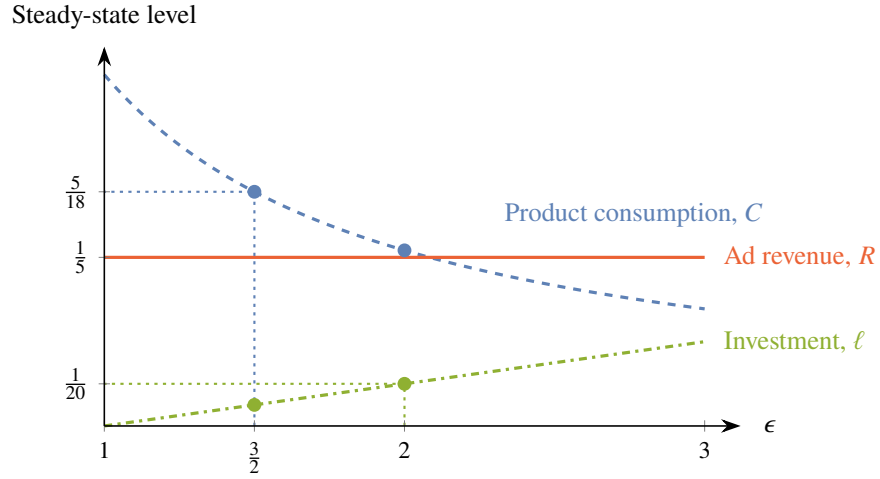


Figure 2: Steady-state product consumption, ad revenue, and platform investment as functions of the elasticity of attention ($s = \frac{1}{2}$).

Transition. Consider an unanticipated increase in ϵ from $\frac{3}{2}$ to 2 at $t = 0$. Platforms cut the ad rate to $\frac{1}{2}$ at once, but consideration sets shrink only gradually, $M_t = \frac{1}{2} + \frac{1}{6}e^{-t}$, so bids rise only gradually. A firm that joins at time t bids $\hat{v} \int_0^\infty \frac{e^{-2u}}{2M_{t+u}\mu_H} du$, and since $\frac{AE[\hat{v}(2)]}{2\mu_H} = \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{5}{3}} = \frac{1}{5}$, ad revenue at time t is $R_t = \frac{1}{5} \int_0^\infty \frac{e^{-2u}}{M_{t+u}} du$. On impact,

$$R_0 = \frac{1}{5} \int_0^\infty \frac{e^{-2u} du}{\frac{1}{2} + \frac{1}{6}e^{-u}} = \frac{36}{5} \left[\frac{1}{6} - \frac{1}{2} \ln \frac{4}{3} \right] \approx 0.164, \quad (13)$$

about 18% below the steady-state level $\frac{1}{5}$. Figure 3 depicts this path (blue curve): ad revenue falls on impact and recovers as consideration sets shrink. Product consumption falls gradually from $\frac{5}{18}$ to $\frac{5}{24}$. In the toy model, investment rises on impact, to about 0.058, and then declines toward $\frac{1}{20}$: the elasticity $\epsilon - 1$ doubles, which outweighs the dip in ad revenue. In the paper, with steeper nuisance costs, the ad rate falls by more on impact, and investment first falls before rising in the long run.

Ad revenue relative to its initial level

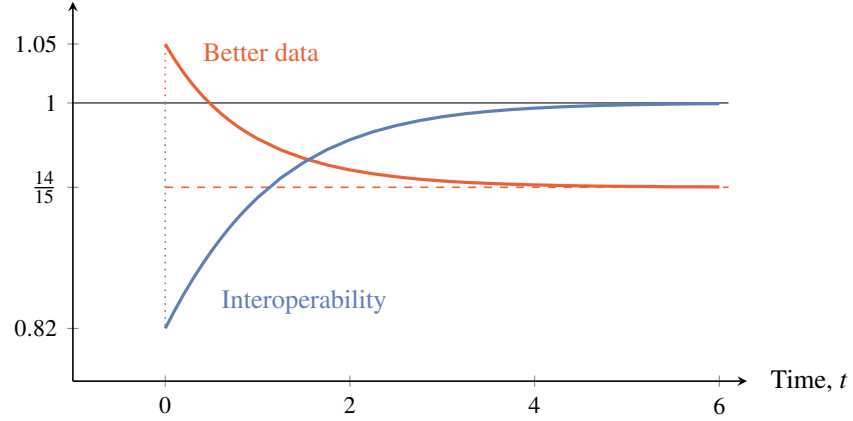


Figure 3: Ad revenue after an unanticipated policy change at $t = 0$, relative to its initial level: better data, from $s = \frac{3}{10}$ to $s = \frac{1}{2}$ with $\epsilon = 2$ (red), and interoperability, from $\epsilon = \frac{3}{2}$ to $\epsilon = 2$ with $s = \frac{1}{2}$ (blue). The dashed line is the long-run level after better data.

7. Step (f): Welfare

General equilibrium. To evaluate welfare, the paper closes the model. Consumers supply $L = 1$ unit of labor, the only input: platforms hire ℓ units for investment, and firms use the rest for production, one unit of labor per unit of output. Consumers own all firms and platforms, and their income equals firms' revenue, $I = 2(1 - \ell)$, the price times production. Ad revenue per unit of income, $\hat{R} = \frac{R}{I} = \frac{1}{4} \cdot \frac{3+2s}{3+4s}$, does not depend on I , so (11) becomes $\ell = \frac{(\epsilon-1)\hat{R}}{4} 2(1 - \ell)$, and

$$\ell = \frac{x}{4+x}, \quad \text{where } x \equiv \frac{\sigma}{\sigma-1} \hat{R} (\epsilon-1) = \frac{(3+2s)(\epsilon-1)}{2(3+4s)}. \quad (14)$$

Planner. Consumers' flow utility is $C^{1-\tau} X^\tau$, with $\tau \in (0, 1)$, where $C = (1 - \ell)M\mu_H$ and $X = v(a)q$. The planner chooses investment and the ad rate, while consumers and firms behave as in the equilibrium. In steady state, the shadow value of quality is $\frac{\tau u}{2q}$, and investing one more unit of labor costs $\frac{(1-\tau)u}{1-\ell}$ in utility and raises quality at rate $\frac{1}{2\sqrt{\ell}}$, so the planner's investment satisfies $\frac{(1-\tau)u}{1-\ell} = \frac{\tau u}{4\ell}$, that is,

$$\ell^* = \frac{\frac{\tau}{1-\tau}}{4 + \frac{\tau}{1-\tau}} = \frac{\tau}{4 - 3\tau}. \quad (15)$$

For the ad rate, the paper compares steady states (its limit $\rho \rightarrow 0$). Since $M = a$ and μ_H does not depend on a , the planner maximizes $a^{1-\tau}(1-a)^\tau$, which gives $a^* = 1 - \tau$.

Result 6 (welfare, Proposition 5 and Theorem 2 of the paper). Comparing the equilibrium with the planner's allocation:

- platforms invest too much if $x > \frac{\tau}{1-\tau}$, and too little if $x < \frac{\tau}{1-\tau}$, where x is the sufficient statistic in (14);
- platforms show too many ads if $\frac{1}{\epsilon} > 1 - \tau$, that is, if $\tau > 1 - \frac{1}{\epsilon}$, and too few ads otherwise;
- a tax on ad revenue at rate $1 - \frac{\tau}{(1-\tau)x}$, rebated to consumers, restores efficient investment (a negative rate is a subsidy).

The two sides of the investment condition reflect different distortions. The statistic x is the product of the markup $\frac{\sigma}{\sigma-1} = 2$, ad revenue per unit of income, and the elasticity $\epsilon - 1$. Ad revenue and $\epsilon - 1$ measure business stealing: platforms invest to take attention, and ad revenue, from their rivals, as in Anderson and Coate (2005). The markup reflects product-market power, which shifts labor toward platform investment. The term $\frac{\tau}{1-\tau}$ measures the value of quality to consumers, which platforms do not collect because their services are free. The ad rate is too high when attention is inelastic, since platforms then care little about the nuisance cost of ads, and too low when consumers value products far more than platform services, since platforms do not collect the value of the products that ads help consumers discover.

Figure 4 depicts these regions in the (ϵ, τ) plane with perfectly informative data, $s = \frac{1}{2}$, where $x = \frac{2(\epsilon-1)}{5}$. Investment is too high below the blue curve, $\tau = \frac{2(\epsilon-1)}{2\epsilon+3}$, and the ad rate is too high above the red curve, $\tau = 1 - \frac{1}{\epsilon}$. The dot marks $\epsilon = 2$ and $\tau = \frac{1}{4}$. There, $x = \frac{2}{5}$, so equilibrium investment is $\frac{1}{11}$, above the planner's $\frac{1}{13}$, and a tax of $\frac{1}{6}$ on ad revenue restores efficiency. The ad rate, $\frac{1}{2}$, is below the planner's $\frac{3}{4}$.

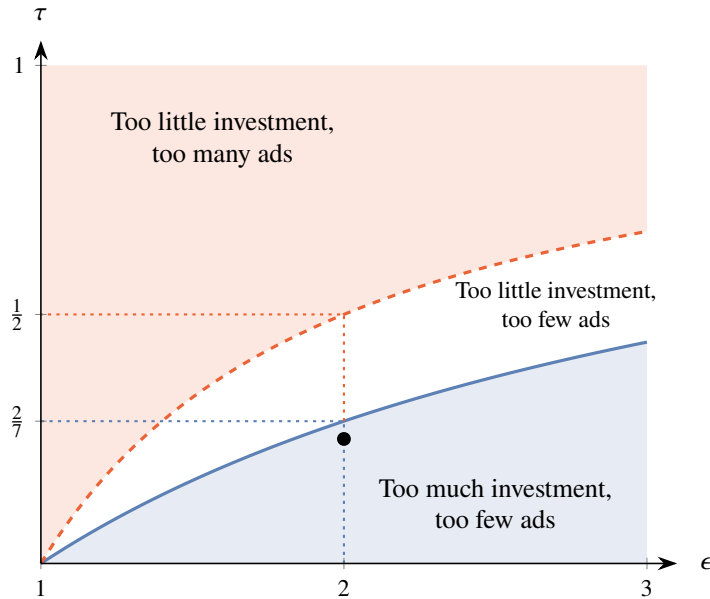


Figure 4: Investment and ad distortions in the (ϵ, τ) plane with $s = \frac{1}{2}$. Investment is too high in the blue region and the ad rate is too high in the red region.

The investment condition also tells us how to evaluate the policies of Steps (d) and (e). With uninformative data, $x = \frac{1}{2}$ at $\epsilon = 2$, so equilibrium investment rises to $\frac{1}{9}$, further above the planner's $\frac{1}{13}$, and the corrective tax rises to $\frac{1}{3}$. Hence, when investment is already too high, a ban on data use moves investment further from the efficient level, as the paper notes. Similarly, interoperability raises x but not the planner's investment ℓ^* , so it moves investment toward the efficient level only when platforms initially invest too little.

8. Step (g): What the Toy Model Keeps and Drops

Table 1 reports the effects of the two policy experiments of Steps (d) and (e).

What the toy model keeps.

- The link between the two markets: ads determine which products consumers discover, and product-market competition determines how much firms pay for ads.

Table 1: Effects of better data (s from $\frac{3}{10}$ to $\frac{1}{2}$, $\epsilon = 2$) and interoperability (ϵ from $\frac{3}{2}$ to 2, $s = \frac{1}{2}$).

Policy	Ad revenue, impact	Ad revenue, long run	Consumption	Investment, long run
Better data	+5%	−7%	$\frac{7}{40}$ to $\frac{5}{24}$ (+19%)	$\frac{3}{56}$ to $\frac{1}{20}$ (−7%)
Interoperability	−18%	0%	$\frac{5}{18}$ to $\frac{5}{24}$ (−25%)	$\frac{1}{40}$ to $\frac{1}{20}$ (+100%)

- The fixed profit pie under CES demand, which explains why better targeting does not raise firms’ total willingness to pay for ads.
- Information rents in second-price auctions, which make ad revenue highest with uninformative data.
- The trade-off between ad load and quality when attention becomes more elastic.
- Short-run effects that differ from long-run effects, because consideration sets adjust gradually while auctions and ad rates respond at once.
- The two welfare distortions, business stealing and unpriced consumer surplus, and the sufficient statistic that signs the investment distortion.

What it drops.

- A finite pool of products. In the paper, forgotten firms can be rediscovered, and fewer ads slow down rediscovery, which raises bids; this channel makes interoperability raise long-run ad revenue.
- Strategic investment by a duopoly of large platforms, which the paper solves numerically and finds to preserve the effects of both policies.
- Data asymmetry between data-rich digital platforms and traditional media. The paper shows that data-rich platforms can gain a larger share of ad revenue while total ad revenue falls, consistent with the decline in advertising’s share of GDP during 2000–2018 documented by Silk and Berndt (2021).
- Network effects, platforms with different productivity, prices for platform services, and entry of firms and platforms, which the paper studies in its online appendix.

References

- Anderson, S. P., and Coate, S. (2005). Market provision of broadcasting: A welfare analysis. *Review of Economic Studies*, 72(4), 947–972.
- Chen, D. (2026). A model of the attention economy: Data, platforms, and discovery. *American Economic Review*, 116(10), 3802–3836.
- Silk, A. J., and Berndt, E. R. (2021). Aggregate advertising expenditure in the US economy: Measurement and growth issues in the digital era. *Foundations and Trends in Marketing*, 15(1), 1–85.

Greener or Cheaper: Economies of Scope in R&D

A step-by-step exercise in the spirit of Strandholm, Espinola-Arredondo, and Munoz-Garcia (2025)

Companion note to Strandholm, J. C., A. Espinola-Arredondo, and F. Munoz-Garcia (2025), “Greener or Cheaper Goods: Economies of Scope in R&D Investments,” *Journal of Environmental Economics and Management* 130: 103107. Full references at the end.

Overview. Firms invest both in cost-reducing R&D, which makes their goods cheaper, and in green R&D (abatement), which makes them cleaner. The literature on environmental policy usually studies one investment at a time: models of abatement with an emission fee, or models of cost-reducing R&D without environmental policy. This note shows that the two decisions interact even when their costs are independent. A firm that invests in R&D produces more, which leads the regulator to raise the emission fee; a firm that invests in abatement pollutes less, which leads the regulator to lower it. As a consequence, the two investments become strategic complements under regulation, and, when economies of scope are weak, a regulator who sets the fee from a model with abatement only underestimates the pollution of the industry. In addition, innovations in one investment can lower the cost of the other (economies of scope), which gives firms a reason to abate even without regulation. Step (a) solves the last two stages, output and the emission fee; Step (b) solves investments without regulation; Step (c) solves investments under regulation; Step (d) compares the fee against that in a model with abatement only; Step (e) considers firms that choose abatement before R&D; and Step (f) lists what the toy model keeps and drops.

1. The Toy Model

Market. Two firms, i and j , sell a homogeneous good with inverse demand $p(Q) = 1 - Q$, where $Q = q_i + q_j$. Firm i 's marginal cost is $\frac{1}{2} - k_i$, where k_i denotes its investment in cost-reducing R&D. Production pollutes one unit of emissions per unit of output, and firm i 's investment in abatement, z_i , lowers its emissions to $e_i = q_i - z_i$. Environmental damage is

$$ED = \frac{3}{2}(Q - Z)^2, \quad (1)$$

where $K = k_i + k_j$ and $Z = z_i + z_j$ denote aggregate R&D and aggregate abatement.

Investment costs. Firm i 's cost of investing (k_i, z_i) is

$$C(k_i, z_i) = k_i^2 + \frac{3}{4}z_i^2 - \lambda k_i z_i, \quad \lambda \in [-1, 1]. \quad (2)$$

When $\lambda = 0$, the two investments have independent costs. When $\lambda > 0$, they exhibit economies of scope: the marginal cost of R&D, $2k_i - \lambda z_i$, decreases in abatement, and the marginal cost of abatement, $\frac{3}{2}z_i - \lambda k_i$, decreases in R&D. When $\lambda < 0$, they exhibit diseconomies of scope. In the paper's notation, the toy model sets $n = 2$ firms, a marginal cost $c = \frac{1}{2}$, an R&D cost parameter $\gamma = 2$, an abatement cost parameter $\alpha = \frac{3}{2}$, and pollution severity $d = \frac{3}{2}$. The interval $\lambda \in [-1, 1]$ guarantees that every firm's investment problem is strictly concave and that all equilibrium investments and fees are positive.

Timing. In the first stage, both firms simultaneously choose (k_i, z_i) . In the second stage, the regulator observes all investments and sets an emission fee t per unit of emissions. In the third stage, firms compete à la Cournot. The regulator maximizes social welfare,

$$SW = CS + PS + T - ED, \quad (3)$$

where $CS = \frac{1}{2}Q^2$ is consumer surplus, PS denotes profits net of fees and investment costs, and $T = t(Q - Z)$ is fee revenue. As a benchmark, Step (b) considers an unregulated industry, $t = 0$.

2. Step (a): Output and the Emission Fee

Third stage. Given investments and fee t , firm i chooses q_i to maximize $(1 - Q)q_i - \left(\frac{1}{2} - k_i\right)q_i - t(q_i - z_i)$. Solving the two first-order conditions yields

$$q_i = \frac{\frac{1}{2} - t + 2k_i - k_j}{3}, \quad \text{and} \quad Q = \frac{1 - 2t + K}{3}. \quad (4)$$

Firm i 's output increases in its own R&D, decreases in its rival's R&D, and decreases in the fee. Its profit, including investment costs, is

$$\pi_i = q_i^2 + tz_i - C(k_i, z_i), \quad (5)$$

where tz_i captures the fee that firm i avoids by abating.

Second stage. An increase in t lowers every firm's output by $\frac{1}{3}$, so the regulator's first-order condition equates price to the average marginal cost plus the marginal damage of emissions,

$$1 - Q = \left(\frac{1}{2} - \frac{K}{2}\right) + 3(Q - Z), \quad \text{which yields} \quad Q = \frac{1 + K + 6Z}{8}. \quad (6)$$

Inserting Q from (4) and solving for t gives the emission fee.

Result 1 (emission fee). The welfare-maximizing fee is

$$t(K, Z) = \frac{5}{16} + \frac{5}{16}K - \frac{9}{8}Z, \quad (7)$$

which increases in aggregate R&D and decreases in aggregate abatement. The fee is separable in K and Z : the effect of one more unit of abatement on the fee does not depend on R&D.

Intuitively, R&D lowers marginal costs and expands output, so the industry pollutes more and the regulator responds with a more stringent fee. Abatement, instead, lowers emissions at every output level, so the regulator can afford a less stringent fee. This gives rise to two externalities across firms that a model with a single investment cannot capture: firm i 's R&D raises the fee that its rival pays, and firm i 's abatement lowers it. For the general model, the fee is $t(K, Z) = \frac{(2dn-1)[n(1-c)+K]-2dn(n+1)Z}{(2d+1)n^2}$, which reduces to (7) at the toy-model parameters.

3. Step (b): Investments without Regulation

Without regulation, $t = 0$, so firm i chooses (k_i, z_i) to maximize $\left(\frac{\frac{1}{2}+2k_i-k_j}{3}\right)^2 - \left(k_i^2 + \frac{3}{4}z_i^2 - \lambda k_i z_i\right)$. The first-order conditions are

$$\frac{4}{9} \left(\frac{1}{2} + 2k_i - k_j\right) - 2k_i + \lambda z_i = 0 \quad \text{and} \quad \lambda k_i - \frac{3}{2}z_i \leq 0. \quad (8)$$

Abatement only lowers the cost of R&D, so $z_i = \frac{2\lambda}{3}k_i$ when $\lambda > 0$, and $z_i = 0$ otherwise. Imposing symmetry yields the next result.

Result 2 (no regulation). Equilibrium investments without regulation are:

- If $\lambda \leq 0$, then $k^{NR} = \frac{1}{7}$ and $z^{NR} = 0$.
- If $\lambda > 0$, then $k^{NR} = \frac{1}{7-3\lambda^2}$ and $z^{NR} = \frac{2\lambda}{3(7-3\lambda^2)}$, both increasing in λ .

Intuitively, an unregulated firm does not pay for its emissions, so abatement has no value of its own. With economies of scope, however, abatement lowers the marginal cost of R&D, and the firm abates to make its R&D cheaper. This motive for green investment does not rely on emission fees, green consumers, or corporate social responsibility.

4. Step (c): Investments under Regulation

In the first stage, firm i anticipates output (4) and fee (7), and chooses (k_i, z_i) to maximize profit (5). Solving the first-order conditions for k_i and z_i , respectively, yields the best responses

$$k_i = \frac{9}{175} + \frac{94 + 128\lambda}{175} z_i - \frac{9}{25} k_j + \frac{54}{175} z_j, \quad (9)$$

$$z_i = \frac{23}{222} + \frac{47 + 64\lambda}{222} k_i - \frac{1}{222} k_j - \frac{9}{37} z_j. \quad (10)$$

Firm i 's two investments are strategic complements: R&D increases in own abatement, and abatement increases in own R&D, whenever $\lambda > -\frac{47}{64}$, which includes the case without economies of scope, $\lambda = 0$. Intuitively, abatement lowers the fee, which raises output and, thus, the return of every unit of cost reduction; and R&D raises the fee, which raises the fee that every unit of abatement avoids. Economies of scope strengthen both effects. In contrast, each investment is a strategic substitute of the rival's investment of the same type.

Imposing symmetry, $k_i = k_j = k$ and $z_i = z_j = z$, in (9) and (10) gives the two loci

$$k = \frac{9 + (148 + 128\lambda)z}{238} \quad \text{and} \quad z = \frac{23 + (46 + 64\lambda)k}{276}, \quad (11)$$

whose intersection is the equilibrium. Figure 1 depicts both loci at $\lambda = 0$ and $\lambda = \frac{1}{2}$. Their intercepts, $k = \frac{9}{238}$ when $z = 0$ and $z = \frac{1}{12}$ when $k = 0$, are the investments of models with a single type of R&D, which Step (d) uses as benchmarks. Both loci are positively sloped, so the equilibrium lies above and to the right of both intercepts. An increase in λ rotates both loci, moving the equilibrium to the northeast.

Result 3 (investments under regulation). Each firm's equilibrium investments and the resulting fee are

$$k^R = \frac{23(2 + \lambda)}{4(115 - 30\lambda - 16\lambda^2)}, \quad z^R = \frac{92 + 9\lambda}{8(115 - 30\lambda - 16\lambda^2)}, \quad t^R = \frac{276 - 133\lambda - 80\lambda^2}{16(115 - 30\lambda - 16\lambda^2)}. \quad (12)$$

Without economies of scope, $\lambda = 0$, they become $k^R = z^R = \frac{1}{10}$ and $t^R = \frac{3}{20}$. Both investments increase in λ . Relative to no regulation:

- Abatement is higher under regulation, $z^R > z^{NR}$, for all $\lambda \in [-1, 1]$.
- R&D is lower under regulation, $k^R < k^{NR}$, if and only if $\lambda < \lambda_k \equiv \frac{\sqrt{58537} - 143}{138} \approx 0.717$.

Figure 2 depicts equilibrium investments with and without regulation as a function of λ . Without regulation, abatement is zero until economies of scope appear, and R&D is flat at $\frac{1}{7}$ for $\lambda \leq 0$. Under regulation, both investments are positive at every λ , and R&D is lower than without regulation unless economies of scope are strong. Intuitively, the fee makes output more expensive, which lowers the return of cost reduction; but it also makes abatement more valuable, and, when λ is high, more abatement makes R&D cheap enough to offset the first effect.

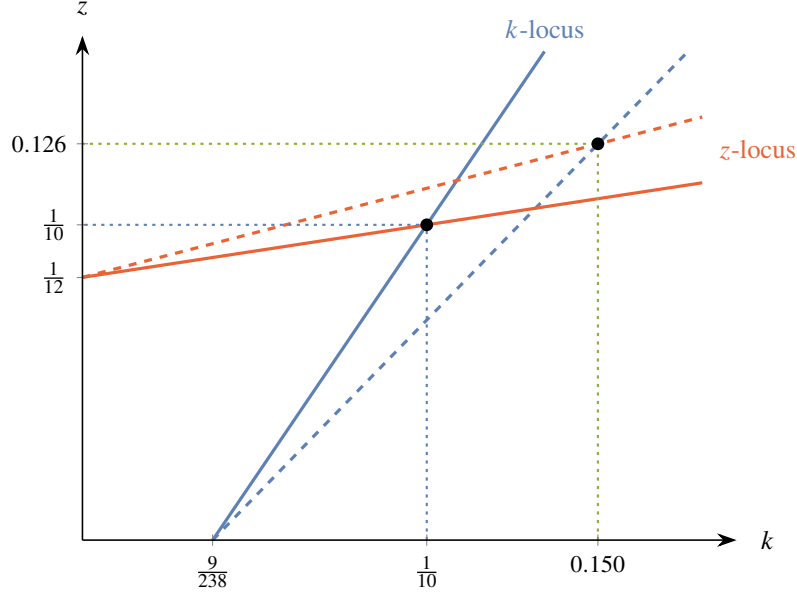


Figure 1: Symmetric best-response loci under regulation, at $\lambda = 0$ (solid) and $\lambda = \frac{1}{2}$ (dashed). Black dots mark the equilibria.

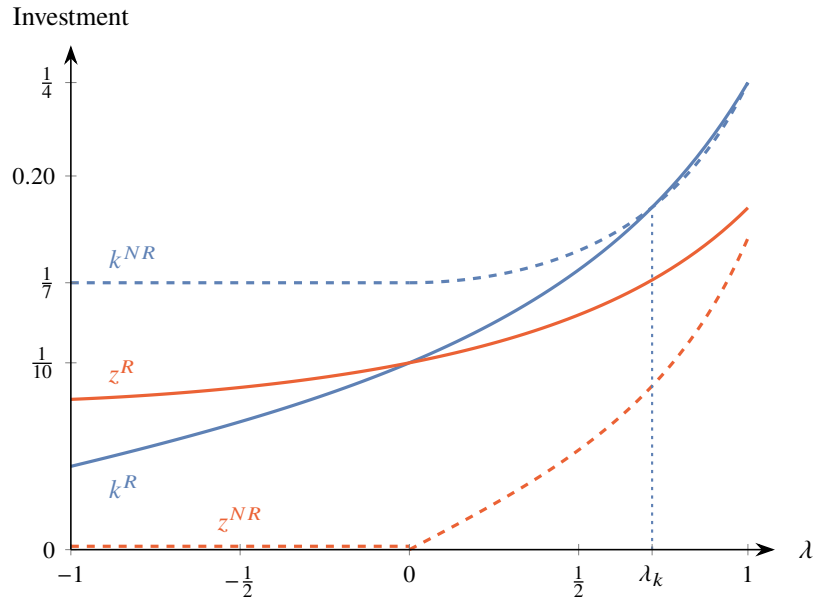


Figure 2: Equilibrium R&D and abatement under regulation (solid) and without regulation (dashed).

5. Step (d): Undertaxation

Consider now a regulator who designs the fee using a standard model where firms only invest in abatement, as in Poyago-Theotoky (2007). Setting $k = 0$ in (11) gives abatement $\bar{z} = \frac{1}{12}$ per firm and, from (7), the fee

$$\bar{t} = \frac{5}{16} - \frac{9}{8} \cdot \frac{1}{6} = \frac{1}{8}, \quad (13)$$

which does not depend on λ , since the cross term $\lambda k_i z_i$ vanishes when $k_i = 0$. Similarly, a model where firms only invest in R&D predicts $\bar{k} = \frac{9}{238} \approx 0.038$. At $\lambda = 0$, both one-investment models underestimate investments, $\bar{z} < z^R$ and $\bar{k} < k^R = \frac{1}{10}$, because they omit the complementarity between them.

Result 4 (undertaxation). The fee gap is

$$t^R - \bar{t} = \frac{(2 + \lambda)(23 - 48\lambda)}{16(115 - 30\lambda - 16\lambda^2)}, \quad (14)$$

so that:

- If $\lambda < \frac{23}{48}$, then $t^R > \bar{t}$: the abatement-only model prescribes a fee that is too lax (undertaxation).
- If $\lambda > \frac{23}{48}$, then $t^R < \bar{t}$: the abatement-only model prescribes a fee that is too stringent (overtaxation).

In addition, t^R is decreasing in λ for all $\lambda > -0.718$.

Figure 3 depicts both fees. At $\lambda = 0$, the gap is $\frac{1}{40}$, and it can be decomposed using (7), which, per firm, reads $t = \frac{5}{16} + \frac{5}{8}k - \frac{9}{4}z$. R&D raises the fee by $\frac{5}{8} \cdot \frac{1}{10} = \frac{1}{16}$, while the extra abatement, $\frac{1}{10} - \frac{1}{12} = \frac{1}{60}$ per firm, lowers it by $\frac{9}{4} \cdot \frac{1}{60} = \frac{3}{80}$, so that $\frac{1}{16} - \frac{3}{80} = \frac{1}{40}$. Intuitively, the abatement-only model ignores the R&D that makes production cheaper and dirtier, so it expects less pollution than the industry generates, and sets a lax fee. As λ increases, both investments increase, but abatement lowers the fee more than R&D raises it: from $\lambda = 0$ to $\lambda = \frac{1}{2}$, k^R grows by 0.050 and z^R by 0.026, which changes the fee by $\frac{5}{8} \cdot 0.050 - \frac{9}{4} \cdot 0.026 \approx -0.027$. Economies of scope, therefore, ameliorate undertaxation, and reverse it when they are strong enough.

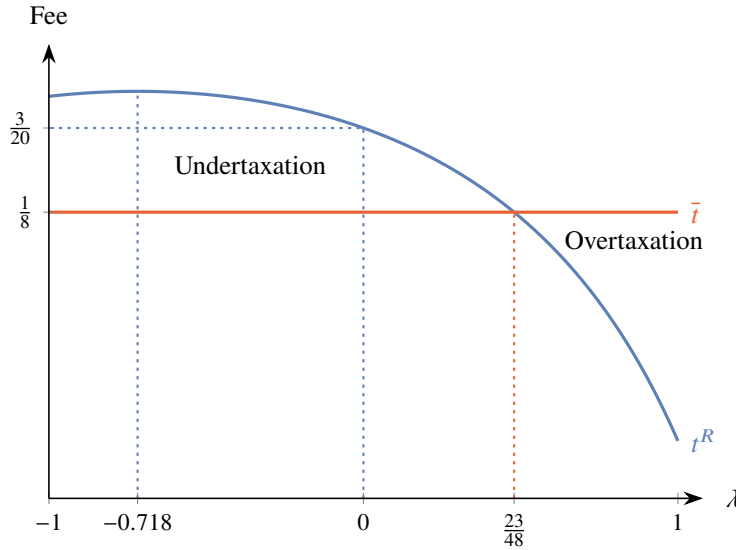


Figure 3: Equilibrium fee t^R and the fee from an abatement-only model, $\bar{t}$.

6. Step (e): Sequential Investments

Consider now that firms choose abatement first, and R&D second, before the regulator sets the fee and firms compete. Abatement then works as a commitment: when firm i chooses z_i , it anticipates how both firms' R&D responds.

Second stage: R&D given abatement. Solving both firms' first-order conditions for R&D given (z_i, z_j) yields

$$k_i^{NR}(z_i, z_j) = \frac{2 + 15\lambda z_i - 6\lambda z_j}{14} \quad \text{and} \quad k_i^R(z_i, z_j) = \frac{18 + (233 + 400\lambda)z_i + (63 - 144\lambda)z_j}{476}, \quad (15)$$

without and with regulation, respectively. Without regulation, firm i 's abatement lowers its rival's R&D, $\frac{\partial k_j^{NR}}{\partial z_i} = -\frac{3\lambda}{7} < 0$ when $\lambda > 0$: it makes firm i 's R&D cheaper, firm i invests more in R&D, and its rival responds with less. With regulation,

$$\frac{\partial k_j^R}{\partial z_i} = \frac{9(7 - 16\lambda)}{476}, \quad (16)$$

which is positive if and only if $\lambda < \frac{7}{16}$. Firm i 's abatement now has a second effect: it lowers the fee that both firms pay, which raises the rival's output and its return to R&D. This effect dominates when economies of scope are weak.

First stage: abatement. By the envelope theorem, firm i 's first-order condition for abatement is

$$\underbrace{\frac{\partial \pi_i}{\partial z_i}}_{\text{Direct effect}} + \underbrace{\frac{\partial \pi_i}{\partial k_j} \frac{\partial k_j}{\partial z_i}}_{\text{Strategic effect}} = 0, \quad (17)$$

where the direct effect alone gives the simultaneous-investment condition. Since the rival's R&D hurts firm i , $\frac{\partial \pi_i}{\partial k_j} < 0$, the strategic effect has the opposite sign of $\frac{\partial k_j}{\partial z_i}$, as in Brander and Spencer (1983).

Result 5 (sequential investments). When firms choose abatement before R&D:

- Without regulation, abatement is $z^{NR,Seq} = \frac{20\lambda}{3(49-23\lambda^2)}$, which exceeds simultaneous abatement z^{NR} for all $\lambda > 0$, and both are zero when $\lambda = 0$.
- With regulation, abatement is $z^{R,Seq} = \frac{5(613+90\lambda)}{31721-9105\lambda-4496\lambda^2}$, which is below simultaneous abatement z^R , and yields a more stringent fee, $t^{R,Seq} > t^R$, if and only if $\lambda < \frac{7}{16}$.

Figure 4 depicts the abatement differential between simultaneous and sequential investments without regulation (panel a) and with regulation (panel b). Without regulation, the differential is negative: firms over-invest in abatement to commit to more R&D, which discourages their rival's R&D. With regulation and weak economies of scope, the differential is positive: an extra unit of abatement lowers the fee and encourages the rival's R&D, so firms abate less when they anticipate this response. At $\lambda = 0$, sequential abatement is $z^{R,Seq} = \frac{3065}{31721} \approx 0.097$, against $z^R = \frac{1}{10}$, and the fee rises from 0.150 to 0.156. When $\lambda > \frac{7}{16}$, the commitment effect of the unregulated industry prevails again.

7. Step (f): What the Toy Model Keeps and Drops

Table 1 collects equilibrium outcomes at $\lambda = 0$ and $\lambda = \frac{1}{2}$.

What the toy model keeps.

- The emission fee increases in R&D and decreases in abatement, creating two externalities across firms.
- Without regulation, firms abate only when the investments exhibit economies of scope.

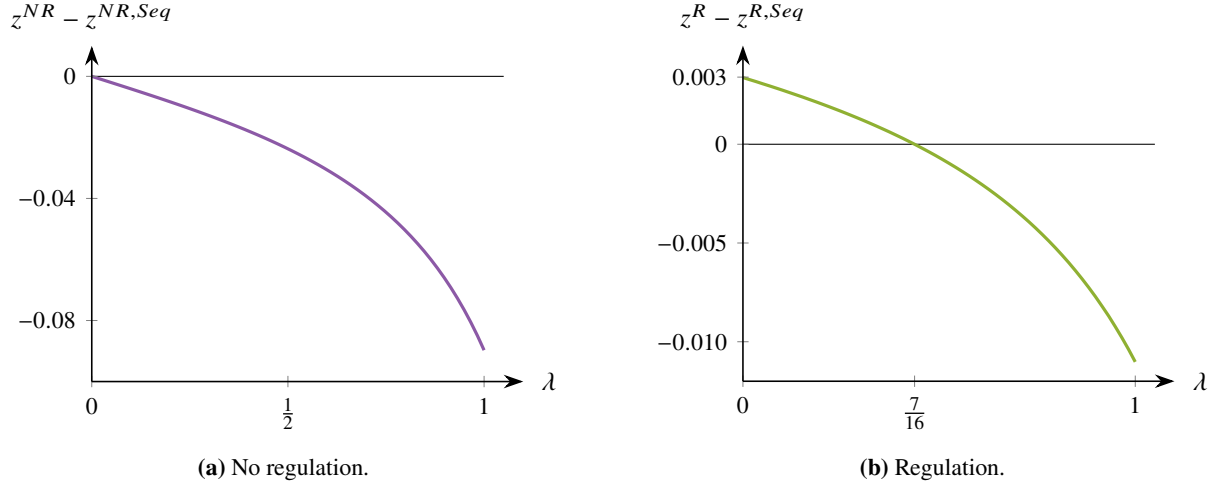


Figure 4: Abatement under simultaneous minus sequential investments.

Table 1: Equilibrium outcomes ($n = 2$, $c = \frac{1}{2}$, $\gamma = 2$, $\alpha = d = \frac{3}{2}$); k and z are per firm.

Setting	λ	k	z	t	$Q - Z$	SW
No regulation	0	0.1429	0	0	0.4286	-0.1327
Regulation, simultaneous	0	0.1000	0.1000	0.1500	0.1000	0.0850
Regulation, sequential	0	0.0979	0.0966	0.1563	0.1012	0.0842
No regulation	$\frac{1}{2}$	0.1600	0.0533	0	0.3333	-0.0200
Regulation, simultaneous	$\frac{1}{2}$	0.1497	0.1257	0.1234	0.0996	0.1018
Regulation, sequential	$\frac{1}{2}$	0.1503	0.1263	0.1222	0.0994	0.1020
Abatement only	any	0	0.0833	0.1250	0.0833	0.0729

- Under regulation, R&D and abatement are strategic complements, even without economies of scope.
- A regulator who uses a model with abatement only sets a lax fee (undertaxation), unless economies of scope are strong.
- Regulation raises abatement and, unless economies of scope are strong, lowers R&D.

What the toy model drops.

- The general number of firms n and the general parameters c , γ , α , and d , so it cannot show how undertaxation varies with market concentration or pollution severity.
- Asymmetric economies of scope, and spillovers or cooperation in R&D, which the paper lists as further research.

References

- Brander, J. A., and Spencer, B. J. (1983). Strategic commitment with R&D: The symmetric case. *Bell Journal of Economics*, 14(1), 225–235.
- Poyago-Theotoky, J. A. (2007). The organization of R&D and environmental policy. *Journal of Economic Behavior and Organization*, 62(1), 63–75.

Strandholm, J. C., Espinola-Arredondo, A., and Munoz-Garcia, F. (2025). Greener or cheaper goods: Economies of scope in R&D investments. *Journal of Environmental Economics and Management*, 130, 103107.