Where Manipulation Hides

Aryan Ayyar*

July 2026

Abstract

A canceled order records an event while leaving its economic source unresolved. Enforcement aimed at visible spoofing patterns therefore changes the tactics traders choose. Some activity exits; the rest migrates toward channels where an execution-based explanation is easier to sustain. A source-evidence platform can recover the missing link among display, market response, and beneficiary trade. Its fixed cost can delay adoption. The model predicts local suppression alongside opaque migration, discontinuous evidence investment, and legitimate-liquidity withdrawal. Apparent enforcement success can conceal greater market-wide harm.

JEL Classification: G12, G14, G18, G28

Keywords: spoofing, market manipulation, surveillance, evidence, displacement, market liquidity

*Draft. Comments are welcome.

1 Introduction

Modern markets record nearly every message and still leave open the central question in a spoofing case: why was the order valuable to the trader who submitted it? A displayed order may have offered a real opportunity to trade when it entered the book. The trader may cancel after prices move, inventory changes, the queue deteriorates, or execution becomes too costly. The same public sequence may arise when the order is valuable because other traders respond to its display and the submitting trader earns a related profit elsewhere. Both orders appear, affect the book, and disappear before execution. Their public histories can be identical.

This ambiguity makes spoofing an evidence problem. The legal question concerns the trader's intent to cancel when the order was submitted. The economic question concerns the source of the order's value. Their relation must be established rather than assumed. A strategy can depend on the market's response when the available evidence fails to establish a cancellation plan. A trader can plan to cancel an order that remains profitable under an ordinary execution rationale. Cancellation records, message counts, short lifetimes, and opposite-side trades can be informative about either question. None reveals the answer by itself.

The distinction becomes more important once traders respond to surveillance. Enforcement changes the expected return to each way of influencing the market. A visible tactic becomes less attractive when the regulator can connect the displayed order to the trade that benefits from the induced response. A similar payoff may remain available through another product, venue, account, or timing pattern where the beneficiary chain is harder to reconstruct. Traders then change where they display, where they trade, and how their activity appears in the data. Apparent enforcement success can conceal displacement.

This paper develops a theory of spoofing enforcement in which evidence quality, tactic choice, and legitimate liquidity are jointly determined. Strategic traders choose between a transparent tactic and an opaque tactic, or stay out of the market. The transparent tactic is easier to investigate. The opaque tactic uses a less visible connection between display and beneficiary execution. Its gross appeal can be lower, yet its evidentiary shelter can make it the preferred refuge after enforcement tightens. Entry into either tactic reduces the response rent available to other entrants, since displayed liquidity loses influence as response-dependent activity becomes more prevalent. A common participation margin links the two tactics and creates equilibrium migration.

The model uses response dependence as its economic classification. A displayed strategy is response dependent when it loses money under its own execution opportunity and becomes profitable through the reaction that its display induces. Execution-rationalizable strategies form the comparison group. This source classification disciplines the payoff analysis without deciding the legal case. Legal intent enters through a separate bridge that describes how often response-dependent and execution-rationalizable conduct is associated with an intent to cancel at submission. The regulator draws an intent inference from source evidence through that bridge. Source dependence carries no legal force when it contains no information about intent. Even perfect source evidence cannot meet a burden that exceeds the strongest intent inference permitted by the bridge.

The timing captures deterrence. The regulator first chooses an evidence architecture and the share of each tactic that will receive scrutiny. Legitimate liquidity providers then decide whether quoting remains worthwhile. Strategic traders observe the resulting expected penalties, decide whether to enter, and select a tactic. The market response and beneficiary trades occur, audits produce noisy source evidence, and the legal rule determines sanctions. Commitment to architecture and coverage before entry gives enforcement its effect on conduct. The regulator thereby chooses an information system, an audit policy, and an equilibrium allocation of activity across tactics.

Public surveillance has a limit. Within a displayed policy that admits ordinary execution and response-dependent strategies, the complete public order history can have the same distribution for both groups. This statement includes prices, sizes, messages, executions, cancellations, timestamps, and the visible path of the book. A classifier confined to that history cannot update the within-policy probability that the strategy depended on the market's response. Public data may still predict source across different policies or market states. The boundary applies within the pooled policy where hidden payoff components leave the same public record.

Richer linkage data solve part of the problem. Beneficial-owner identifiers, synchronized clocks, and complete order lifecycles can connect a displayed order in one product to an execution in another. That connection establishes provenance and leaves causal attribution unresolved. A common information shock can generate the same display, response, execution, and cancellation sequence. The model demonstrates this observational equivalence with two economies that produce the same passive record and imply different effects of visibility.

A feasible visibility experiment supplies the missing comparison. Candidate orders are assigned to visible and hidden public-feed conditions after their policies and beneficiary actions are locked, while matching and priority remain unchanged. Randomization occurs at a predetermined cluster level so that interference is contained within the assigned unit. This is a partial-interference design in the sense of Hudgens and Halloran (2008). The experiment identifies the average effect of visibility on locked beneficiary outcomes for the enrolled stratum. Individual source labels require validated cases or a separately identified latent-source model. Legal intent further requires evidence on how source dependence predicts intent at submission. This sequence keeps causal identification, predictive classification, and legal inference separate.

Enforcement then changes routing. The equilibrium has four possible regions: no strategic entry, transparent activity only, opaque activity only, or both tactics active. The allocation is unique and changes continuously as expected penalties change, with slope changes when a tactic enters or exits. Within the region where both tactics are used, a higher penalty on one tactic suppresses that tactic, raises activity in the other, and lowers total strategic participation. The suppressed mass divides between true exit and refuge migration. The relative abundance of marginal entrants and the attractiveness of the refuge determine that split.

This decomposition changes the interpretation of enforcement data. A fall in the targeted signature combines two economically different outcomes. Some traders leave because the remaining opportunities no longer cover their participation cost. Others retain the response rent and acquire it through the refuge. Counts in the targeted product reveal the sum of these movements. Global deterrence depends on the exit component.

The routing result has a sharp implication for harm. Total response-dependent activity falls when enforcement tightens within an active region. Social harm can rise when the refuge causes enough harm per unit and attracts enough of the suppressed activity. A regulator that ranks targets by suspicious message counts, predicted spoofing probability, or ease of detection can then allocate audits in the wrong order. The relevant object is the reduction in harm after the equilibrium response of every tactic. An opaque tactic with fewer visible incidents can deserve more enforcement when suppressing it produces substantial exit and little migration. A transparent tactic with many incidents can be a poor target when its users move into a more damaging opaque strategy.

The regulator's problem also includes the supply of legitimate liquidity. Bona fide providers differ in the private cost of quoting. A false source certificate imposes legal, reputational, and operational losses, reducing the surplus from participation. Marginal providers leave as audit exposure rises. Their exit removes private trading surplus and a market-quality benefit created by displayed liquidity. Rapid quote revision and unexecuted limit orders can contribute to liquidity and price discovery (Hendershott et al., 2011; Conrad et al., 2015; Brogaard et al., 2019). The model also accounts for response-dependent traders who lacked an intent to cancel. A source certificate is accurate for their economic type and may still produce a legal error. This distinction prevents the false-sanction cost from being scaled only by the mass of ordinary liquidity providers.

Evidence quality has a nonmonotone effect on false certification. When a demanding test is nearly uninformative, it certifies few cases from either group. Initial gains in precision can raise certification in both groups. Once the evidence becomes sufficiently informative, further precision separates them and reduces false certification. At a fixed deterrence target, better evidence always reduces bona fide providers' exposure, supporting legitimate entry and market quality.

The platform has two complementary tasks. Provenance narrows the beneficiary trades that could be linked to a displayed order. Response reconstruction separates the effect of visibility from common shocks and ordinary hedging. Narrowing the candidate set makes a cleaner response signal more valuable, and a cleaner signal makes provenance more useful. Either partial system may fail to cover its fixed cost even when the completed chain can do so.

Fixed evidence costs govern when the platform is built. While both tactics are active, the value of an opaque-tactic platform rises with the mass available to deter. Below a threshold, the regulator tolerates the refuge. At the threshold, the platform is built and its first optimal use removes the opaque tactic. Part of the displaced activity returns to the transparent tactic and the rest exits. With a larger initial opaque mass, audit coverage eventually reaches its unconstrained level and some opaque activity can remain.

The post-exit problem is different. Continued legacy enforcement can remove the transparent tactic before the evidence platform is adopted. Opaque mass then stops growing. Platform adoption may still occur later. The force behind this delay is enforcement leakage. Auditing the opaque tactic can make the transparent tactic attractive again, which reduces the marginal value of opaque enforcement. As legacy enforcement becomes stronger, transparent re-entry requires a larger policy change and the leakage weakens. The platform can cross its adoption threshold even though the stock of opaque activity has remained constant.

This mechanism produces a distinctive transition at delayed adoption. The first platform investment after transparent exit generally reactivates the transparent tactic or leaves it at the edge of re-entry, while only partially suppressing opaque activity. Complete removal occurs in a terminal corner where neither tactic remains attractive. Platform adoption can therefore leave the conduct that motivated it in place at first. Its value comes from changing the mix of tactics and reducing harm over the full equilibrium path.

These cases form a phase diagram in legacy enforcement and platform cost. Low platform cost brings investment during coexistence. Intermediate cost delays it until transparent activity has exited and re-entry leakage has fallen. High cost prevents investment. Section 5 and the Appendix characterize the resulting active sets and alternative parameter orderings.

The final policy comparison concerns a cancellation rule. Holding the full vector of expected tactic penalties fixed makes routing and manipulation harm identical, which isolates implementation costs. Source enforcement pays for evidence, audit coverage, false sanctions, and any loss of legitimate liquidity. A cancellation rule carries its own administrative burden, exceptions, queue costs, and effect on bona fide cancellations. Source enforcement has higher welfare when its information advantage saves enough legitimate liquidity and legal error to cover the platform and audit costs. A cancellation rule can prevail when evidence is expensive, weakly related to legal intent, or narrowly targeted at a tactic whose users migrate to a harmful refuge. A broad rule can also reduce refuge opportunities by affecting several tactics together. The welfare ranking depends on these primitives.

These mechanisms yield three empirical signatures. Targeted activity falls more than global response-dependent activity when traders migrate to a refuge. Tactic masses remain continuous at ordinary entry and exit boundaries, whereas platform adoption can move them discretely. After transparent activity disappears, later platform adoption should follow a decline in transparent re-entry risk rather than renewed growth in opaque activity. Section 7 develops the designs and data requirements.

Research on spoofing establishes its private return and its sensitivity to market design. Lee et al. (2013) use account-level orders and a disclosure-rule change to document strategic spoofing. Cartea et al. (2020) solve a trader's spoofing decision under an expected fine, while Williams and Skrzypacz (2025) show when spoofing survives in equilibrium and how cancellation restrictions trade deterrence against legitimate trading. Cartea et al. (2025) study when learning algorithms discover order-book manipulation. This paper makes evidence architecture part of the strategic environment. The architecture changes entry, routing, and the population of tactics that the regulator observes.

Empirical work identifies suspicious order-book patterns and measures market-quality effects. Do and Putniņš (2023) build detection methods from prosecuted cases, while Brogaard et al. (2026) use identified order-book data to estimate the effect of spoofing on market quality. Enforcement cases reveal a selected portion of underlying manipulation (Comerton-Forde and Putniņš, 2014). The theory asks what additional variation identifies the economic source of profit, how source evidence maps into a legal inference, and how enforcement changes the tactics represented in both detected and undetected activity.

The distinction between economic source and legal intent responds to a long-standing difficulty in separating manipulation from facially valid trading (Fischel and Ross, 1991; Kyle and Viswanathan, 2008). Allen and Gale (1992) and Aggarwal and Wu (2006) study when profitable manipulation survives. Positions across related markets create the beneficiary channel in models of cash-settlement and derivative manipulation (Kumar and Seppi, 1992; Zhang, 2022), and Stenfors et al. (2023) study cross-market spoofing. Brunnermeier and Pedersen (2005) show how a trader can profit from another trader's response and how that response propagates across markets. The induced response creates the rent here; the evidence system determines where that rent can be earned.

The routing mechanism also belongs to the economics of enforcement. Offenders can spend resources to reduce detection (Malik, 1990), and a penalty schedule changes which harmful act they choose (Mookherjee and Png, 1994). In finance, tighter regulation can shift activity toward a less regulated and less transparent sector (Plantin, 2015; Buchak et al., 2018). The refuge in this paper is a trading tactic. Its appeal comes from the difficulty of reconstructing the beneficiary chain.

Evidence design connects this routing problem to proof burdens and market quality. Kaplow (2011) studies deterrence and the chilling of desirable conduct when proof burdens and enforcement effort are chosen jointly. Research on securities regulation shows that rule content, implementation, and enforcement all matter for liquidity (Cumming et al., 2011; Christensen et al., 2016). The present model identifies the information system through which enforcement changes both strategic conduct and ordinary quoting.

Three boundaries guide the interpretation. Response dependence is an economic source classification, and legal intent remains a separate state. A visibility experiment identifies an average causal effect within its enrolled population, while individual source likelihoods require validated labels or an identified latent model. The platform problem is a stationary comparison across enforcement environments with an amortized fixed cost. Irreversible investment, learning over time, and strategic destruction of evidence would add dynamic forces beyond the current model. These boundaries locate the claims that the evidence can support.

Section 2 describes the beneficiary chain, public pooling, and the source-to-intent bridge. Section 3 solves entry, routing, and displacement. Section 4 develops the evidence technology and legitimate-liquidity margin. Section 5 characterizes platform adoption. Section 6 compares enforcement designs, and Section 7 develops the empirical implications. All proofs, boundary cases, and active-set solutions appear in the Appendix.

2 The Evidence Problem

The order book records submission, execution, modification, and cancellation. Its record omits the payoff decomposition that made display valuable. This section describes that missing object and the evidence required to recover it.

2.1 A candidate order and its beneficiary trade

Consider a trader who displays a large order in product A. Other traders respond by revising quotes, submitting marketable orders, or moving liquidity away from the displayed side. During that response, the same beneficial owner receives a favorable execution in related product B. The candidate order in A is then withdrawn. The public sequence is suggestive because display precedes response, the response precedes the beneficiary execution, and withdrawal follows the execution.

Positions in one market can make price-moving activity in another market profitable in models of cash-settlement and derivative manipulation (Kumar and Seppi, 1992; Zhang, 2022). Cross-market spoofing uses the same broad beneficiary channel (Stenfors et al., 2023). The question here is what evidence attributes the beneficiary payoff to the displayed order.

The sequence remains consistent with an ordinary trading strategy. The candidate order may have offered execution value when it was submitted. A common information shock may then have changed prices in both products, produced the execution in B, and made the order in A stale. Order-lifecycle linkage and synchronized timestamps would connect the events when the records also establish common ownership. The Consolidated Audit Trail provides a partial analogue for U.S. NMS securities by linking reportable events through an order's lifecycle, attaching persistent account and discretionary-trader codes, and requiring synchronized clocks (U.S. Securities and Exchange Commission, 2012). It does not by itself span every related product, recover locked beneficiary exposure, or identify the visibility counterfactual.

The economically relevant counterfactual asks whether the beneficiary trade would retain its value if responders could not observe the candidate order. Let $M = 1$ denote a response-necessary strategy, one whose profit requires the market response induced by display. Let $M = 0$ denote an execution-rationalizable strategy. The classification concerns the source of value. It is sharper than a label based on cancellation because strategies in both classes can submit and cancel the same displayed order.

Legal intent is a separate state. Let $I = 1$ denote an intent to cancel at submission. The four combinations of economic source and legal intent are feasible. In particular, a response-necessary strategy can lack the relevant intent, and a trader with a cancellation plan can submit an order that remains profitable under an execution rationale. This separation follows the legal importance placed on intent at submission in U.S. futures enforcement (Commodity Futures Trading Commission, 2013) and the broader difficulty of distinguishing manipulation from ordinary trading (Fischel and Ross, 1991; Kyle and Viswanathan, 2008). The model treats the legal standard abstractly and makes no claim about the resolution of any individual case.

Figure 1 separates the event sequence from the inference built on it. The exchange observes the display, the market response, the beneficiary execution, and the withdrawal. Provenance identifies whose execution belongs in the chain. Visibility variation asks whether the display caused the beneficiary outcome. The source-to-intent bridge determines what that causal evidence implies for the legal state. A complete record of Panel A can therefore leave every question in Panel B open.

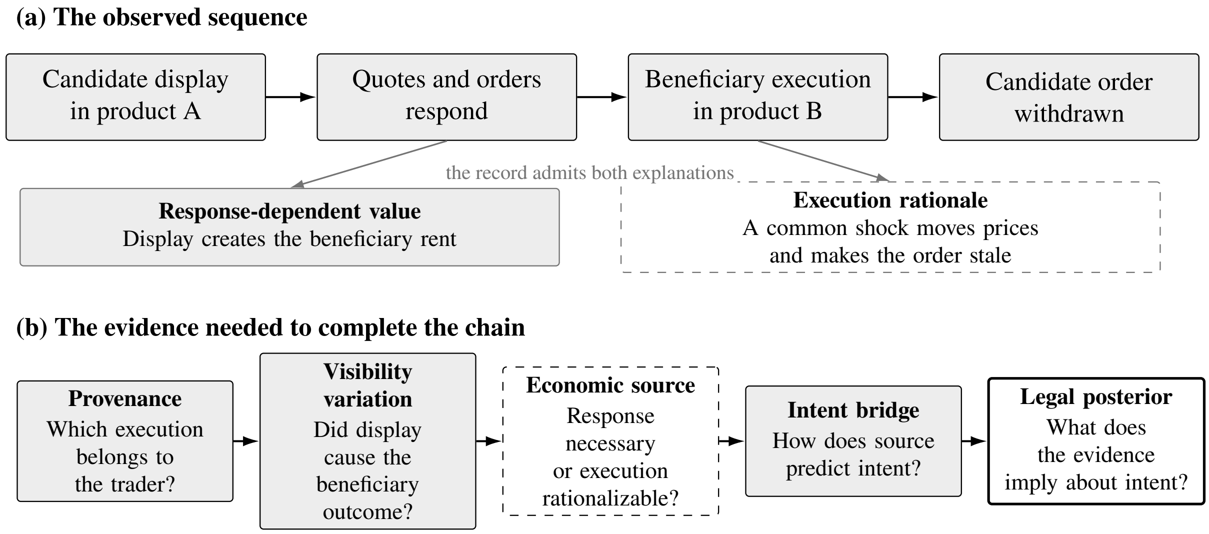

Figure 1: From an observed sequence to a legal inference

Notes: Panel A records the public sequence around a candidate order. That sequence is consistent with response-dependent value and with a common shock followed by ordinary execution and cancellation. Panel B shows the three evidentiary links needed to move from the candidate order to a legal inference. Provenance links the order to a beneficiary trade, visibility variation identifies the effect of display on the beneficiary outcome, and the bridge maps economic source evidence into an intent posterior.

2.2 What the public history can reveal

Fix a displayed policy p, such as a rule for size, price, queue position, modification, and cancellation. Suppose both source types choose it and generate the same conditional distribution of public histories. The policy then pools economically different traders. Differences across policies remain informative, while the history of a trader inside this pool carries no source likelihood.

Proposition 1 (Public histories and source evidence)**.** *Suppose both source types select the same displayed policy and have the same conditional public-history law. Public history then carries no source likelihood within that policy. Random assignment of visibility under the stated design conditions identifies the average effect of visibility on beneficiary outcomes for the enrolled stratum. Response necessity also requires the strategy-surplus comparison. Legal intent additionally requires the source-to-intent bridge.*

Proposition 1 marks the limit of pattern surveillance. A cancellation score can remain useful for screening across policies or market states. Within the pooled policy, the score cannot learn which hidden payoff source generated the order. More public messages refine the path of the same observable object.

Passive linkage fills a different gap. It establishes which beneficiary trade belongs to the candidate order. The causal gap remains because visibility is endogenous to the trading episode. Appendix B constructs two economies with the same joint distribution of visibility, response, beneficiary outcome, withdrawal, ownership, and time, yet with different visibility effects.

2.3 A feasible source experiment

The missing comparison can be produced by varying public visibility after the relevant decisions have been locked. Candidate orders are enrolled before their outcomes are known and assigned at the level of predetermined interference clusters. Both arms face the same feasibility, matching, and priority rules; assignment changes only what other traders can see. Treatment can spill over inside a cluster while remaining isolated across clusters, the partial-interference structure of Hudgens and Halloran (2008). The visible-hidden difference in beneficiary outcomes identifies the average effect of visibility for the enrolled stratum. Individual source labels still require validated cases or an identified latent-source model. Mapping source evidence to legal intent requires one further step.

Write π for the source prevalence in a preregistered reference stratum and let ρ_m be the probability of legal intent conditional on source state m. Under conditional independence of source evidence E and intent given source, Bayes' rule yields

$$\mathbb{P}(I = 1 \mid E = e) = \frac{\rho_1 \pi \Lambda(e) + \rho_0(1 - \pi)}{\pi \Lambda(e) + 1 - \pi}, \tag{1}$$

where $\Lambda(e)$ is the source likelihood ratio. When $\rho_1 > \rho_0$, stronger source evidence raises the intent posterior. Its range remains bounded by ρ_0 and ρ_1. Source evidence is legally uninformative when the two bridge probabilities are equal, and it cannot meet a posterior burden above ρ_1. Appendix B derives every ordering and equality case.

The beneficiary chain therefore has three links. Provenance connects the candidate to the beneficiary. Visibility variation identifies the market response caused by display. The bridge describes the legal force of that source evidence. Losing any link changes what the evidence can establish.

3 Enforcement and the Location of Manipulation

Source evidence matters because traders choose their exposure to it. A surveillance system aimed at one visible pattern changes the relative return to every substitute that can earn the same response rent. Costly efforts to reduce detection make avoidance part of the enforcement problem (Malik, 1990). This section solves that routing decision.

3.1 Tactics, entry, and congestion

There are two tactics. The transparent tactic, denoted T, leaves a beneficiary chain that the current enforcement system can reconstruct at relatively low cost. The opaque tactic, denoted O, preserves more execution-based explanations by changing the product, venue, account structure, timing pattern, or order design. These labels describe evidentiary exposure. An opaque tactic may create a conspicuous order while concealing the source of its profit.

Tactic $x \in \{\mathrm{T}, \mathrm{O}\}$ has pre-enforcement attractiveness v_x and expected penalty τ_x. Its net attractiveness is $A_x = v_x - \tau_x$. The opportunity weight $w_x > 0$ measures how much activity the tactic can absorb before its response rent is crowded out. The routing masses contain response-necessary strategies; execution-rationalizable liquidity is tracked separately in the regulator's error and participation problem. Potential strategic entrants have a constant density $\lambda > 0$ over private participation costs. In equilibrium, all active tactics offer the same cutoff surplus z, and entrants with costs below z participate. The allocation solves

$$m_x = w_x \left[A_x - z\right]_+, \qquad m_\mathrm{T} + m_\mathrm{O} = \lambda z. \tag{2}$$

The first relation says that congestion absorbs the gap between a tactic's attractiveness and the common cutoff. The second clears the participation margin. A larger cutoff draws in more strategic traders. The positive parts allow no entry, either tactic alone, or both tactics together.

The system has a unique solution because aggregate tactic demand falls strictly with the common cutoff while entry supply rises. When both tactics are active, define $D = \lambda + w_\mathrm{T} + w_\mathrm{O}$. The masses are

$$m_\mathrm{T} = \frac{w_\mathrm{T}\{(\lambda + w_\mathrm{O})A_\mathrm{T} - w_\mathrm{O}A_\mathrm{O}\}}{D}, \tag{3}$$

$$m_\mathrm{O} = \frac{w_\mathrm{O}\{(\lambda + w_\mathrm{T})A_\mathrm{O} - w_\mathrm{T}A_\mathrm{T}\}}{D}. \tag{4}$$

Appendix D gives the necessary and sufficient inequalities for all four regions and verifies continuity at every boundary.

3.2 Exit, migration, and harm

Consider an increase in the expected penalty on active target j, with the other active tactic k serving as the refuge. Differentiating equations (3) and (4) gives

$$\frac{\partial m_j}{\partial \tau_j} = -\frac{w_j(\lambda + w_k)}{D}, \qquad \frac{\partial m_k}{\partial \tau_j} = \frac{w_j w_k}{D}. \tag{5}$$

The targeted tactic contracts and the refuge expands. Total strategic activity falls by $\lambda w_j / D$ per unit of penalty. For every unit removed from the target, the migration and exit shares are

$$\underbrace{\frac{w_k}{\lambda + w_k}}_{\text{migration}} \quad \text{and} \quad \underbrace{\frac{\lambda}{\lambda + w_k}}_{\text{exit}}. \tag{6}$$

The decomposition is finite as well as local because the masses are affine within the coexistence region. It remains valid until the target reaches its active-set boundary.

The migration share is the model's marginal-deterrence object. A change in the relative penalty schedule reallocates conduct among harmful acts before it removes conduct from the market, a general force studied by Mookherjee and Png (1994).

Let $h_x > 0$ be harm per unit of tactic x and let $H = h_{\mathrm{T}} m_{\mathrm{T}} + h_{\mathrm{O}} m_{\mathrm{O}}$. The displacement-adjusted harm reduction from penalizing target j is

$$G_j = \frac{w_j \{h_j(\lambda + w_k) - h_k w_k\}}{D}, \qquad \frac{\partial H}{\partial \tau_j} = -G_j. \tag{7}$$

An enforcement action can therefore lower total activity and raise harm. This occurs when the refuge attracts enough displaced mass and causes sufficiently greater harm per unit.

Proposition 2 (Enforcement and migration)**.** *The routing system in equation* (2) *has a unique, continuous, piecewise-affine equilibrium. In a strict coexistence region, a higher penalty on one tactic reduces the target, expands the refuge, and lowers total strategic participation according to equation* (5)*. The removed target mass divides into the migration and exit shares in equation* (6)*. Enforcement raises aggregate harm exactly when $G_j < 0$. In a single-tactic region, a higher penalty on the active tactic produces exit until either the tactic is eliminated or the refuge enters.*

Figure 2 displays the proposition's three economic objects. The boundaries in Panel A show that a penalty can move activity through several active sets even though the equilibrium remains continuous. Panel B follows that movement along a horizontal slice. When the opaque tactic enters, half of each additional unit removed from the transparent tactic migrates and half exits under the symmetric calibration. Panel C adds harm. Equal harm per unit makes aggregate harm fall with total activity. Twice the harm in the refuge makes enforcement locally harm neutral. A still more harmful refuge turns the slope positive.

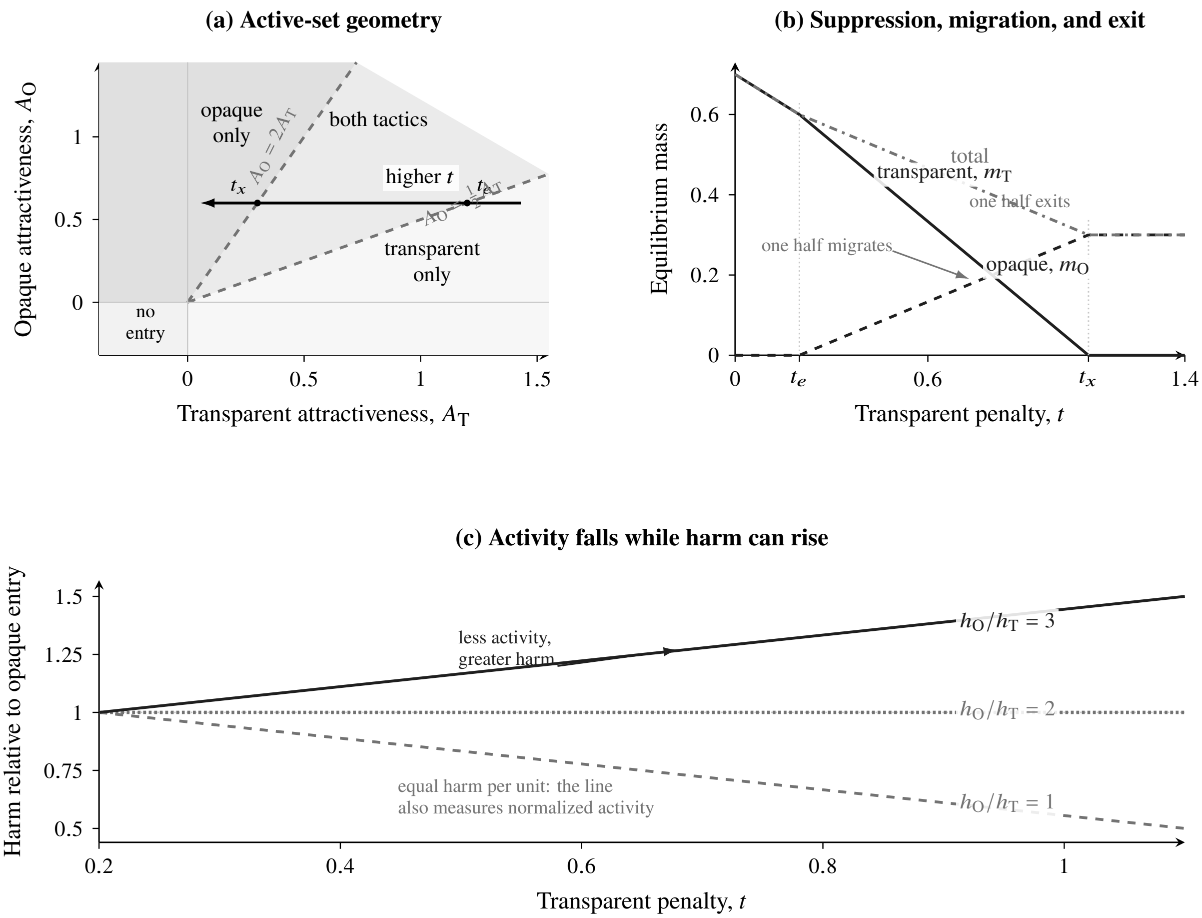

Figure 2: Routing, migration, and harm

Notes: Panel A gives the active-set regions in tactic-attractiveness space and traces a rise in the transparent penalty t. Panel B plots the resulting tactic masses and total strategic activity. The opaque tactic enters at $t_e = 0.2$, and the transparent tactic exits at $t_x = 1.1$. Panel C normalizes harm to one at opaque entry and varies opaque harm relative to transparent harm. The calibration sets $\lambda = w_T = w_O = h_T = 1$, $A_T = 1.4 - t$, and $A_O = 0.6$. All curves are analytical solutions of equation (2).

The proposition changes how local enforcement statistics should be read. A decline in the targeted signature reports target suppression. Its migration component remains active elsewhere. The exit component measures global deterrence. A regulator can record fewer suspicious orders in the monitored product while the market inherits a more harmful tactic mix.

The same logic changes audit priorities. Raw incident counts give weight to where activity is visible. The welfare gain from an audit uses G_j, the probability that an audit produces usable evidence, the cost of scrutiny, and the legal and liquidity losses created by error. An opaque cell with fewer detectable incidents can have the larger enforcement value because its suppression produces more exit and less migration. Section 6 makes this comparison explicit.

4 Producing Source Evidence

The routing result treats expected penalties as given. A regulator must first produce evidence that can support them. Information quality determines what an audit can learn. Coverage determines how often the evidence process is applied. Keeping these margins separate is central to enforcement design. Their joint effect on deterrence and legitimate conduct parallels the economics of proof burdens (Kaplow, 2011).

4.1 Information and coverage

Let E be the evidence produced by an audited case. Its conditional laws under response-necessary and execution-rationalizable conduct are P_1 and P_0. A case receives a source certificate when its log likelihood ratio crosses a fixed hurdle $b > 0$. Conditional on audit, define

$$\eta = \mathbb{P}(\text{certificate} \mid M = 1), \qquad \alpha = \mathbb{P}(\text{certificate} \mid M = 0). \tag{8}$$

The first rate governs deterrence of response-necessary conduct. The second governs exposure of execution-rationalizable conduct. If coverage of tactic x is a_x and a certified case loses $F > 0$, the expected tactic penalty is

$$\tau_x = \tau_x^0 + F a_x \eta. \tag{9}$$

Here τ_x^0 is legacy enforcement. Information changes η and α. Coverage scales the population exposed to those rates.

For a transparent benchmark, suppose the source log likelihood ratio has information index $J > 0$ and follows

$$L \mid M = 1 \sim \mathcal{N}(J/2, J), \qquad L \mid M = 0 \sim \mathcal{N}(-J/2, J). \tag{10}$$

This experiment arises from two normal signals with a common variance. The certification rates are

$$\eta(J, b) = \Phi\left(\frac{\sqrt{J}}{2} - \frac{b}{\sqrt{J}}\right), \qquad \alpha(J, b) = \Phi\left(-\frac{\sqrt{J}}{2} - \frac{b}{\sqrt{J}}\right). \tag{11}$$

True-source certification rises with information. False-source certification first rises and then falls, reaching its maximum at $J = 2b$. When the experiment contains almost no information and the hurdle is demanding, it certifies almost nobody. Early improvements move observations from both states across the hurdle. Stronger experiments eventually separate the states.

The ratio relevant for legitimate burden behaves monotonically. Suppose the regulator holds the expected strategic penalty at a target $t > 0$, so required coverage is $a = t/(F\eta)$. If a false certificate imposes private loss f_x, a bona fide provider bears

$$p_x(J;t) = \frac{f_x t}{F}\frac{\alpha(J,b)}{\eta(J,b)}. \tag{12}$$

The ratio α/η falls strictly with information. A regulator can therefore preserve deterrence while reducing the burden on ordinary liquidity supply.

4.2 Completing the beneficiary chain

The evidence platform performs two tasks. Provenance narrows the set of candidate beneficiary trades; source reconstruction cleans the signal that attributes their outcomes to display. Under the maintained Gaussian candidate-average architecture, each task increases the information value of the other. Provenance makes reconstruction more useful because the causal signal is spread across fewer candidates. Reconstruction makes provenance more useful because a matched candidate carries a cleaner source signal. Appendix C derives the cross-difference and classifies the correlated-signal cases.

Proposition 3 (Evidence precision)**.** *Under the Gaussian experiment in equation* (10) *with a fixed positive hurdle, true-source certification rises with J, false-source certification peaks at $J = 2b$, and η/α rises strictly for every $J > 0$. At a feasible fixed deterrence target, the private burden in equation* (12) *falls strictly with information. Under the conditionally Gaussian candidate-average architecture, provenance and source reconstruction are strict information complements.*

Figure 3 shows why the false-certification result and the liquidity result differ. Panel A holds the hurdle fixed. A nearly uninformative experiment rarely crosses it under either source state, so initial improvements can raise false certification. Beyond $J = 2b$, separation dominates and the false rate falls. Panel B holds deterrence fixed instead. Better information requires less coverage and lowers false exposure per unit of deterrence throughout. Panel C carries that ratio into the participation decision: the high-information system preserves more legitimate supply at every positive target penalty.

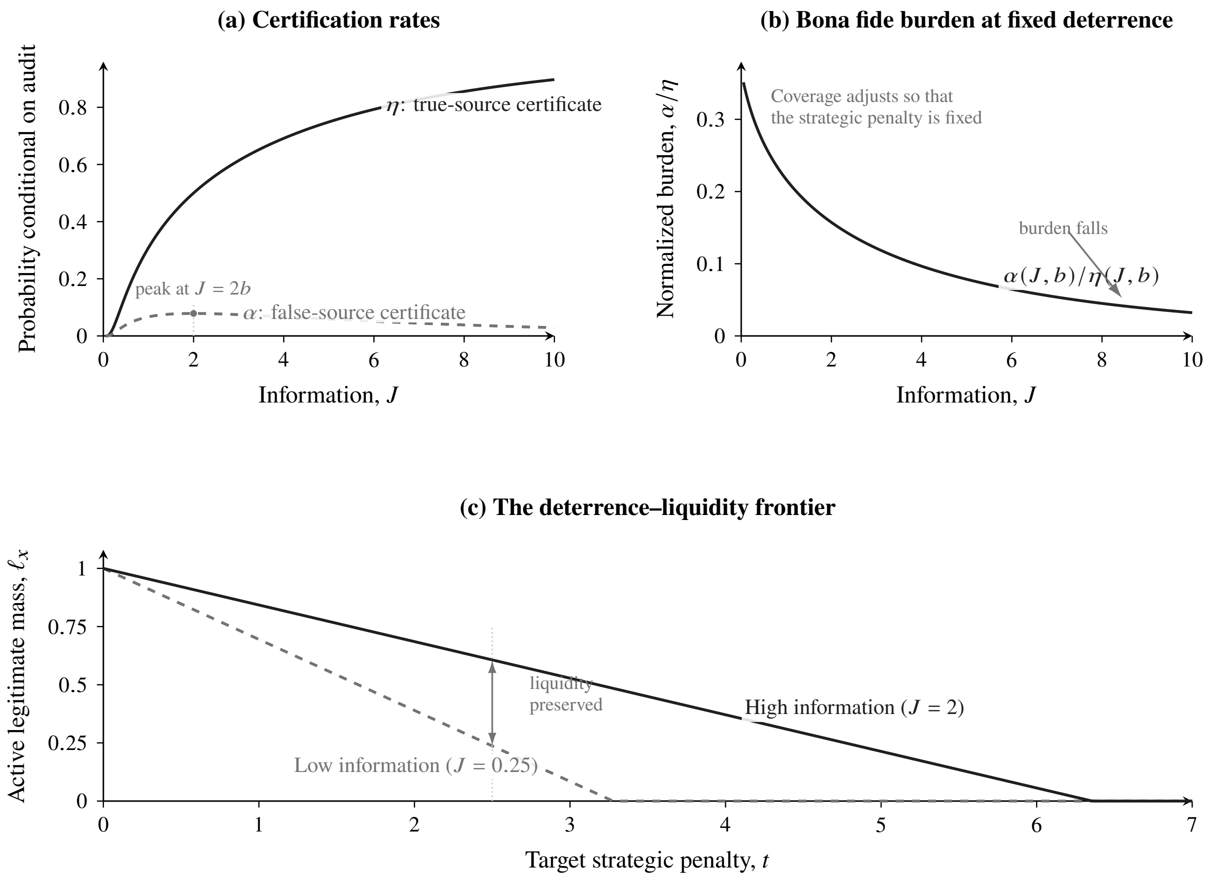

Figure 3: Information, certification, and legitimate liquidity

Notes: Panel A plots the true-source rate η and false-source rate α from equation (11) with hurdle $b = 1$. Panel B plots α/η, which is proportional to bona fide burden at a fixed deterrence target. Panel C normalizes legitimate mass to one at zero burden and compares information indices $J = 0.25$ and $J = 2$. The target penalty changes required coverage according to equation (12).

The complementarity statement belongs to the maintained compression. If the platform retains a full vector of correlated signals, two evidence legs can be complements, substitutes, or redundant. Appendix C gives the general two-channel Gaussian classification. Informational complementarity also differs from investment complementarity. A large fixed cost, a binding coverage cap, or an enforcement target that has already exited can prevent an informative bundle from creating operating value.

4.3 The regulator's operating problem

The regulator chooses architecture and coverage before traders route. Architecture determines information and fixed cost; coverage determines penalties, routing, and legitimate exposure. Section 6 introduces legitimate participation and legal-error costs, while Appendix E gives the full objective and boundary conditions.

Inside a fixed routing region, the marginal harm benefit of coverage is $F\eta G_x$, where G_x is the displacement-adjusted gradient from equation (7). This term already includes refuge migration. The evidence platform becomes valuable when that benefit, net of error and coverage costs, is large enough to cover its fixed cost. The next section characterizes when that crossing occurs.

5 When Evidence Is Built

Evidence infrastructure is costly before a single case is audited. A regulator can therefore tolerate an opaque refuge even when the platform would improve each audited case. Adoption occurs when the activity exposed to the new evidence, together with the routing response to its enforcement, makes the operating value large enough. The closed-form threshold holds screened legitimate and response-necessary non-intender pools fixed, which keeps net marginal coverage benefits constant within each routing cell. Section 6 and Appendix E give the endogenous-liquidity extension and its boundary candidates.

5.1 Adoption while both tactics are active

Consider a platform aimed at the opaque tactic during coexistence. Let $\varphi = F\eta$ be the penalty created by one unit of coverage. From equation (5), opaque mass falls by

$$Q_B = \varphi \frac{w_{\mathrm{O}}(\lambda + w_{\mathrm{T}})}{D} \tag{13}$$

per unit of coverage. If m is the counterfactual opaque mass before the platform is installed, useful coverage cannot exceed m/Q_B. Let $B_B > 0$ be the net marginal value of coverage after displacement-adjusted harm, direct audit expense, false-source loss, and legal error have been combined. With quadratic cost $\kappa a^2/2$, gross operating value is

$$S(B_B, m) = \max_{0 \le a \le \min\{1, m/Q_B\}} \left\{ B_B a - \frac{\kappa}{2} a^2 \right\}. \tag{14}$$

The platform with fixed cost K is installed when $S(B_B, m) \ge K$. Define $\widehat{a} = \min\{B_B/\kappa, 1\}$. If K is no larger than the maximum operating value, the first coverage level that can cover the cost is

$$a_K^- = \frac{B_B - \sqrt{B_B^2 - 2\kappa K}}{\kappa}, \qquad m_K = Q_B a_K^-. \tag{15}$$

The economically relevant root lies on the rising side of the coverage-value curve. Adoption begins when opaque mass reaches m_K.

At that point, the mass cap binds: $m_K = Q_B a_K^-$. The platform's first positive-cost use removes the opaque tactic. The routing result then sends the fraction $w_T/(\lambda + w_T)$ of removed opaque activity into the transparent tactic. The remaining fraction exits. With further growth in counterfactual opaque activity, coverage eventually reaches $\widehat{a}$ and the opaque tactic can return.

5.2 Adoption after transparent exit

The platform problem changes once legacy enforcement has removed the transparent tactic. Let t be the legacy transparent penalty, let $v_T - t$ be transparent net attractiveness, and hold pre-platform opaque net attractiveness at $u > 0$. Without the platform, transparent activity exits at

$$t_x = v_T - \frac{w_O u}{\lambda + w_O}. \tag{16}$$

For $t \geq t_x$, opaque activity is the only active tactic and its mass is constant as t rises.

Opaque coverage can make the transparent tactic attractive again. Starting from the opaque-only region, the coverage required for transparent re-entry is

$$r(t) = \frac{\lambda + w_O}{\varphi w_O}(t - t_x). \tag{17}$$

Below this boundary, opaque suppression produces pure exit. Above it, part of further suppression leaks into transparent entry. Stronger legacy enforcement moves the boundary outward and allows more opaque coverage to operate before leakage begins.

Let B_0 be net marginal coverage value while the opaque tactic is alone and let B_B be its value after transparent re-entry. The benchmark ordering is $0 < B_B < B_0$: migration absorbs part of the harm benefit in coexistence. The regulator's optimized operating value then rises with $r(t)$ even though opaque mass is unchanged. A platform that failed to cover its cost at transparent exit can become worthwhile later because enforcement creates more true exit and less re-entry.

Proposition 4 (Evidence investment)**.** *Suppose the fixed-pool benchmark applies and the regulator faces a positive fixed platform cost and quadratic coverage cost.*

(i) During coexistence, adoption follows the mass threshold in equation (15)*. At first adoption the opaque tactic is removed, with its users divided between transparent migration and exit.*

(ii) After transparent exit, suppose $0 < B_B < B_0$ and the unconstrained coverage choices lie inside the relevant re-entry and tactic-kill boundaries. Let $S_B = B_B^2/(2\kappa)$ be operating value when re-entry begins at once, and let S_0 be maximum operating value when transparent re-entry is unavailable. If $K \leq S_B$, the platform is built during coexistence. If $S_B < K \leq S_0$,

adoption occurs after transparent exit as the re-entry boundary rises. If $K > S_0$, the platform is not built along the stationary enforcement path. These build statements are intersected with the policy domain $t \geq 0$. A negative implied crossing means that the platform is already installed at $t = 0$.

(iii) *First delayed adoption generally leaves opaque activity positive. It either reactivates the transparent tactic or places that tactic at its entry boundary. Complete elimination at first delayed adoption occurs only at the terminal corner in which both tactics have zero net attractiveness.*

The delayed result is the central implication of the platform model. The regulator can invest later with no renewed growth in the target population. The value rises because the same opaque audit produces a different equilibrium response. Looking only at opaque counts would miss the force that triggers adoption.

Figure 4 exposes the force behind the delay. Before transparent re-entry, an opaque audit produces true exit and its marginal value follows the upper branch in Panel A. Once coverage reaches $r(t)$, transparent activity returns and absorbs part of the suppression. The marginal benefit drops. Stronger legacy enforcement moves this drop to the right. Panel B shows the resulting rise in optimized operating value. The value crosses the fixed platform cost even though the pre-platform opaque mass remains fixed. Panel C shows the first policy change at that crossing. Opaque activity falls, some transparent activity returns, and total strategic activity declines.

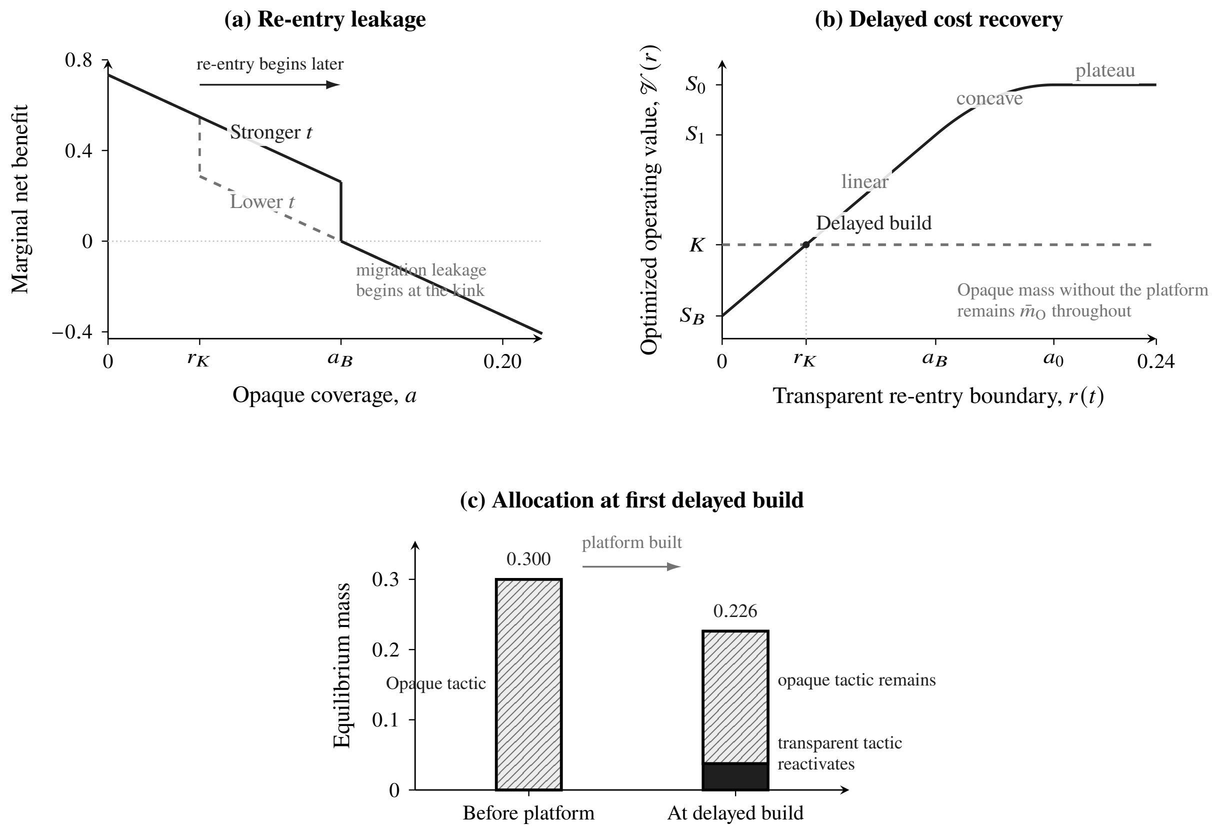

Figure 4: Why evidence investment can be delayed

Notes: Panel A plots marginal operating value against opaque coverage. Its downward jump at $r(t)$ reflects transparent re-entry. Panel B gives the optimized value $V(r)$ and a fixed cost $K = 0.04$. Panel C compares tactic masses immediately before and after the first delayed build. The calibration sets $B_B = 0.4722$, $B_0 = 0.7334$, $\kappa = 4$, $\lambda = w_T = w_O = 1$, $u = 0.6$, and $\varphi = 1.5675$. The build occurs at $r_K = 0.0464$; opaque mass falls from 0.3000 to 0.1888, transparent mass rises from zero to 0.0374, and total activity falls to 0.2262.

Figure 5 places every first-build threshold in the (t, K) plane. Low platform cost produces adoption while the tactics coexist. At S_B, the first-build date jumps because a platform that cannot recover its cost in coexistence must wait until the transparent tactic has exited. The delayed frontier then rises as stronger legacy enforcement pushes the re-entry boundary outward. Its kink at S_1 marks the point where the constrained optimum passes the coexistence coverage choice. At S_0, even the largest operating value available without re-entry fails to support a more expensive platform.

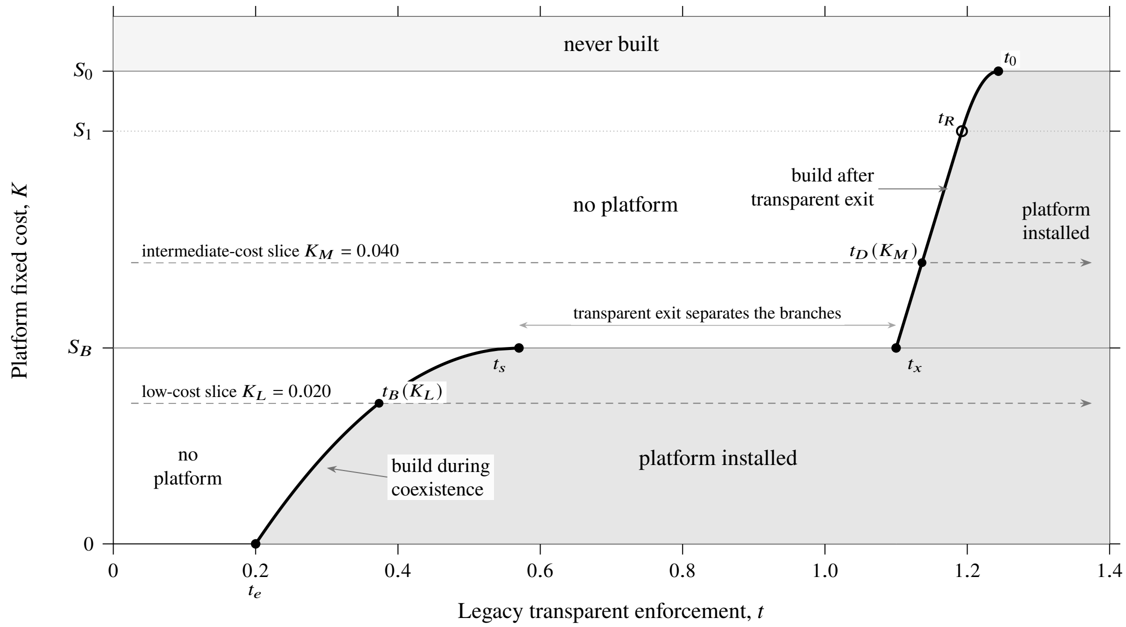

Figure 5: Legacy enforcement, fixed cost, and the timing of evidence investment

Notes: The figure gives the first platform build in legacy transparent enforcement t and fixed cost K. The platform is installed to the right and below the frontier. The lower branch describes adoption during coexistence; the upper branch describes delayed adoption after transparent exit. Thresholds are $t_e = 0.2$, $t_s = 0.5701$, $t_x = 1.1$, $t_R = 1.1925$, $t_0 = 1.2437$, $S_B = 0.0279$, $S_1 = 0.0587$, and $S_0 = 0.0672$. The remaining parameters match Figure 4.

Appendix F proves the threshold rules, enumerates every coverage corner, derives the exact delayed-adoption date, and gives the complete stationary phase theorem. Other orderings of B_0 and B_B can generate adoption only after exit, adoption only during coexistence, a bounded adoption interval, or no useful coverage. The aligned ordering in Proposition 4 gives a clean benchmark rather than a universal sequence.

6 Market Quality and Policy Design

The regulator affects response-necessary display and bona fide liquidity supply. Expected sanctions reduce the first; evidentiary error can reduce the second by making ordinary quoting costly. A policy that ignores the second margin can improve its enforcement statistic while weakening the market it is meant to protect.

This liquidity margin is economically material. Algorithmic trading can narrow spreads and improve quote informativeness (Hendershott et al., 2011). High-frequency quote revision is associated with more efficient prices and lower trading costs (Conrad et al., 2015), and unexecuted limit orders carry substantial price discovery (Brogaard et al., 2019). Evidence from securities regulation likewise distinguishes the rules on the books from their implementation and enforcement (Cumming et al., 2011; Christensen et al., 2016). The model gives these effects a common margin: evidentiary error changes the participation value of ordinary quoting.

6.1 Legitimate participation

Let a bona fide provider in tactic cell x earn gross quoting surplus s_x and draw a private participation cost from a distribution with constant density ν_x. The expected private burden from source enforcement is $p_x = f_x a_x \alpha$. Active legitimate mass is

$$\ell_x(a_x, J) = \nu_x \left[s_x - f_x a_x \alpha(J, b)\right]_+ . \tag{18}$$

If each active provider creates external market-quality value q_x^L, the sector's private surplus and external value at burden p are

$$\mathcal{V}_x(p) = \nu_x \left\{ \frac{[s_x - p]_+^2}{2} + q_x^L \left[s_x - p\right]_+ \right\} . \tag{19}$$

Chilling loss is $\mathcal{V}_x(0) - \mathcal{V}_x(p_x)$. Better information lowers p_x at a fixed strategic penalty by Proposition 3, which raises legitimate mass and value before participation constraints bind.

The legal error ledger also includes source-accurate certificates issued to response-necessary traders who lacked an intent to cancel. If $n_{10,x}$ is their screened mass, expected false legal sanctions in that cell scale with $a_x \eta n_{10,x}$. Bona fide false-source errors scale with $a_x \alpha \ell_x$. Conflating these groups would price legal error using the wrong certification rate.

Within a smooth routing and participation region, the regulator compares the harm benefit $F\eta G_x$ with four marginal costs: direct audit expense, legitimate participation loss, false-source adjudication, and sanctions of response-necessary non-intenders. Existence follows from continuity and compact coverage choices. The regional first-order system and boundary comparison appear in Appendix E.

In a fixed legitimate-pool benchmark, define the net coverage benefit in routing region S as

$$B_x^S = F\eta G_x^S - c_x - \chi_x \alpha \bar{\ell}_x - \chi_{10,x} \eta \bar{n}_{10,x}, \tag{20}$$

where c_x is marginal audit cost and the last two terms price legal error. With quadratic curvature κ_x, no binding cap, and zero coverage feasible, operating value is

$$V_x^S = \frac{\left[B_x^S\right]_+^2}{2\kappa_x}. \tag{21}$$

Audit priority follows this value. Visibility and raw mass affect the comparison only through their effect on the terms in equation (20). A small opaque cell can merit the first audit when it has a favorable harm gradient and a precise evidence process.

6.2 Source evidence and cancellation rules

At a common tactic-penalty vector, source enforcement and a cancellation rule generate the same routing and manipulation harm. Source enforcement pays for the evidence platform, audit coverage, false sanctions, and legitimate-liquidity loss. A cancellation rule carries its own administrative burden, exceptions, queue costs, and effects on ordinary cancellation. The label "blunt" alone does not determine those costs.

Source enforcement has higher welfare when its savings in legitimate liquidity and legal error exceed its platform and audit costs. A cancellation rule has higher welfare when its broader reach against refuge tactics offsets its burden on ordinary cancellations. When penalty vectors differ, the displacement-adjusted harm gradients determine the value of each additional penalty. Appendix G gives the exact cost inequalities.

Proposition 5 (Enforcement design)**.** *At a common feasible tactic-penalty vector, source enforcement has higher welfare exactly when its combined platform, audit, legitimate-liquidity, and legal-error costs are lower. Componentwise savings in bona fide burden, administration inclusive of platform cost, and legal error are sufficient for source dominance. The reversed comparisons are sufficient for cancellation-rule dominance. When the policies create different penalty vectors in a coexistence region, each additional penalty is valued by its displacement-adjusted harm gradient. A broader cancellation rule can overcome a source policy's implementation-cost advantage when it suppresses a sufficiently harmful refuge. Neither policy has an unconditional welfare ranking.*

The comparison clarifies the role of precision. At a common deterrence target, information reduces the source policy's burden on bona fide liquidity. Fixed platform cost works in the opposite direction. When source enforcement concentrates on one tactic, displacement can add a third force by changing the penalty vector across tactics.

The model consequently gives no presumption that a more targeted rule has higher welfare. Its advantage comes from the quality of the distinction it can make and the liquidity it preserves. Its limitation comes from the tactics outside its evidentiary reach. A broad rule reaches more substitutes and can burden ordinary cancellations across the same range. Policy design must price both margins in the same equilibrium.

7 Empirical Implications

The model separates local suppression, refuge migration, true exit, evidence investment, and legitimate-liquidity withdrawal. Each margin has a different empirical signature. Their joint measurement is more informative than a count of targeted cancellations.

7.1 Suppression and migration

A tactic-specific enforcement shock should reduce the targeted signature. Once a substitute tactic is active, the same shock should increase activity in products, venues, book locations, or order forms that preserve the response rent with weaker evidentiary exposure. Migration should be stronger where the refuge has greater capacity, where cross-product pass-through is high, and where traders face a relatively inelastic participation margin. The cross-sectional interaction is therefore between pre-policy exposure to the target and the economic substitutability of the refuge.

A practical design can use a dated change in order-to-trade rules, surveillance coverage, cancellation fees, or persistent-noise-creator regimes. Target and refuge signatures should be fixed before examining post-policy outcomes. Synchronized market-by-order data can measure public displacement across tactics and venues. Stable participant or beneficial-owner identifiers are required to trace the same trader across accounts and products.

7.2 Exit and harm

The decline in the target should exceed the decline in total response-dependent activity. Their gap rises with refuge attractiveness. In data without validated source labels, the corresponding exercise uses a prespecified family of target and refuge signatures and treats the result as reduced-form displacement rather than a count of legal violations.

Market quality can deteriorate after total suspicious activity falls. The relevant environments are those in which the refuge has greater price impact, induces more volatility, weakens price discovery, or withdraws depth from a more important product. An empirical system should therefore estimate targeted activity, refuge activity, spreads, depth, price impact, reversals, volatility, and resiliency together. A policy that lowers the first outcome and worsens the remaining market-quality outcomes is consistent with a negative displacement-adjusted harm gradient.

7.3 Entry boundaries and policy kinks

Ordinary tactic entry and exit create continuity in activity levels and changes in slopes. A refuge may show little response to enforcement until its entry boundary is crossed, followed by a linear increase over the local coexistence region. Regression designs that impose one average response across the full policy range can obscure this threshold.

Platform adoption is discrete and can create jumps. During coexistence, first adoption can sharply reduce the opaque tactic and increase the transparent tactic. After transparent exit, first delayed adoption can bring transparent activity back while opaque activity remains positive. A data-linkage rollout, audit-platform launch, or cross-venue information agreement supplies a candidate event. The model predicts the sign of each tactic's jump from its pre-adoption active set.

7.4 Evidence completion

Owner linkage and synchronized clocks should produce their largest enforcement gain when combined with records or variation that recover locked beneficiary exposure and isolate the effect of visibility. The relevant empirical interaction joins two data improvements: candidate provenance and source reconstruction. Under the candidate-average benchmark, each raises the marginal value of the other.

Predictive accuracy on prosecuted cases does not establish the causal visibility effect. Case labels reflect selection into investigation and adjudication; detected and prosecuted manipulation is a small selected share of underlying conduct (Comerton-Forde and Putniņš, 2014). Estimating the source experiment requires randomized visibility, a valid instrument, or a separately defended natural experiment. Estimating the legal bridge requires outcomes that identify intent and a reference prevalence appropriate for deployment.

7.5 Legitimate liquidity

At a common deterrence target, an improvement in source information should reduce audit exposure of bona fide providers and support displayed liquidity. The predicted outcomes are greater quoting participation, depth, fill probability, and resiliency, together with narrower spreads when the added supply is valuable. The identifying comparison holds the expected penalty on response-necessary conduct fixed. A simultaneous change in coverage and information otherwise confounds precision with enforcement intensity.

Blunt cancellation restrictions should have larger effects on ordinary modification and cancellation among providers whose inventory, queue position, and stale-quote risk create legitimate reasons to withdraw. Comparing their response with that under a source-evidence rollout helps measure the liquidity side of Proposition 5.

7.6 Delayed investment

After legacy enforcement has removed transparent activity, the no-platform model predicts a flat opaque mass as enforcement continues to rise. The platform can still be installed later. Its timing should coincide with a higher transparent re-entry barrier, not with renewed opaque growth. At the adoption event, transparent activity can return because opaque enforcement releases part of the response rent back to the old tactic.

This prediction separates the model from a scale-only account of technology adoption. Under a scale account, delayed investment follows renewed growth in the target population. Here the target stock is stable and the enforcement leakage changes. Observing the active-set transition is therefore essential for interpreting the platform decision.

7.7 Data boundaries

Anonymous market-by-order data identify public histories, order lifecycles, book responses, tactic-level signatures, and market-quality outcomes. They can reveal local suppression and public migration. They cannot establish beneficial ownership across accounts, reconstruct locked inventory, identify individual source necessity, or infer legal intent.

Stable owner identifiers supply provenance. Complete synchronized lifecycles establish event order. Inventory and beneficiary records measure the payoff connection. Source-isolating variation supplies the visibility counterfactual. Intent evidence or validated adjudicative outcomes estimate the legal bridge. The model assigns a distinct empirical claim to each record type so that the strength of the conclusion tracks the strength of the evidence.

8 Conclusion

A displayed order attracts a response, a related trade captures the benefit, and the order disappears. The public record preserves that sequence with remarkable precision. It can still leave unresolved whether the display earned value through execution or through the response it induced.

That ambiguity shapes the market before any case is brought. Traders compare the expected enforcement exposure of available tactics and route toward the one that preserves a profitable explanation. Enforcement of a transparent tactic produces a mixture of exit and migration. Total manipulation can fall while harm rises because the remaining activity moves into a more damaging refuge.

Source evidence changes this allocation by completing the beneficiary chain. Provenance links the candidate order to the beneficiary trade. Valid visibility variation recovers the missing counterfactual. A separate bridge determines the legal force of the source conclusion. Better information makes a fixed amount of deterrence less costly to legitimate liquidity.

The platform that produces this evidence has its own equilibrium timing. A low fixed cost supports adoption while transparent and opaque tactics coexist. An intermediate cost can delay adoption until transparent activity has exited. Opaque activity need not grow during the delay. The platform becomes valuable because stronger legacy enforcement reduces the leakage created when opaque audits reactivate the transparent tactic.

The regulator's task is therefore wider than detecting a visible pattern. It must measure where suppressed activity goes, how much truly exits, which tactic carries the resulting harm, and how the evidence process affects ordinary liquidity supply. Apparent success in one order book can conceal a deterioration elsewhere. The location of manipulation is an outcome of enforcement design.

A Primitives, Timing, and Assumptions

This appendix collects the formal environment used in the proofs. The main text introduces each object when its economic role first appears.

A.1 Economic and legal states

The source state is $M \in \{0, 1\}$, where $M = 1$ means that the strategy is profitable only because other traders respond to its display. The legal-intent state is $I \in \{0, 1\}$, where $I = 1$ means that the trader intended cancellation at submission. Let

$$\pi = \mathbb{P}(M = 1), \qquad \rho_m = \mathbb{P}(I = 1 \mid M = m). \tag{22}$$

The benchmark has $0 < \pi < 1$ and $0 \leq \rho_0 < \rho_1 \leq 1$.

Source evidence E has conditional laws P_1 and P_0, with densities f_1 and f_0 on a common dominating measure. Its likelihood ratio is $\Lambda(e) = f_1(e)/f_0(e)$. The bridge restriction is

$$I \perp E \mid M. \tag{23}$$

A legal decision rule compares $\mathbb{P}(I = 1 \mid E = e)$ with an abstract posterior burden $\bar{\rho} \in (0, 1)$.

A.2 Tactics and routing

There are two tactics $x \in \{\mathrm{T}, \mathrm{O}\}$. Tactic x has pre-enforcement attractiveness v_x, expected penalty τ_x, net attractiveness $A_x = v_x - \tau_x$, opportunity weight $w_x > 0$, equilibrium mass m_x, and harm $h_x > 0$ per unit. The masses m_x contain response-necessary entrants from both legal-intent cells. Execution-rationalizable actors enter the public-pooling and legal-error blocks and are not assigned the same routing equation. Potential strategic entrants are nonatomic and have constant cost density $\lambda > 0$ on the solved support. The common entry cutoff is z. The routing system is

$$m_x = w_x \, [A_x - z]_+ \,, \qquad m_\mathrm{T} + m_\mathrm{O} = \lambda z, \qquad z \geq 0. \tag{24}$$

Total strategic mass and harm are

$$\mathcal{M} = m_\mathrm{T} + m_\mathrm{O} = \lambda z, \qquad H = h_\mathrm{T} m_\mathrm{T} + h_\mathrm{O} m_\mathrm{O}. \tag{25}$$

A.3 Evidence, coverage, and legitimate liquidity

An evidence architecture q has information $J(q) \geq 0$ and amortized fixed cost $K(q) \geq 0$. The regulator chooses coverage $a_x \in [0, 1]$. Conditional source-certificate rates are $\eta(J, b)$ under $M = 1$ and $\alpha(J, b)$ under $M = 0$. A certificate imposes private loss $F > 0$ on the strategic trader. Expected penalty is

$$\tau_x = \tau_x^0 + F a_x \eta(J(q), b). \tag{26}$$

Bona fide liquidity providers in cell x have gross quote surplus $s_x > 0$, constant private-cost density $\nu_x > 0$, private loss $f_x > 0$ after a false source certificate, and external market-quality contribution $q_x^L \geq 0$. Their private burden, active mass, and sector value are

$$p_x = f_x a_x \alpha, \tag{27}$$

$$\ell_x(a_x, J) = \nu_x [s_x - p_x]_+, \tag{28}$$

$$\mathcal{V}_x(p_x) = \nu_x \left\{ \frac{[s_x - p_x]_+^2}{2} + q_x^L [s_x - p_x]_+ \right\}. \tag{29}$$

Chilling loss is

$$C_x^{chill}(a_x, J) = \mathcal{V}_x(0) - \mathcal{V}_x(p_x). \tag{30}$$

If $\chi_x \geq 0$ is additional loss per false source certificate among bona fide providers, then

$$D_x^{FP}(a_x, J) = \chi_x a_x \alpha \ell_x(a_x, J). \tag{31}$$

Let $n_{10,x}$ denote screened mass in the $(M, I) = (1, 0)$ cell. If $\chi_{10,x} \geq 0$ is loss per false legal sanction in that cell,

$$D_x^{10}(a, J) = \chi_{10,x} a_x \eta n_{10,x}(a). \tag{32}$$

The general benchmark uses $n_{10,x}(a) = \omega_{10,x} m_x(a)$ for a fixed composition $\omega_{10,x} \in [0, 1]$, so coverage in one tactic can change the non-intender mass routed to the other. Closed-form operating results use policy-invariant screened masses $\bar{\ell}_x$ and $\bar{n}_{10,x}$.

Direct audit cost is $c_x a_x + \kappa_x a_x^2/2$, where $c_x \geq 0$ and $\kappa_x > 0$. The regulator minimizes

$$\mathcal{L}(q, a) = H(m(\tau(q, a))) + K(q) + \sum_x \left\{ c_x a_x + \frac{\kappa_x}{2} a_x^2 + C_x^{chill} + D_x^{FP} + D_x^{10} \right\}. \tag{33}$$

A.4 Timing

The posterior burden and legacy penalties are fixed first. The regulator then chooses and commits to an architecture and coverage. Legitimate providers enter or exit after observing those choices.

Strategic traders observe expected penalties, decide whether to enter, and route across tactics. Orders, responses, beneficiary executions, and withdrawals occur. Audited cases produce source evidence, the posterior rule is applied, and payoffs are realized. Commitment before strategic entry gives equation (26) deterrent content.

A.5 Assumptions used by the closed-form results

A1. Probative bridge. The main case has $0 \leq \rho_0 < \rho_1 \leq 1$. Appendix B also classifies equality and reverse-order cases.

A2. Source experiment. Conditional independence equation (23) holds. Direct records of a cancellation plan form a separate intent experiment.

A3. Reference stratum. The source prior π is the true prevalence in a prespecified observable deployment stratum. A tactic-specific prior that changes with policy requires a tactic-specific hurdle and a new coverage solution.

A4. Domains. All opportunity weights, harm coefficients, densities, sanction magnitudes, private losses, quote surpluses, and audit curvatures stated as positive are strictly positive. Coverage lies in the unit interval. Fixed, linear, external, and adjudicative costs are nonnegative.

A5. Routing. Entrants are nonatomic, their cost density is constant on the solved support, congestion is affine, and the outside option is zero.

A6. Commitment. Architecture and coverage are observed and committed before entry and routing.

A7. Source laws. The source laws are mutually absolutely continuous on the certification support. Calibration uses independently validated source labels, controlled ground-truth episodes, or a separately identified latent-source model.

A8. Visibility design. Assignment is randomized with overlap after order policies and beneficiary positions are locked. Matching, priority, and preassignment feasibility rules are common across arms. Interference is contained within declared clusters, and inference uses the same clusters.

A9. Candidate-average architecture. There are $n \geq 1$ independent records. Conditional on source state and architecture, the candidate scores within each record are independent Gaussian variables with a common state-invariant variance. Exactly one candidate has nonzero source shift θ, and all candidate counts and score variances are positive. Provenance lowers the candidate count and reconstruction lowers variance.

A10. Legitimate supply. The provider cost support contains every stated interior cutoff. Positive parts govern exit.

A11. Fixed-pool operating benchmark. Screened legitimate and response-necessary non-intender masses are policy invariant when the closed-form audit and adoption formulas invoke bars over those masses.

A12. Aligned phase ordering. The global adoption theorem adds the displayed ordering of post-exit and coexistence harm benefits and the stated slack-cap inequalities. Other orderings use the finite candidate solution in Appendix F.

B Identification and the Source-to-Intent Bridge

B.1 The exact bridge

Lemma 1 (Source-to-intent bridge)**.** *Under equations* (22) *and* (23)*,*

$$\mathbb{P}(I = 1 \mid E = e) = \frac{\rho_1 \pi \Lambda(e) + \rho_0(1 - \pi)}{\pi \Lambda(e) + 1 - \pi}. \tag{34}$$

If $\rho_0 < \bar{\rho} < \rho_1$, the posterior burden is met exactly when

$$\log \Lambda(e) \geq b_I := \log \left\{ \frac{(1 - \pi)(\bar{\rho} - \rho_0)}{\pi(\rho_1 - \bar{\rho})} \right\}. \tag{35}$$

The complete boundary classification is given in the proof.

Proof. Conditional independence gives

$$\begin{aligned}
\mathbb{P}(I = 1 \mid e) &= \sum_{m \in \{0,1\}} \mathbb{P}(I = 1 \mid M = m, e)\mathbb{P}(M = m \mid e) && (36)\\
&= \rho_1 \mathbb{P}(M = 1 \mid e) + \rho_0 \mathbb{P}(M = 0 \mid e) && (37)\\
&= \rho_0 + (\rho_1 - \rho_0)\mathbb{P}(M = 1 \mid e). && (38)
\end{aligned}$$

Bayes' rule and the definition of Λ give

$$\mathbb{P}(M = 1 \mid e) = \frac{\pi f_1(e)}{\pi f_1(e) + (1 - \pi) f_0(e)} \tag{39}$$

$$= \frac{\pi\Lambda(e)}{\pi\Lambda(e) + 1 - \pi}. \tag{40}$$

Substituting equation (40) into equation (38) yields equation (34).

Write the right side of equation (34) as $R(\Lambda)$. Differentiation gives

$$R'(\Lambda) = \frac{\rho_1\pi\{\pi\Lambda + 1 - \pi\} - \pi\{\rho_1\pi\Lambda + \rho_0(1-\pi)\}}{\{\pi\Lambda + 1 - \pi\}^2} \tag{41}$$

$$= \frac{\pi(1-\pi)(\rho_1 - \rho_0)}{\{\pi\Lambda + 1 - \pi\}^2}. \tag{42}$$

Thus R rises when $\rho_1 > \rho_0$, is constant when they are equal, and falls when $\rho_1 < \rho_0$. Its endpoint values are

$$R(0) = \rho_0, \qquad \lim_{\Lambda\to\infty} R(\Lambda) = \rho_1. \tag{43}$$

For $\rho_1 > \rho_0$, the inequality $R(\Lambda) \geq \bar{\rho}$ is

$$\pi\Lambda(\rho_1 - \bar{\rho}) \geq (1-\pi)(\bar{\rho} - \rho_0). \tag{44}$$

If $\bar{\rho} \leq \rho_0$, both the range argument and equation (44) show that every evidence realization meets the burden. If $\rho_0 < \bar{\rho} < \rho_1$, division by the positive coefficient $\pi(\rho_1 - \bar{\rho})$ gives the likelihood-ratio threshold in equation (35). If $\bar{\rho} = \rho_1$, equation (44) requires the source posterior to equal one, which occurs only at an infinite likelihood ratio or under a degenerate experiment that identifies $M = 1$ with probability one. If $\bar{\rho} > \rho_1$, the burden lies above the range in equation (43) and cannot be met by source evidence.

For $\rho_1 = \rho_0 = \rho$, equation (34) equals ρ for every e. All evidence realizations meet the burden when $\bar{\rho} \leq \rho$, and none does when $\bar{\rho} > \rho$.

For $\rho_1 < \rho_0$, the posterior falls from ρ_0 to ρ_1. If $\bar{\rho} \leq \rho_1$, every realization meets the burden. If $\bar{\rho} > \rho_0$, none does. For $\rho_1 < \bar{\rho} < \rho_0$, rearrangement of the same posterior inequality gives

$$\Lambda(e) \leq \frac{(1-\pi)(\rho_0 - \bar{\rho})}{\pi(\bar{\rho} - \rho_1)}. \tag{45}$$

At $\bar{\rho} = \rho_0$, only an evidence realization with source posterior zero can meet the burden. This exhausts every ordering and equality case. □

B.2 Public pooling

Let H^{pub} be the complete public history generated by displayed policy p. The pooling restriction states that the hidden payoff decomposition does not enter the observable event kernel. For every measurable event A and hidden state ξ,

$$\mathbb{P}(H^{pub} \in A \mid p, M, I, \xi) = Q_p(A). \tag{46}$$

Lemma 2 (Exact public pooling)**.** *Suppose positive masses of both source types select policy p and equation* (46) *holds for every measurable event. Then*

$$\mathbb{P}(H^{pub} \in A \mid p, M = 0) = \mathbb{P}(H^{pub} \in A \mid p, M = 1) = Q_p(A) \tag{47}$$

for every measurable A. The public likelihood ratio for source equals one almost surely within p, and the source posterior equals its within-policy prior.

Proof. Fix $m \in \{0, 1\}$. Apply the law of iterated probabilities to the latent legal state and hidden economic state:

$$\begin{aligned}
\mathbb{P}(H^{pub} \in A \mid p, M = m) &= \int \mathbb{P}(H^{pub} \in A \mid p, M = m, I = i, \xi)\,\mathrm{d}P(i, \xi \mid p, M = m) && (48)\\
&= \int Q_p(A)\,\mathrm{d}P(i, \xi \mid p, M = m) && (49)\\
&= Q_p(A) \int \mathrm{d}P(i, \xi \mid p, M = m) && (50)\\
&= Q_p(A). && (51)
\end{aligned}$$

This proves equation (47). Both conditional measures are the same probability measure. Their Radon-Nikodym derivative is therefore one on the common support. Bayes' rule multiplies the prior odds by one, leaving the within-policy prior unchanged. □

Equality of topological supports would not establish the lemma. The result uses equality of conditional laws within the pooled policy.

B.3 The beneficiary chain and passive linkage

For candidate c, let $D_A(c)$ denote public display in product A, $R_A(c)$ the market response, $X_B(c)$ the locked beneficiary execution in product B, $W_A(c)$ withdrawal in A, and $\Pi(c)$ strategy profit. The maintained causal chain is

$$D_A(c) \longrightarrow R_A(c) \longrightarrow X_B(c) \longrightarrow W_A(c), \qquad X_B(c) \longrightarrow \Pi(c), \qquad W_A(c) \longrightarrow \Pi(c). \tag{52}$$

The opaque routing primitive has a direct payoff interpretation. Suppose display changes responder flow by

$$\delta(m_O) = \delta_0 - \delta_1 m_O, \qquad \delta_0, \delta_1 > 0, \tag{53}$$

and the beneficiary gain is

$$\Delta_B(m_O) = q_B g \beta(\delta_0 - \delta_1 m_O), \tag{54}$$

where beneficiary quantity q_B, per-unit surplus g, and cross-product pass-through β are positive. If visible A-side value excluding this channel is s_A^V and implementation cost is k_O, a type with private cost c earns

$$u_O(c, m_O) = s_A^V + \Delta_B(m_O) - k_O - \tau_O - c \tag{55}$$

$$= \underbrace{s_A^V + q_B g \beta \delta_0 - k_O}_{v_O} - \underbrace{q_B g \beta \delta_1}_{1/w_O} m_O - \tau_O - c. \tag{56}$$

Thus the affine opportunity weight is $w_O = 1/(q_B g \beta \delta_1)$. The routed population is response necessary when its cost support satisfies

$$s_A^V - k_O - c < 0 \leq s_A^V + \Delta_B(m_O) - k_O - c. \tag{57}$$

Within-product and downstream source certificates need not be nested. If s_A^0 is pre-certificate A-side surplus, Δ_A is the within-A response increment, and Δ_B is the downstream increment, define

$$C_A^0 = \mathbf{1}\{s_A^0 < 0 \leq s_A^0 + \Delta_A\}, \tag{58}$$

$$C_{AB} = \mathbf{1}\{s_A^0 + \Delta_A < 0 \leq s_A^0 + \Delta_A + \Delta_B\}. \tag{59}$$

The values $(s_A^0, \Delta_A, \Delta_B) = (-2, 3, 0)$ give $(C_A^0, C_{AB}) = (1, 0)$. The values $(-2, 0, 3)$ give $(0, 1)$. Each certificate can hold without the other.

Lemma 3 (Passive linkage boundary)**.** *There are two structural causal models with the same joint passive law of display D, response R, beneficiary outcome Y, withdrawal W, owner identifiers, and timestamps, but different average causal effects of visibility on Y.*

Proof. Let U be Bernoulli with probability one half. In the causal-chain economy set

$$D = U, \qquad R = D, \qquad Y = R, \qquad W = Y. \tag{60}$$

In the common-shock economy set

$$D = U, \qquad R = U, \qquad Y = U, \qquad W = U. \tag{61}$$

In both economies the passive record equals $(0,0,0,0)$ with probability one half and $(1,1,1,1)$ with probability one half. Add the same constant owner identifier and timestamps to both records. The full passive laws remain equal.

Under the intervention $do(D = d)$, the causal-chain economy has $Y = d$, so

$$\mathbb{E}[Y \mid do(D = 1)] - \mathbb{E}[Y \mid do(D = 0)] = 1. \tag{62}$$

In the common-shock economy, $Y = U$ after intervening on D, so

$$\mathbb{E}[Y \mid do(D = 1)] - \mathbb{E}[Y \mid do(D = 0)] = 0. \tag{63}$$

The same passive law is therefore compatible with distinct causal effects. □

B.4 Visibility identification

Let candidates be partitioned into predetermined clusters g. Assignment $Z_g \in \{0,1\}$ is randomized conditional on preassignment strata W_g within enrolled population $\mathcal{P}$. Let $Y_B(c,z)$ be candidate c's potential beneficiary outcome under cluster exposure z. Assume positivity, consistency, perfect implementation, common matching and priority rules, preassignment feasibility screening, and no interference across clusters. Formally,

$$Z_g \perp \{Y_B(c,0), Y_B(c,1) : c \in g\} \mid W_g, \mathcal{P}. \tag{64}$$

Lemma 4 (Visibility experiment)**.** *Under equation* (64) *and the stated design conditions,*

$$\begin{aligned} &\mathbb{E}[Y_B(c,1) - Y_B(c,0) \mid W_g = w, \mathcal{P}] \\ &\quad = \mathbb{E}[Y_B \mid Z_g = 1, W_g = w, \mathcal{P}] - \mathbb{E}[Y_B \mid Z_g = 0, W_g = w, \mathcal{P}]. \end{aligned} \tag{65}$$

Proof. Conditional randomization implies, for $z \in \{0,1\}$,

$$\mathbb{E}[Y_B(c,z) \mid Z_g = z, W_g = w, \mathcal{P}] = \mathbb{E}[Y_B(c,z) \mid W_g = w, \mathcal{P}]. \tag{66}$$

Consistency and perfect implementation give

$$Y_B = Y_B(c,z) \quad \text{when } Z_g = z. \tag{67}$$

Substituting observed outcomes into the left side of equation (66) for each arm and subtracting the $z = 0$ equality from the $z = 1$ equality yields equation (65). Positivity ensures that both conditional observed means are defined from the enrolled data. □

The estimand is an average total visibility effect for the enrolled stratum. It is not an individual effect, a mediated response effect without further structure, a response-necessity certificate without the surplus comparison, or a legal-intent finding.

Proof of Proposition 1. Lemma 2 proves equality of the two public-history laws and the unit public likelihood ratio. Lemma 3 proves that adding owner identifiers and timestamps does not resolve the visibility counterfactual. Lemma 4 establishes the conditional average visibility effect under the stated cluster design. Equation equation (57) supplies the surplus comparison required to classify the source as response necessary. Lemma 1 proves the separate mapping from source evidence to a legal-intent posterior, including every boundary and reverse-order case. These five results establish each clause of Proposition 1. □

C Information, Certification, and Evidence Architecture

C.1 Likelihood-ratio certification

For mutually absolutely continuous source laws P_1 and P_0, let

$$\Lambda(e) = \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(e), \qquad C_b = \{e : \log \Lambda(e) \geq b\}. \tag{68}$$

Conditional certification rates are

$$\eta = P_1(C_b), \qquad \alpha = P_0(C_b). \tag{69}$$

For screened source masses m^{scr} and ℓ^{scr} and coverage a, expected true and false source certificates are $a\eta m^{scr}$ and $a\alpha\ell^{scr}$. The expected strategic penalty is $Fa\eta$. These quantities keep information, coverage, and sanction magnitude separate.

C.2 The Gaussian experiment

Suppose a raw statistic satisfies

$$X \mid M = 1 \sim N(\mu_1, \sigma^2), \qquad X \mid M = 0 \sim N(\mu_0, \sigma^2), \qquad \mu_1 > \mu_0. \tag{70}$$

Taking the log ratio of the two densities gives

$$L(X) = -\frac{(X-\mu_1)^2}{2\sigma^2} + \frac{(X-\mu_0)^2}{2\sigma^2} \tag{71}$$

$$= \frac{(X^2 - 2\mu_0 X + \mu_0^2) - (X^2 - 2\mu_1 X + \mu_1^2)}{2\sigma^2} \tag{72}$$

$$= \frac{\mu_1 - \mu_0}{\sigma^2}\left(X - \frac{\mu_1 + \mu_0}{2}\right). \tag{73}$$

Define

$$J = \frac{(\mu_1 - \mu_0)^2}{\sigma^2} > 0. \tag{74}$$

Under $M = 1$, the mean of equation (73) is

$$\frac{\mu_1 - \mu_0}{\sigma^2}\left(\mu_1 - \frac{\mu_1 + \mu_0}{2}\right) = \frac{(\mu_1 - \mu_0)^2}{2\sigma^2} = \frac{J}{2}. \tag{75}$$

Under $M = 0$, the same calculation gives $-J/2$. In either state, its variance is

$$\left(\frac{\mu_1 - \mu_0}{\sigma^2}\right)^2 \sigma^2 = J. \tag{76}$$

Thus

$$L \mid M = 1 \sim N(J/2, J), \qquad L \mid M = 0 \sim N(-J/2, J). \tag{77}$$

Lemma 5 (Gaussian rates, derivatives, and limits)**.** *At fixed hurdle $b > 0$,*

$$\eta(J, b) = \Phi\left(\frac{\sqrt{J}}{2} - \frac{b}{\sqrt{J}}\right), \tag{78}$$

$$\alpha(J, b) = \Phi\left(-\frac{\sqrt{J}}{2} - \frac{b}{\sqrt{J}}\right). \tag{79}$$

Their derivatives are

$$\eta_J = \phi\left(\frac{\sqrt{J}}{2} - \frac{b}{\sqrt{J}}\right)\frac{J + 2b}{4J^{3/2}} > 0, \tag{80}$$

$$\alpha_J = \phi\left(-\frac{\sqrt{J}}{2} - \frac{b}{\sqrt{J}}\right)\frac{2b - J}{4J^{3/2}}. \tag{81}$$

The false-source rate rises on $0 < J < 2b$, peaks at $J = 2b$, and falls for $J > 2b$. Moreover,

$$\lim_{J \downarrow 0}(\eta, \alpha) = (0, 0), \qquad \lim_{J \to \infty}(\eta, \alpha) = (1, 0). \tag{82}$$

Proof. Under $M = 1$, standardization of equation (77) gives

$$\eta = \mathbb{P}(L \geq b \mid M = 1) \tag{83}$$
$$= 1 - \Phi\left(\frac{b - J/2}{\sqrt{J}}\right) \tag{84}$$
$$= \Phi\left(\frac{J/2 - b}{\sqrt{J}}\right) = \Phi\left(\frac{\sqrt{J}}{2} - \frac{b}{\sqrt{J}}\right). \tag{85}$$

Under $M = 0$,

$$\alpha = 1 - \Phi\left(\frac{b + J/2}{\sqrt{J}}\right) \tag{86}$$
$$= \Phi\left(-\frac{\sqrt{J}}{2} - \frac{b}{\sqrt{J}}\right). \tag{87}$$

Define

$$g(J) = \frac{1}{2}J^{1/2} - bJ^{-1/2}, \qquad r(J) = -\frac{1}{2}J^{1/2} - bJ^{-1/2}. \tag{88}$$

Term-by-term differentiation gives

$$g'(J) = \frac{1}{4}J^{-1/2} + \frac{b}{2}J^{-3/2} = \frac{J + 2b}{4J^{3/2}}, \tag{89}$$
$$r'(J) = -\frac{1}{4}J^{-1/2} + \frac{b}{2}J^{-3/2} = \frac{2b - J}{4J^{3/2}}. \tag{90}$$

The chain rule yields equation (81). Since the standard normal density is positive, the signs follow from the displayed numerators. As $J \downarrow 0$, both $g(J)$ and $r(J)$ tend to negative infinity. As $J \to \infty$, $g(J)$ tends to positive infinity and $r(J)$ to negative infinity. Applying the limits of Φ proves the endpoint claims. $\square$

Lemma 6 (Information improves the true-to-false ratio)**.** *For $b > 0$, $\eta(J, b)/\alpha(J, b)$ is strictly increasing in $J > 0$.*

Proof. For $J \geq 2b$, Lemma 5 gives $\eta_J > 0$ and $\alpha_J \leq 0$. Since both rates are positive at finite J, their ratio rises strictly.

Now take $0 < J < 2b$ and put $s = \sqrt{J}$. Define

$$x = \frac{b}{s} - \frac{s}{2} > 0, \qquad y = \frac{b}{s} + \frac{s}{2} > x. \tag{91}$$

Then $\eta = \bar{\Phi}(x)$ and $\alpha = \bar{\Phi}(y)$, where $\bar{\Phi} = 1 - \Phi$. Differentiation gives

$$\frac{\mathrm{d}x}{\mathrm{d}s} = -\frac{y}{s}, \qquad \frac{\mathrm{d}y}{\mathrm{d}s} = -\frac{x}{s}. \tag{92}$$

Let $R(u) = \phi(u)/\bar{\Phi}(u)$ be the inverse Mills ratio. Using equation (92),

$$\frac{\mathrm{d}}{\mathrm{d}s} \log \frac{\eta}{\alpha} = -R(x)\frac{\mathrm{d}x}{\mathrm{d}s} + R(y)\frac{\mathrm{d}y}{\mathrm{d}s} \tag{93}$$

$$= \frac{yR(x) - xR(y)}{s}. \tag{94}$$

It remains to show $yR(x) > xR(y)$.

For $u > 0$, integration by parts gives

$$\bar{\Phi}(u) = \int_u^\infty \phi(t)\,\mathrm{d}t = \int_u^\infty \frac{1}{t}\{t\phi(t)\}\,\mathrm{d}t \tag{95}$$

$$= \frac{\phi(u)}{u} - \int_u^\infty \frac{\phi(t)}{t^2}\,\mathrm{d}t. \tag{96}$$

Since $t > u$ inside the final integral except at its lower endpoint,

$$\int_u^\infty \frac{\phi(t)}{t^2}\,\mathrm{d}t < \frac{\bar{\Phi}(u)}{u^2}. \tag{97}$$

Substitution into equation (96) yields

$$\bar{\Phi}(u) > \frac{u}{1+u^2}\phi(u), \qquad R(u) < u + \frac{1}{u}. \tag{98}$$

Direct differentiation of R gives $R'(u) = R(u)\{R(u) - u\}$. Therefore

$$\frac{\mathrm{d}}{\mathrm{d}u}\frac{R(u)}{u} = \frac{uR'(u) - R(u)}{u^2} \tag{99}$$

$$= \frac{R(u)}{u^2}\left[u\{R(u) - u\} - 1\right] < 0, \tag{100}$$

where the strict inequality uses equation (98). Since $0 < x < y$,

$$\frac{R(x)}{x} > \frac{R(y)}{y}, \tag{101}$$

which is equivalent to $yR(x) > xR(y)$. Equation equation (94) is therefore positive. The ratio rises with s and hence with $J = s^2$ on $0 < J < 2b$. Combining the two regions completes the proof. □

At fixed strategic penalty $t = Fa\eta$, feasible coverage is $a = t/(F\eta)$. A bona fide provider's private burden is

$$p_x(J;t) = f_x a\alpha = \frac{f_x t}{F}\frac{\alpha(J,b)}{\eta(J,b)}. \tag{102}$$

Lemma 6 makes this burden strictly decreasing in information. The positive-part supply and value functions in equation (29) then weakly rise, with strict increase before their caps bind.

C.3 Candidate-average complementarity

Suppose n independent records each contain $N(P)$ candidate scores that, conditional on source state and architecture, are independent Gaussian variables with common state-invariant variance $\sigma^2(R) > 0$. Exactly one candidate has source-induced mean shift $\theta \neq 0$. The retained candidate average is therefore Gaussian, with mean shift $\theta/N(P)$. Within-record independence gives variance

$$\mathrm{Var}\left(\frac{1}{N(P)}\sum_{j=1}^{N(P)} Y_j\right) = \frac{1}{N(P)^2}\sum_{j=1}^{N(P)} \mathrm{Var}(Y_j) = \frac{\sigma^2(R)}{N(P)}. \tag{103}$$

Per-record information is therefore

$$\frac{\{\theta/N(P)\}^2}{\sigma^2(R)/N(P)} = \frac{\theta^2}{N(P)\sigma^2(R)}. \tag{104}$$

The common Gaussian variance makes the squared mean separation divided by variance the exact per-record likelihood information used in the certification experiment. Independence makes information additive across records:

$$J(P,R) = \frac{n\theta^2}{N(P)\sigma^2(R)}. \tag{105}$$

Lemma 7 (Architecture complementarity)**.** *If $N(1) < N(0)$ and $\sigma^2(1) < \sigma^2(0)$, provenance and reconstruction have a strictly positive information cross-difference.*

Proof. Substitution of the four architecture values gives

$$\Delta_{PR} J = n\theta^2 \left\{\frac{1}{N(1)\sigma^2(1)} - \frac{1}{N(1)\sigma^2(0)} - \frac{1}{N(0)\sigma^2(1)} + \frac{1}{N(0)\sigma^2(0)}\right\} \tag{106}$$

$$= n\theta^2 \left\{\frac{1}{N(1)} - \frac{1}{N(0)}\right\}\left\{\frac{1}{\sigma^2(1)} - \frac{1}{\sigma^2(0)}\right\}. \tag{107}$$

Both bracketed differences are positive, and $n\theta^2 > 0$. Their product is strictly positive. $\square$

C.4 Correlated evidence channels

Let $Y \mid M = 0 \sim N(0, \Sigma)$ and $Y \mid M = 1 \sim N(\delta, \Sigma)$, where both channels have unit variance and correlation $\varrho \in (-1, 1)$. Then

$$\Sigma^{-1} = \frac{1}{1-\varrho^2} \begin{pmatrix} 1 & -\varrho \\ -\varrho & 1 \end{pmatrix}. \tag{108}$$

Writing $\delta = (\delta_P, \delta_R)'$, joint information is

$$J_{PR} = \delta' \Sigma^{-1} \delta \tag{109}$$

$$= \frac{\delta_P^2 + \delta_R^2 - 2\varrho\delta_P\delta_R}{1-\varrho^2}. \tag{110}$$

Singleton information is $J_P = \delta_P^2$ and $J_R = \delta_R^2$. Hence

$$J_{PR} - J_P - J_R = \frac{\varrho^2(\delta_P^2 + \delta_R^2) - 2\varrho\delta_P\delta_R}{1-\varrho^2}, \tag{111}$$

whose sign can be positive or negative. The incremental information from reconstruction conditional on provenance is

$$J_{PR} - J_P = \frac{\delta_P^2 + \delta_R^2 - 2\varrho\delta_P\delta_R - (1-\varrho^2)\delta_P^2}{1-\varrho^2} \tag{112}$$

$$= \frac{(\delta_R - \varrho\delta_P)^2}{1-\varrho^2}. \tag{113}$$

It is zero exactly when $\delta_R = \varrho\delta_P$. Thus a second evidence leg can be complementary, substitutive in cross-difference, or conditionally redundant.

Proof of Proposition 3. Lemma 5 proves the certification formulas, strict increase of the true-source rate, the false-source hump, and all endpoint limits. Lemma 6 proves strict increase of η/α for every positive information level. Substitution into the fixed-deterrence burden proves the legitimate-liquidity conclusion. Lemma 7 proves the positive candidate-average information cross-difference. Equation equation (111) supplies the stated qualification for correlated evidence channels. These results prove Proposition 3. □

D Global Routing and Displacement

D.1 The four active sets

For an active set $S \subseteq \{\mathrm{T}, \mathrm{O}\}$, define

$$D_S = \lambda + \sum_{x \in S} w_x. \tag{114}$$

Theorem 1 (Unique global routing equilibrium)**.** *The system in equation* (24) *has a unique solution. It belongs to exactly one of the following regions, with equality assigned to the adjacent single-active region.*

No entry. *If $A_\mathrm{T} \leq 0$ and $A_\mathrm{O} \leq 0$, then*

$$z = m_\mathrm{T} = m_\mathrm{O} = 0. \tag{115}$$

Transparent only. *If*

$$A_\mathrm{T} > 0, \qquad A_\mathrm{O} \leq \frac{w_\mathrm{T} A_\mathrm{T}}{\lambda + w_\mathrm{T}}, \tag{116}$$

then

$$z = \frac{w_\mathrm{T} A_\mathrm{T}}{\lambda + w_\mathrm{T}}, \qquad m_\mathrm{T} = \frac{\lambda w_\mathrm{T} A_\mathrm{T}}{\lambda + w_\mathrm{T}}, \qquad m_\mathrm{O} = 0. \tag{117}$$

Opaque only. *If*

$$A_\mathrm{O} > 0, \qquad A_\mathrm{T} \leq \frac{w_\mathrm{O} A_\mathrm{O}}{\lambda + w_\mathrm{O}}, \tag{118}$$

then

$$z = \frac{w_\mathrm{O} A_\mathrm{O}}{\lambda + w_\mathrm{O}}, \qquad m_\mathrm{O} = \frac{\lambda w_\mathrm{O} A_\mathrm{O}}{\lambda + w_\mathrm{O}}, \qquad m_\mathrm{T} = 0. \tag{119}$$

Both active. *Let $D = \lambda + w_\mathrm{T} + w_\mathrm{O}$. If*

$$(\lambda + w_\mathrm{O}) A_\mathrm{T} > w_\mathrm{O} A_\mathrm{O}, \qquad (\lambda + w_\mathrm{T}) A_\mathrm{O} > w_\mathrm{T} A_\mathrm{T}, \tag{120}$$

then

$$z = \frac{w_\mathrm{T} A_\mathrm{T} + w_\mathrm{O} A_\mathrm{O}}{D}, \tag{121}$$

$$m_\mathrm{T} = \frac{w_\mathrm{T}\{(\lambda + w_\mathrm{O}) A_\mathrm{T} - w_\mathrm{O} A_\mathrm{O}\}}{D}, \tag{122}$$

$$m_\mathrm{O} = \frac{w_\mathrm{O}\{(\lambda + w_\mathrm{T}) A_\mathrm{O} - w_\mathrm{T} A_\mathrm{T}\}}{D}. \tag{123}$$

Proof. Move the right side of the market-clearing equation to the left and define

$$F(z) = w_{\mathrm{T}}\,[A_{\mathrm{T}} - z]_+ + w_{\mathrm{O}}\,[A_{\mathrm{O}} - z]_+ - \lambda z. \tag{124}$$

The function is continuous. On every interval that does not contain A_{T} or A_{O}, its slope is

$$F'(z) = -\lambda - \sum_{x:A_x>z} w_x < 0. \tag{125}$$

Continuity and the strictly negative segment slopes make F strictly decreasing on $[0,\infty)$. Also, $F(z) \to -\infty$ as $z \to \infty$.

If both net attractiveness values are nonpositive, $F(0) = 0$ and strict decrease gives the unique nonnegative root $z = 0$. The positive parts then give equation (115). If at least one A_x is positive, $F(0) > 0$. The intermediate value theorem and strict decrease give exactly one positive root.

Fix a candidate active set S. Removing the positive parts for active tactics and setting inactive masses to zero gives

$$\sum_{x\in S} w_x(A_x - z) = \lambda z. \tag{126}$$

Collecting the terms in z yields

$$z_S = \frac{\sum_{x\in S} w_x A_x}{D_S}. \tag{127}$$

Validity requires $A_x > z_S$ for each active x and $A_y \le z_S$ for each inactive y.

For $S = \{\mathrm{T}\}$, equation (127) gives $z = w_{\mathrm{T}}A_{\mathrm{T}}/(\lambda + w_{\mathrm{T}})$. The active inequality $A_{\mathrm{T}} > z$ is equivalent to $A_{\mathrm{T}} > 0$. The inactive inequality $A_{\mathrm{O}} \le z$ gives equation (116). Substituting z into $m_{\mathrm{T}} = w_{\mathrm{T}}(A_{\mathrm{T}} - z)$ gives

$$m_{\mathrm{T}} = w_{\mathrm{T}}A_{\mathrm{T}}\left(1 - \frac{w_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}}\right) = \frac{\lambda w_{\mathrm{T}}A_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}}. \tag{128}$$

This proves equation (117). Exchanging tactic labels gives equations (118) and (119).

For $S = \{\mathrm{T},\mathrm{O}\}$, equation (127) gives the first line of equation (123). The transparent active inequality is

$$A_{\mathrm{T}} > \frac{w_{\mathrm{T}}A_{\mathrm{T}} + w_{\mathrm{O}}A_{\mathrm{O}}}{D} \tag{129}$$

$$DA_{\mathrm{T}} > w_{\mathrm{T}}A_{\mathrm{T}} + w_{\mathrm{O}}A_{\mathrm{O}} \tag{130}$$

$$(\lambda + w_{\mathrm{O}})A_{\mathrm{T}} > w_{\mathrm{O}}A_{\mathrm{O}}. \tag{131}$$

The opaque inequality gives the second condition in equation (120). Finally,

$$m_{\mathrm{T}} = w_{\mathrm{T}} \left(A_{\mathrm{T}} - \frac{w_{\mathrm{T}} A_{\mathrm{T}} + w_{\mathrm{O}} A_{\mathrm{O}}}{D} \right) \tag{132}$$

$$= \frac{w_{\mathrm{T}} \{ (\lambda + w_{\mathrm{O}}) A_{\mathrm{T}} - w_{\mathrm{O}} A_{\mathrm{O}} \}}{D}, \tag{133}$$

and the opaque expression follows by exchanging labels. The unique root established above makes these regions exhaustive and rules out two distinct valid solutions. □

D.2 Boundary continuity

Lemma 8 (Continuous boundaries)**.** *The functions $z, m_{\mathrm{T}}, m_{\mathrm{O}}$ are continuous and piecewise affine. On the transparent-only/coexistence boundary, where $A_{\mathrm{T}} > 0$ and*

$$A_{\mathrm{O}} = \frac{w_{\mathrm{T}} A_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}}, \tag{134}$$

the both-active solution equals the transparent-only solution and $m_{\mathrm{O}} = 0$. The corresponding opaque-only/coexistence statement holds where $A_{\mathrm{O}} > 0$ and

$$A_{\mathrm{T}} = \frac{w_{\mathrm{O}} A_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}}. \tag{135}$$

At $(A_{\mathrm{T}}, A_{\mathrm{O}}) = (0, 0)$, both single-active masses converge to the no-entry solution. Algebraic extensions of the two displayed lines into negative net attractiveness are not active-set boundaries.

Proof. Substitute equation (134) into the both-active cutoff:

$$z_B = \frac{w_{\mathrm{T}} A_{\mathrm{T}} + w_{\mathrm{O}} w_{\mathrm{T}} A_{\mathrm{T}} / (\lambda + w_{\mathrm{T}})}{D} \tag{136}$$

$$= \frac{w_{\mathrm{T}} A_{\mathrm{T}} \{ \lambda + w_{\mathrm{T}} + w_{\mathrm{O}} \}}{(\lambda + w_{\mathrm{T}}) D} \tag{137}$$

$$= \frac{w_{\mathrm{T}} A_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}} = z_T. \tag{138}$$

The opaque numerator in equation (123) becomes

$$(\lambda + w_{\mathrm{T}}) \frac{w_{\mathrm{T}} A_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}} - w_{\mathrm{T}} A_{\mathrm{T}} = 0, \tag{139}$$

so $m_{\mathrm{O}} = 0$. Since total mass equals λz, the remaining mass is $m_{\mathrm{T}} = \lambda z_T$, which is the transparent-only formula. Exchanging labels proves continuity at equation (135). At $A_x = 0$ in a single-active region, the active mass is proportional to A_x and converges to zero. Every formula is affine within its region, completing the proof. □

D.3 Displacement in the coexistence region

Lemma 9 (Displacement matrix, finite kill, and harm)**.** *Inside the strict both-active region,*

$$\frac{\partial m_{\mathrm{T}}}{\partial \tau_{\mathrm{T}}} = -\frac{w_{\mathrm{T}}(\lambda + w_{\mathrm{O}})}{D}, \qquad \frac{\partial m_{\mathrm{O}}}{\partial \tau_{\mathrm{T}}} = \frac{w_{\mathrm{T}} w_{\mathrm{O}}}{D}, \tag{140}$$

$$\frac{\partial m_{\mathrm{O}}}{\partial \tau_{\mathrm{O}}} = -\frac{w_{\mathrm{O}}(\lambda + w_{\mathrm{T}})}{D}, \qquad \frac{\partial m_{\mathrm{T}}}{\partial \tau_{\mathrm{O}}} = \frac{w_{\mathrm{T}} w_{\mathrm{O}}}{D}. \tag{141}$$

For target j and refuge $k \neq j$, the migration and exit shares of target suppression are $w_k/(\lambda + w_k)$ and $\lambda/(\lambda + w_k)$. The finite penalty required to remove target mass $m_j > 0$ is

$$\Delta\tau_j^{kill} = \frac{m_j D}{w_j(\lambda + w_k)}. \tag{142}$$

At that boundary,

$$\Delta m_k = \frac{w_k}{\lambda + w_k} m_j, \qquad -\Delta\mathcal{M} = \frac{\lambda}{\lambda + w_k} m_j. \tag{143}$$

Finally,

$$\frac{\partial H}{\partial \tau_j} = -G_j, \qquad G_j = \frac{w_j\{h_j(\lambda + w_k) - h_k w_k\}}{D}. \tag{144}$$

Proof. Since $A_x = v_x - \tau_x$, $\partial A_x/\partial \tau_x = -1$ and the other tactic's attractiveness is unchanged. Differentiate each numerator in equation (123). For example,

$$\frac{\partial m_{\mathrm{T}}}{\partial \tau_{\mathrm{T}}} = \frac{w_{\mathrm{T}}(\lambda + w_{\mathrm{O}})}{D} \frac{\partial A_{\mathrm{T}}}{\partial \tau_{\mathrm{T}}} = -\frac{w_{\mathrm{T}}(\lambda + w_{\mathrm{O}})}{D}, \tag{145}$$

$$\frac{\partial m_{\mathrm{O}}}{\partial \tau_{\mathrm{T}}} = -\frac{w_{\mathrm{O}} w_{\mathrm{T}}}{D} \frac{\partial A_{\mathrm{T}}}{\partial \tau_{\mathrm{T}}} = \frac{w_{\mathrm{T}} w_{\mathrm{O}}}{D}. \tag{146}$$

The other two derivatives follow by exchanging labels, proving equation (141).

The magnitude of target suppression per unit penalty is $w_j(\lambda + w_k)/D$. Refuge expansion per unit penalty is $w_j w_k/D$. Dividing the second coefficient by the first gives

$$\frac{w_j w_k/D}{w_j(\lambda + w_k)/D} = \frac{w_k}{\lambda + w_k}. \tag{147}$$

Adding the two mass derivatives gives

$$\frac{\partial \mathcal{M}}{\partial \tau_j} = -\frac{w_j(\lambda + w_k)}{D} + \frac{w_j w_k}{D} = -\frac{\lambda w_j}{D}. \tag{148}$$

Dividing the magnitude of this total decline by target suppression gives $\lambda/(\lambda+w_k)$. The two shares sum to one.

Within the affine region, target mass after a penalty increase $\Delta\tau_j$ is

$$m_j(\Delta\tau_j) = m_j - \frac{w_j(\lambda + w_k)}{D}\Delta\tau_j. \tag{149}$$

Setting this expression to zero and solving gives equation (142). Multiplying the penalty change by the refuge and total-mass derivatives gives equation (143). Lemma 8 verifies that these endpoint masses equal the adjacent single-active solution.

For harm,

$$\frac{\partial H}{\partial \tau_j} = h_j \left[-\frac{w_j(\lambda + w_k)}{D}\right] + h_k \left[\frac{w_j w_k}{D}\right] \tag{150}$$

$$= -\frac{w_j\{h_j(\lambda + w_k) - h_k w_k\}}{D} = -G_j. \tag{151}$$

Thus harm rises exactly when $G_j < 0$. □

D.4 Single-active regions

In the transparent-only region, differentiation of equation (117) gives

$$\frac{\partial m_{\mathrm{T}}}{\partial \tau_{\mathrm{T}}} = -\frac{\lambda w_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}}, \qquad \frac{\partial m_{\mathrm{O}}}{\partial \tau_{\mathrm{T}}} = 0. \tag{152}$$

An opaque penalty has no local effect while opacity remains inactive. If the opaque tactic stays inactive, transparent activity reaches zero when the remaining net attractiveness A_{T} is removed. In the opaque-only region,

$$\frac{\partial m_{\mathrm{O}}}{\partial \tau_{\mathrm{O}}} = -\frac{\lambda w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}}, \qquad \frac{\partial m_{\mathrm{T}}}{\partial \tau_{\mathrm{O}}} = 0, \tag{153}$$

with the symmetric kill condition. At a refuge-entry boundary, the relevant derivative switches to equation (141).

Proof of Proposition 2. Existence, uniqueness, and the four active sets follow from Theorem 1. Lemma 8 proves continuity and piecewise affinity. Lemma 9 proves all coexistence derivatives, migration and exit shares, finite boundary behavior, and the harm condition. Equations equations (152) and (153) establish the single-active claims. These cases exhaust the routing system. □

E The Regulator's Coverage Problem

E.1 Existence and regional first-order conditions

For active routing set S, define the displacement-adjusted harm gradient for active target $x \in S$ by

$$G_x^S = w_x \left\{ h_x - \frac{\sum_{y \in S} h_y w_y}{D_S} \right\}. \tag{154}$$

Set $G_x^S = 0$ locally when $x \notin S$.

Theorem 2 (Coverage existence and regional KKT system)**.** *A global minimizer of equation* (33) *exists over the finite architecture set and coverage square. Inside a fixed routing set S with $f_x a_x \alpha < s_x$, the coverage stationarity condition is*

$$\begin{aligned} 0 = & -F\eta G_x^S + c_x + \kappa_x a_x + \nu_x f_x \alpha (s_x + q_x^L - f_x \alpha a_x) \\ & + \chi_x \nu_x \alpha (s_x - 2 f_x \alpha a_x) + \eta \sum_y \chi_{10,y} \left\{ \mathbf{1}\{y = x\} n_{10,y}(a) + a_y \frac{\partial n_{10,y}(a)}{\partial a_x} \right\} - \mu_x^0 + \mu_x^1, \end{aligned} \tag{155}$$

with

$$\mu_x^0 a_x = 0, \qquad \mu_x^1 (1 - a_x) = 0, \qquad \mu_x^0, \mu_x^1 \geq 0. \tag{156}$$

At routing and legitimate-exit boundaries, the stationarity equation is replaced by the directional conditions derived below.

Proof. Lemma 8 makes the routing masses continuous in penalties and hence in coverage. The positive-part functions in equation (29) are continuous. Under the fixed intent-composition relation $n_{10,x}(a) = \omega_{10,x} m_x(a)$, the non-intender mass is also continuous. Every term in equation (33) is therefore continuous on $[0, 1]^2$ for each architecture. The coverage square is compact, so the Weierstrass theorem gives a minimizer for each architecture. The architecture set is finite, so at least one of those minimizers has the smallest global value.

Inside a routing region, expected penalty changes with coverage at rate $F\eta$. By definition of the harm gradient,

$$\frac{\partial H}{\partial a_x} = \frac{\partial H}{\partial \tau_x} \frac{\partial \tau_x}{\partial a_x} = (-G_x^S)(F\eta). \tag{157}$$

For $p_x = f_x a_x \alpha < s_x$, legitimate value is

$$\mathcal{V}_x(p_x) = \nu_x \left\{ \frac{(s_x - f_x \alpha a_x)^2}{2} + q_x^L (s_x - f_x \alpha a_x) \right\}. \tag{158}$$

Differentiate chilling loss $\mathcal{V}_x(0) - \mathcal{V}_x(p_x)$:

$$\frac{\partial C_x^{chill}}{\partial a_x} = -\nu_x \left\{(s_x - f_x\alpha a_x)(-f_x\alpha) + q_x^L(-f_x\alpha)\right\} \tag{159}$$

$$= \nu_x f_x\alpha(s_x + q_x^L - f_x\alpha a_x). \tag{160}$$

False-source adjudication loss is

$$D_x^{FP} = \chi_x \nu_x \alpha a_x(s_x - f_x\alpha a_x). \tag{161}$$

Its derivative is

$$\frac{\partial D_x^{FP}}{\partial a_x} = \chi_x \nu_x \alpha(s_x - 2f_x\alpha a_x). \tag{162}$$

For each tactic y, the product rule gives

$$\frac{\partial D_y^{10}}{\partial a_x} = \chi_{10,y}\eta \frac{\partial}{\partial a_x}\{a_y n_{10,y}(a)\} \tag{163}$$

$$= \chi_{10,y}\eta \left\{\mathbf{1}\{y = x\} n_{10,y}(a) + a_y \frac{\partial n_{10,y}(a)}{\partial a_x}\right\}. \tag{164}$$

Summing equation (164) over y captures both the target cell and any covered refuge whose non-intender mass changes through displacement. Adding this sum to equations (157), (160) and (162), the derivative of direct audit cost, and the lower- and upper-bound multipliers yields equation (155). The complementarity equations in equation (156) are the KKT conditions for $a_x \geq 0$ and $a_x \leq 1$. At a kink, a two-sided derivative need not exist; optimality requires a nonpositive directional derivative toward the candidate from one side and a nonnegative derivative away from it on the other side.

To verify equation (154), use the active-set cutoff

$$z_S = \frac{\sum_{y\in S} w_y A_y}{D_S}. \tag{165}$$

For $x \in S$ and $y \in S$,

$$\frac{\partial m_y}{\partial \tau_x} = w_y \left\{-\mathbf{1}\{x = y\} + \frac{w_x}{D_S}\right\}. \tag{166}$$

Weighting by h_y and summing gives

$$\frac{\partial H}{\partial \tau_x} = -h_x w_x + \frac{w_x}{D_S}\sum_{y\in S} h_y w_y = -G_x^S. \tag{167}$$

This completes the regional derivation. □

E.2 Boundary conditions and the global candidate comparison

Let $\mathcal{R}$ index the finitely many cells generated by the four routing regions, the legitimate-participation inequalities $f_x \alpha a_x < s_x$, and the coverage bounds. For a cell $R \in \mathcal{R}$, let $g^R(a)$ be the gradient of the regulator objective on its interior. Its x component is the left side of equation (155) before adding bound multipliers, with the routing set and legitimate-participation state associated with R.

The objective is continuous across every cell boundary. Routing masses are continuous by Lemma 8. At a legitimate-exit point $a_x^L = s_x/(f_x \alpha)$,

$$\ell_x(a_x^L, J) = 0, \qquad C_x^{chill}(a_x^L, J) = \mathcal{V}_x(0), \qquad D_x^{FP}(a_x^L, J) = 0, \tag{168}$$

which equal the post-exit values. The non-intender term is continuous because routing mass is continuous.

For a feasible direction d from boundary point a into adjoining cell R, the one-sided directional derivative for the minimization problem is

$$\mathcal{L}'(a; d) = g^R(a)' d. \tag{169}$$

A boundary point is locally optimal only if

$$g^R(a)' d \geq 0 \tag{170}$$

for every feasible direction d and every adjoining cell entered by that direction. Conversely, if the objective is convex on each adjoining cell and equation (170) holds for every such direction, the boundary point is a local minimum over their union.

For a one-dimensional crossing in a_x from cell R^- to cell R^+ as coverage rises, equation (170) becomes

$$g_x^{R^-}(a) \leq 0 \leq g_x^{R^+}(a). \tag{171}$$

At $a_x = 0$, it becomes $g_x^{R^+}(a) \geq 0$; at $a_x = 1$, it becomes $g_x^{R^-}(a) \leq 0$. At legitimate exit, the derivative on the pre-exit side is equation (155) without multipliers. On the post-exit side, the two bona fide terms vanish, giving

$$g_{x,post}^S(a) = -F\eta G_x^S + c_x + \kappa_x a_x + \eta \sum_y \chi_{10,y} \left\{ \mathbf{1}\{y = x\} n_{10,y}(a) + a_y \frac{\partial n_{10,y}(a)}{\partial a_x} \right\}. \tag{172}$$

At a routing boundary, use the same formula on each side with $G_x^{S^-}$, $G_x^{S^+}$, and the corresponding one-sided routing derivatives of $n_{10,y}$.

The global solution can be found without assuming differentiability at a boundary. On each closed cell, evaluate every feasible interior stationary point, every stationary point on each open face, and every vertex. Apply the same procedure recursively to intersections of faces. Because there are finitely many routing and legitimate-participation cells and each regional objective is quadratic under fixed intent composition, this procedure yields a finite candidate set unless a quadratic is flat on a face. In the flat case, every point on that face has the same value and either endpoint represents it in the global comparison. Continuity and compactness ensure that the smallest candidate value is the global minimum. This is the boundary comparison used by the closed-form one-instrument results below.

A sufficient condition for a unique global optimum is that every feasible regional cell has a positive-definite Hessian, every regional constrained minimizer is unique, and one regional minimizer has a strictly lower objective value than all others. Without the final strict comparison, multiple global optima can lie on an architecture or active-set boundary.

E.3 Closed-form coverage within a fixed cell

Fix architecture J, routing set S, and one coverage instrument a_x on a maximal interval

$$\mathcal{I}_x^S = [\underline{a}_x^S, \overline{a}_x^S] \subseteq [0,1] \tag{173}$$

that preserves both the routing set and interior legitimate participation. Hold the screened response-necessary non-intender mass at $\bar{n}_{10,x}$. Define

$$\bar{c}_x = c_x + \chi_{10,x}\eta\bar{n}_{10,x}, \tag{174}$$

$$\widetilde{B}_x^S = F\eta G_x^S - \bar{c}_x - \nu_x\alpha\{f_x(s_x + q_x^L) + \chi_x s_x\}, \tag{175}$$

$$\widetilde{\kappa}_x = \kappa_x - \nu_x\alpha^2(f_x^2 + 2\chi_x f_x). \tag{176}$$

Lemma 10 (Interior-cell coverage)**.** *Up to a cell-specific constant, regulator gain is*

$$\Delta W_x(a_x) = \widetilde{B}_x^S a_x - \frac{\widetilde{\kappa}_x}{2}a_x^2. \tag{177}$$

If $\widetilde{\kappa}_x > 0$, the unique cell optimum is

$$a_{x,S}^* = \Pi_{\mathcal{I}_x^S}\left(\frac{\widetilde{B}_x^S}{\widetilde{\kappa}_x}\right), \tag{178}$$

where Π denotes Euclidean projection onto the complete interval.

Proof. For $p = f\alpha a < s$, expand legitimate value around zero burden:

$$C^{chill} = \nu \left\{ \frac{s^2 - (s - f\alpha a)^2}{2} + q^L f\alpha a \right\} \tag{179}$$

$$= \nu \left\{ \frac{2sf\alpha a - f^2\alpha^2 a^2}{2} + q^L f\alpha a \right\} \tag{180}$$

$$= \nu f\alpha(s + q^L)a - \frac{\nu f^2\alpha^2}{2}a^2. \tag{181}$$

False-source adjudication loss expands as

$$D^{FP} = \chi\nu\alpha a(s - f\alpha a) \tag{182}$$

$$= \chi\nu\alpha s a - \chi\nu f\alpha^2 a^2. \tag{183}$$

Harm falls linearly at rate $F\eta G_x^S$ within the routing cell. Subtract direct audit cost, equations (181) and (183), and $\chi_{10}\eta\bar{n}_{10}a$. The coefficient on a is $\widetilde{B}_x^S$. The coefficient on $-a^2/2$ is

$$\kappa_x - \nu_x f_x^2\alpha^2 - 2\chi_x\nu_x f_x\alpha^2 = \widetilde{\kappa}_x. \tag{184}$$

This proves equation (177). When $\widetilde{\kappa}_x > 0$, the gain is strictly concave. Its unconstrained derivative is

$$\frac{\partial \Delta W_x}{\partial a_x} = \widetilde{B}_x^S - \widetilde{\kappa}_x a_x, \tag{185}$$

which vanishes at $\widetilde{B}_x^S/\widetilde{\kappa}_x$. Projection onto the closed feasible interval gives equation (178). Strict concavity makes the projected maximizer unique. □

If the interval contains zero, its upper endpoint is the smallest of the unit cap, the next routing boundary, and the legitimate-exit boundary $s_x/(f_x\alpha)$. Only in that cell can equation (178) be written as a capped positive part. If $\widetilde{\kappa}_x = 0$, gain is linear and an endpoint is optimal. If $\widetilde{\kappa}_x < 0$, gain is strictly convex and its maximum over the closed interval is also at an endpoint.

At $a_x^L = s_x/(f_x\alpha)$, legitimate mass reaches zero. Beyond that point, chilling loss is the constant $\mathcal{V}_x(0)$ and false-source loss among active bona fide providers is zero. On a post-exit routing cell, gain has stationary candidate

$$a_{x,S}^{post} = \Pi_{\mathcal{I}_{x,S}^{post}} \left(\frac{F\eta G_x^S - \bar{c}_x}{\kappa_x} \right), \tag{186}$$

where the interval begins at a_x^L and ends at the next routing or unit boundary. The complete one-instrument solution compares the pre-exit candidate, a_x^L, each post-exit candidate, all routing boundaries, and the unit endpoints. This comparison allows optimal coverage to continue beyond complete legitimate exit when manipulation harm is large.

E.4 Fixed-pool audit ranking

For the closed-form adoption benchmark, hold screened legitimate and non-intender masses fixed. Define

$$B_x^S = F\eta G_x^S - c_x - \chi_x \alpha \bar{\ell}_x - \chi_{10,x}\eta \bar{n}_{10,x}. \tag{187}$$

Lemma 11 (Fixed-pool coverage and audit ranking)**.** *On a routing-cell interval $\mathcal{I}_x^S = [\underline{a}_x^S, \overline{a}_x^S]$, optimal coverage is*

$$a_{x,S}^* = \Pi_{\mathcal{I}_x^S}\left(\frac{B_x^S}{\kappa_x}\right). \tag{188}$$

In a zero-containing cell whose upper cap does not bind at the optimum, operating value is

$$V_x^S = \frac{\left[B_x^S\right]_+^2}{2\kappa_x}. \tag{189}$$

When both tactic-specific problems satisfy those conditions, opaque auditing has greater value exactly when

$$\frac{\left[B_{\mathrm{O}}^S\right]_+^2}{\kappa_{\mathrm{O}}} > \frac{\left[B_{\mathrm{T}}^S\right]_+^2}{\kappa_{\mathrm{T}}}. \tag{190}$$

For a coexistence target j with refuge k, target-kill coverage is

$$a_j^{kill} = \frac{m_j D}{F\eta w_j(\lambda + w_k)}. \tag{191}$$

In a j-only cell, kill coverage is $A_j/(F\eta)$. If refuge k has net attractiveness $A_k > 0$, its entry boundary is

$$a_k^{entry} = \frac{A_j - (\lambda + w_j)A_k/w_j}{F\eta}, \tag{192}$$

when this expression lies in the current cell. The first routing cap is the smallest positive feasible value among tactic kill, refuge entry, and the unit bound. With a binding cap or a cell that begins above zero, audit ranking uses the attained cell values and the finite adjoining-cell candidate comparison rather than equation (190)*.*

Proof. Within routing cell S, a unit of coverage creates penalty $F\eta$ and reduces harm at rate $F\eta G_x^S$. Fixed screened pools create linear costs $\chi_x \alpha \bar{\ell}_x a_x$ and $\chi_{10,x}\eta \bar{n}_{10,x} a_x$. Adding direct cost $c_x a_x + \kappa_x a_x^2/2$, regulator gain relative to the cell's lower-bound continuation value is

$$\zeta_x^S + B_x^S a_x - \frac{\kappa_x}{2} a_x^2, \tag{193}$$

where ζ_x^S makes value continuous with the preceding cell. If zero belongs to the cell and value is measured relative to zero coverage, $\zeta_x^S = 0$.

The derivative of equation (193) is $B_x^S - \kappa_x a_x$, and its second derivative is $-\kappa_x < 0$. The unconstrained maximizer is B_x^S/κ_x, so projection onto the complete closed interval proves equation (188). In a zero-containing cell with upper endpoint $\bar{a}$, the cases are

$$a_{x,S}^* = \begin{cases} 0, & B_x^S \leq 0, \\ B_x^S/\kappa_x, & 0 < B_x^S < \kappa_x \bar{a}, \\ \bar{a}, & B_x^S \geq \kappa_x \bar{a}. \end{cases} \tag{194}$$

In the first case, value is zero. In the uncapped interior case, substitution gives

$$B_x^S \frac{B_x^S}{\kappa_x} - \frac{\kappa_x}{2}\left(\frac{B_x^S}{\kappa_x}\right)^2 = \frac{(B_x^S)^2}{2\kappa_x}. \tag{195}$$

Combining the nonpositive and positive cases gives equation (189). Comparing the two uncapped values and multiplying by the positive constant two gives equation (190). If either cap binds, its value is instead $B_x^S \bar{a} - \kappa_x \bar{a}^2/2$, so the uncapped inversion inequality no longer applies.

In coexistence, Lemma 9 gives

$$\frac{\partial m_j}{\partial a_j} = -F\eta \frac{w_j(\lambda + w_k)}{D}. \tag{196}$$

Starting from mass m_j, the affine path reaches zero at the coverage in equation (191). In a j-only cell, net attractiveness becomes $A_j - F\eta a_j$, so target mass reaches zero at

$$a_j^{kill} = \frac{A_j}{F\eta}. \tag{197}$$

The refuge remains inactive while the single-active condition in equation (116) or equation (118) holds. With generic labels, equality is

$$A_k = \frac{w_j(A_j - F\eta a_j)}{\lambda + w_j}. \tag{198}$$

Multiplying by $\lambda + w_j$, moving the coverage term to the other side, and dividing by $F\eta w_j$ gives equation (192). The active set changes at the first positive equality reached, so the operative cap is the minimum of entry, kill, and one.

At a cap, Lemma 8 makes routing and objective value continuous. The next cell begins with an intercept chosen to equal the value at that boundary. Applying equation (188) in each adjoining interval and comparing every projected candidate and shared endpoint therefore exhausts the one-instrument compact domain. This proves the capped-cell and ranking statements. □

F Platform Adoption and the Global Enforcement Path

F.1 Adoption during coexistence

Target the opaque tactic and use the fixed-pool benchmark. Define

$$\varphi = F\eta, \qquad \sigma_B = \frac{w_{\mathrm{O}}(\lambda + w_{\mathrm{T}})}{D}, \qquad Q_B = \varphi\sigma_B. \tag{199}$$

Let $m > 0$ be counterfactual opaque mass before installing the platform. While both tactics remain active,

$$m_{\mathrm{O}}(a) = m - Q_B a. \tag{200}$$

Let B_B denote the net marginal coverage benefit in coexistence, as defined by equation (187), and write $\kappa = \kappa_{\mathrm{O}}$. Useful coverage is bounded by

$$\bar{a}(m) = \min\{1, m/Q_B\}. \tag{201}$$

Operating value before fixed cost is

$$S(B_B, m) = \max_{0 \le a \le \bar{a}(m)} \left\{ B_B a - \frac{\kappa}{2} a^2 \right\}. \tag{202}$$

Theorem 3 (Both-active adoption)**.** *Let $K \ge 0$, $\kappa > 0$, and $Q_B > 0$.*

(i) If $B_B \le 0$, positive coverage is not optimal and an architecture with positive fixed cost is not adopted.

(ii) If $B_B > 0$, define

$$\widehat{a}_B = \min\{B_B/\kappa, 1\}, \qquad S_B^{max} = B_B \widehat{a}_B - \frac{\kappa}{2}\widehat{a}_B^2. \tag{203}$$

If $K > S_B^{max}$, the platform is not adopted in this cell for any target mass.

(iii) If $B_B > 0$ and $0 \le K \le S_B^{max}$, define

$$a_K^- = \frac{B_B - \sqrt{B_B^2 - 2\kappa K}}{\kappa}, \qquad m_K = Q_B a_K^-. \tag{204}$$

The platform is adopted exactly when $m \ge m_K$, with indifference at equality. Optimal coverage is

$$a^*(m) = \min\{\widehat{a}_B, m/Q_B\}. \tag{205}$$

For $K > 0$, first adoption has $a^* = a_K^- = m_K/Q_B$, so opaque activity reaches zero. Of the *removed mass, $w_T/(\lambda + w_T)$ migrates to transparency and $\lambda/(\lambda + w_T)$ exits.*

Proof. The derivative of the objective inside equation (202) is $B_B - \kappa a$.

If $B_B \leq 0$, this derivative is nonpositive for every $a \geq 0$. The maximizer is $a = 0$ and gross operating value is zero. A positive fixed cost cannot be recovered.

Now let $B_B > 0$. The unconstrained maximizer is B_B/κ. Imposing the unit cap gives $\widehat{a}_B$. Substitution yields equation (203). To see the two familiar branches, if $B_B < \kappa$ then $\widehat{a}_B = B_B/\kappa$ and

$$S_B^{max} = \frac{B_B^2}{2\kappa}. \tag{206}$$

If $B_B \geq \kappa$, then $\widehat{a}_B = 1$ and $S_B^{max} = B_B - \kappa/2$. The mass cap can only lower value, so $K > S_B^{max}$ rules out adoption.

Take $K \leq S_B^{max}$. On the rising branch, the cost-recovery equation is

$$B_B a - \frac{\kappa}{2} a^2 = K. \tag{207}$$

Multiplication by two and rearrangement gives

$$\kappa a^2 - 2B_B a + 2K = 0. \tag{208}$$

The quadratic formula gives two roots

$$a_K^{\pm} = \frac{B_B \pm \sqrt{B_B^2 - 2\kappa K}}{\kappa}. \tag{209}$$

The discriminant is nonnegative. When $B_B < \kappa$, this follows from $K \leq B_B^2/(2\kappa)$. When $B_B \geq \kappa$,

$$B_B^2 - 2\kappa K \geq B_B^2 - 2\kappa(B_B - \kappa/2) \tag{210}$$

$$= (B_B - \kappa)^2 \geq 0. \tag{211}$$

The smaller root lies in $[0, \widehat{a}_B]$. The larger root lies on the declining branch or beyond the unit cap. Since coverage is chosen to maximize value, cost recovery begins at the smaller root.

The available coverage cap m/Q_B reaches a_K^- exactly when $m = Q_B a_K^-$. Below that mass, the highest feasible point on the rising branch has value below K. At or above it, the feasible set contains a point with value at least K. This proves the adoption rule. Maximizing the concave quadratic subject to the mass and unit caps gives equation (205).

If $K > 0$, $a_K^- > 0$. At first adoption $m = m_K$ and equation (200) gives

$$m_O(a_K^-) = m_K - Q_B a_K^- = 0. \tag{212}$$

The migration and exit shares follow from Lemma 9 with the transparent tactic as refuge. □

F.2 The no-platform legacy path

Let the legacy transparent penalty be t, hold opaque net attractiveness at $u > 0$, and set

$$A_{\mathrm{T}}(t) = v_{\mathrm{T}} - t, \qquad A_{\mathrm{O}} = u. \tag{213}$$

The entry and exit thresholds are

$$t_e = v_{\mathrm{T}} - \frac{\lambda + w_{\mathrm{T}}}{w_{\mathrm{T}}} u, \qquad t_x = v_{\mathrm{T}} - \frac{w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}} u. \tag{214}$$

Their difference is

$$t_x - t_e = u \left\{ \frac{\lambda + w_{\mathrm{T}}}{w_{\mathrm{T}}} - \frac{w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}} \right\} \tag{215}$$

$$= \frac{\lambda(\lambda + w_{\mathrm{T}} + w_{\mathrm{O}})u}{w_{\mathrm{T}}(\lambda + w_{\mathrm{O}})} > 0. \tag{216}$$

Theorem 1 then gives transparent-only routing for $t \leq t_e$, coexistence for $t_e < t < t_x$, and opaque-only routing for $t \geq t_x$, after intersecting these ranges with the policy domain $t \geq 0$. In coexistence,

$$m^0_{\mathrm{O}}(t) = \frac{w_{\mathrm{T}} w_{\mathrm{O}}}{D}(t - t_e). \tag{217}$$

After transparent exit,

$$m^0_{\mathrm{O}}(t) = \bar{m}_{\mathrm{O}} = \frac{\lambda w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}} u, \tag{218}$$

which no longer changes with t.

F.3 Opaque-only regulation and transparent re-entry

Opaque coverage changes its net attractiveness to $u - pa$. Starting from $t \geq t_x$, the transparent tactic remains inactive while

$$A_{\mathrm{T}}(t) \leq \frac{w_{\mathrm{O}}(u - pa)}{\lambda + w_{\mathrm{O}}}. \tag{219}$$

Solving equality for coverage gives the re-entry boundary

$$r(t) = \frac{u - \frac{\lambda + w_{\mathrm{O}}}{w_{\mathrm{O}}} A_{\mathrm{T}}(t)}{\varphi} = \frac{\lambda + w_{\mathrm{O}}}{p w_{\mathrm{O}}}(t - t_x). \tag{220}$$

If $A_{\mathrm{T}}(t) > 0$, opaque activity reaches zero in coexistence when

$$(\lambda + w_{\mathrm{T}})(u - pa) = w_{\mathrm{T}} A_{\mathrm{T}}(t). \tag{221}$$

The corresponding kill boundary is

$$k(t) = \frac{u - \frac{w_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}} A_{\mathrm{T}}(t)}{\varphi}. \tag{222}$$

For $A_{\mathrm{T}} > 0$,

$$k(t) - r(t) = \frac{A_{\mathrm{T}}(t)}{\varphi} \left\{ \frac{\lambda + w_{\mathrm{O}}}{w_{\mathrm{O}}} - \frac{w_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}} \right\} > 0. \tag{223}$$

If $A_{\mathrm{T}} \leq 0$, transparent re-entry is unavailable and opaque kill occurs at

$$L = \frac{u}{p}. \tag{224}$$

Let

$$G_0 = \frac{h_{\mathrm{O}} \lambda w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}}, \qquad G_B = G_{\mathrm{O}}^{\{\mathrm{T},\mathrm{O}\}}, \tag{225}$$

and collect marginal implementation and error costs in

$$d = c_{\mathrm{O}} + \chi_{\mathrm{O}} \alpha \bar{\ell}_{\mathrm{O}} + \chi_{10,\mathrm{O}} \eta \bar{n}_{10,\mathrm{O}}. \tag{226}$$

Net coverage benefits before and after transparent re-entry are

$$B_0 = pG_0 - d, \qquad B_B = pG_B - d. \tag{227}$$

Up to opaque kill, exact operating value is

$$W(a;t) = \begin{cases} B_0 a - \kappa a^2/2, & 0 \leq a \leq r(t), \\ B_B a + (B_0 - B_B) r(t) - \kappa a^2/2, & r(t) \leq a \leq k(t). \end{cases} \tag{228}$$

The second line adds value $B_0 r$ from the opaque-only segment and value $B_B(a - r)$ after re-entry, then subtracts total quadratic cost.

Lemma 12 (All post-exit coverage candidates)**.** *Let*

$$\mathcal{I}_0 = [0, \min\{1, r(t), k(t)\}], \tag{229}$$

$$\mathcal{I}_B = [\max\{0, r(t)\}, \min\{1, k(t)\}], \tag{230}$$

with the second interval omitted when empty. The global operating optimum is among

$$\Pi_{\mathcal{I}_0}(B_0/\kappa), \qquad \Pi_{\mathcal{I}_B}(B_B/\kappa), \tag{231}$$

and the feasible endpoints $0, r(t), k(t), 1$. *If* $A_\mathrm{T} \le 0$*, replace* $k(t)$ *with* L *and omit* $\mathcal{I}_B$. *An optimum exists and is unique unless distinct listed candidates have the same value.*

Proof. On each nonempty branch of equation (228), the second derivative is $-\kappa < 0$. The first branch has stationary point B_0/κ and the second has stationary point B_B/κ. A strictly concave quadratic on a closed interval is maximized at the projection of its stationary point onto that interval. The feasible set is the union of the two closed branch intervals. A global maximizer is therefore one of the two branch maximizers or a shared and outer endpoint. Compactness gives existence. Strict concavity gives uniqueness within each branch, leaving multiplicity only when candidates from different branches tie. When $A_\mathrm{T} \le 0$, there is one branch on $[0, \min\{1, L\}]$ and the same projection argument applies. □

F.4 Exact delayed adoption

Define

$$\vartheta = \frac{w_\mathrm{T} w_\mathrm{O}}{(\lambda + w_\mathrm{T})(\lambda + w_\mathrm{O})}, \qquad k_0 = (1 - \vartheta)L. \tag{232}$$

Maintain

$$0 < B_B < B_0, \tag{233}$$

and define

$$a_B = \frac{B_B}{\kappa} < \min\{1, k_0\}, \qquad a_0 = \min\{B_0/\kappa, 1, L\} > a_B. \tag{234}$$

Let

$$S_B = \frac{B_B^2}{2\kappa}, \quad \Delta = B_0 - B_B, \quad S_1 = S_B + \Delta a_B, \quad S_0 = B_0 a_0 - \frac{\kappa}{2} a_0^2. \tag{235}$$

Theorem 4 (Delayed adoption after transparent exit)**.** *Under equations* (233) *and* (234)*, optimized operating value as a function of the re-entry boundary is*

$$\mathscr{V}(r) = \begin{cases} S_B + \Delta r, & 0 \le r \le a_B, \\ B_0 r - \kappa r^2/2, & a_B \le r \le a_0, \\ S_0, & r \ge a_0. \end{cases} \tag{236}$$

If $S_B < K \le S_0$, the platform first becomes worthwhile after transparent exit. The threshold is

$$r_K = \begin{cases} (K - S_B)/\Delta, & S_B < K \le S_1, \\ \{B_0 - \sqrt{B_0^2 - 2\kappa K}\}/\kappa, & S_1 < K \le S_0, \end{cases} \tag{237}$$

and

$$t_D = t_x + \frac{p w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}} r_K. \tag{238}$$

If $K < S_1$, first delayed coverage is $a_B > r_K$, so transparency re-enters and opacity is partially suppressed. If $K = S_1$, coverage equals the re-entry boundary. If $S_1 < K \le S_0$, coverage is r_K and transparency remains at its boundary. Opaque activity survives except when $K = S_0$, $a_0 = L$, and both tactics reach zero net attractiveness.

Proof. At $t = t_x$, equation (220) gives $r = 0$. At $t = v_{\mathrm{T}}$, it gives $r = L$. To express the kill boundary in terms of r, use

$$t = t_x + \frac{p w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}} r \tag{239}$$

and

$$v_{\mathrm{T}} - t_x = \frac{w_{\mathrm{O}} u}{\lambda + w_{\mathrm{O}}} = \frac{p w_{\mathrm{O}} L}{\lambda + w_{\mathrm{O}}}. \tag{240}$$

Then

$$A_{\mathrm{T}}(t) = \frac{p w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}} (L - r). \tag{241}$$

Substitution into equation (222) yields

$$k(r) = L - \frac{w_{\mathrm{T}} w_{\mathrm{O}}}{(\lambda + w_{\mathrm{T}})(\lambda + w_{\mathrm{O}})} (L - r) \tag{242}$$

$$= (1 - \vartheta) L + \vartheta r. \tag{243}$$

Condition $a_B < k_0$ makes a_B feasible at $r = 0$, and equation (243) keeps it feasible for every $r \ge 0$ considered below.

Take $0 \le r \le a_B$. The stationary point on the second branch of equation (228) is $a_B = B_B/\kappa$, which is feasible because $a_B \ge r$ and $a_B < k(r)$. Substitution gives

$$\mathscr{V}(r) = B_B a_B + \Delta r - \frac{\kappa}{2} a_B^2 \tag{244}$$

$$= \frac{B_B^2}{\kappa} + \Delta r - \frac{B_B^2}{2\kappa} \tag{245}$$

$$= S_B + \Delta r. \tag{246}$$

Take $a_B \le r \le a_0$. At the kink $a = r$, the derivative from the first branch is

$$B_0 - \kappa r \ge B_0 - \kappa a_0 \ge 0, \tag{247}$$

where the final inequality follows from $a_0 \le B_0/\kappa$. The derivative from the second branch is

$$B_B - \kappa r \le B_B - \kappa a_B = 0. \tag{248}$$

Thus value rises up to the kink and falls after it. The unique optimum is $a = r$, giving $B_0 r - \kappa r^2/2$.

For $r \ge a_0$, the first branch contains the constrained opaque-only maximizer a_0. It yields S_0, and additional movement of the re-entry boundary does not change the chosen coverage. The three expressions agree at $r = a_B$ because

$$B_0 a_B - \frac{\kappa}{2} a_B^2 = B_B a_B + \Delta a_B - \frac{\kappa}{2} a_B^2 = S_B + \Delta a_B = S_1. \tag{249}$$

They agree at $r = a_0$ by the definition of S_0. This proves equation (236).

Since $\mathscr{V}(0) = S_B$, the strict inequality $K > S_B$ rules out adoption at transparent exit. The value rises continuously to S_0. On the first branch, solving $S_B + \Delta r = K$ gives the first line of equation (237); this solution lies in $(0, a_B]$ exactly when $K \le S_1$. On the second branch, solve

$$B_0 r - \frac{\kappa}{2} r^2 = K. \tag{250}$$

The smaller root is the second line of equation (237); it lies in $(a_B, a_0]$ when $S_1 < K \le S_0$. Inverting equation (220) gives equation (238).

When $K < S_1$, the threshold has $r_K < a_B$, and the first-region optimum is a_B. Coverage exceeds the re-entry boundary, so the equilibrium returns to coexistence. When $K = S_1$, $r_K = a_B$. For $K > S_1$, the optimum at first adoption is the kink $a = r_K$. Finally, equation (243) gives $k(r) > r$ for every $r < L$, so coverage at or just above the re-entry boundary leaves opaque activity positive. Equality $k(r) = r$ requires $r = L$. At first adoption this occurs only at the stated terminal corner. □

F.5 The global stationary phase theorem

Define

$$t_s = t_e + \frac{\varphi(\lambda + w_{\mathrm{T}})}{w_{\mathrm{T}}} a_B, \qquad t_R = t_x + \frac{\varphi w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}} a_B, \qquad t_0 = t_x + \frac{\varphi w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}} a_0. \tag{251}$$

The slack conditions imply

$$t_e < t_s < t_x < t_R < t_0 \le v_{\mathrm{T}}. \tag{252}$$

The inequalities $t_e < t_s$, $t_x < t_R$, and $t_R < t_0$ follow from positive coefficients and $a_0 > a_B$. For $t_s < t_x$, first use $a_B < k_0 = (1-\vartheta)L$:

$$
\begin{aligned}
t_s - t_e &= \frac{\varphi(\lambda + w_\mathrm{T})}{w_\mathrm{T}} a_B && (253)\\
&< \frac{\varphi(\lambda + w_\mathrm{T})}{w_\mathrm{T}} \left\{1 - \frac{w_\mathrm{T} w_\mathrm{O}}{(\lambda + w_\mathrm{T})(\lambda + w_\mathrm{O})}\right\} L && (254)\\
&= u \left\{\frac{\lambda + w_\mathrm{T}}{w_\mathrm{T}} - \frac{w_\mathrm{O}}{\lambda + w_\mathrm{O}}\right\} && (255)\\
&= t_x - t_e, && (256)
\end{aligned}
$$

where $\varphi L = u$ and the last equality uses equation (214). Hence $t_s < t_x$. To verify $t_0 \leq v_\mathrm{T}$, use $a_0 \leq L$:

$$
t_0 - t_x \leq \frac{\varphi w_\mathrm{O}}{\lambda + w_\mathrm{O}} L = \frac{w_\mathrm{O} u}{\lambda + w_\mathrm{O}} = v_\mathrm{T} - t_x. \tag{257}
$$

Theorem 5 (Global build and routing phases)**.** *Maintain equations* (233) *and* (234)*, positive fixed cost, and a build-on-tie convention. Every phase interval below is intersected with the policy domain $t \geq 0$. If a computed first-build threshold is negative, the stationary platform is already installed at $t = 0$ in the phase prescribed by the theorem, and there is no in-domain build event.*

(A) If $0 < K \leq S_B$, define

$$
t_B = t_e + \frac{\varphi(\lambda + w_\mathrm{T})}{w_\mathrm{T}} a_K^-, \tag{258}
$$

where a_K^- uses B_B in equation (204)*. The path is transparent only before t_e, coexistence until t_B, transparent only under kill-cap coverage until t_s, coexistence under coverage a_B until t_R, opaque only with coverage tracking the re-entry boundary until t_0, and opaque only with coverage a_0 thereafter, subject to the terminal no-entry condition below.*

(B) If $S_B < K \leq S_0$, there is no build before transparent exit. The no-platform path reaches opaque-only routing, and the platform is built at t_D from equation (238)*. When $K < S_1$, this investment reactivates transparency until t_R. When $K \geq S_1$, coverage remains on the re-entry boundary. Coverage tracks that boundary until t_0 and equals a_0 afterward.*

(C) If $K > S_0$, the platform is never adopted along the stationary path and routing follows the no-platform regions.

At high legacy enforcement, terminal opaque mass is

$$m_{\mathrm{O}}^{\infty}=\begin{cases}0, & K\le S_0 \text{ and } a_0=L=u/\varphi\le 1,\\ \dfrac{\lambda w_{\mathrm{O}}}{\lambda+w_{\mathrm{O}}}(u-\varphi a_0), & K\le S_0 \text{ and } a_0<L,\\ \dfrac{\lambda w_{\mathrm{O}}}{\lambda+w_{\mathrm{O}}}u, & K>S_0.\end{cases} \tag{259}$$

Thus terminal no entry occurs exactly in the first case.

Proof. The no-platform routing phases and opaque mass path follow from equations (214), (217) and (218).

During coexistence, the counterfactual opaque mass is

$$m_{\mathrm{O}}^{0}(t)=\frac{w_{\mathrm{T}}w_{\mathrm{O}}}{D}(t-t_e). \tag{260}$$

The mass reduction per unit opaque coverage is

$$Q_B=\varphi\frac{w_{\mathrm{O}}(\lambda+w_{\mathrm{T}})}{D}. \tag{261}$$

Equating counterfactual mass to the first-adoption mass $Q_B a_K^-$ gives

$$\frac{w_{\mathrm{T}}w_{\mathrm{O}}}{D}(t_B-t_e)=\varphi\frac{w_{\mathrm{O}}(\lambda+w_{\mathrm{T}})}{D}a_K^-, \tag{262}$$

$$t_B-t_e=\frac{\varphi(\lambda+w_{\mathrm{T}})}{w_{\mathrm{T}}}a_K^-, \tag{263}$$

which proves equation (258). When $K\le S_B$, the smaller root satisfies $a_K^-\le a_B$, so $t_B\le t_s$. Theorem 3 shows that coverage tracks the opaque kill cap from t_B until the cap reaches the unconstrained coexistence optimum a_B at t_s. Routing is transparent only on this segment because opacity is held at zero.

For $t>t_s$, chosen coverage stays at a_B. Counterfactual opaque mass continues to grow, so opacity reappears continuously and both tactics are active until the no-platform transparent-exit point. After t_x, the transparent tactic is active under the platform exactly while $a_B>r(t)$. Solving $r(t)=a_B$ with equation (220) gives t_R. From t_R onward, Theorem 4 shows that optimal coverage tracks $r(t)$ until $r(t)=a_0$, which gives t_0. Coverage then stays at a_0. This proves the early-build phase sequence.

If $S_B<K\le S_0$, coexistence operating value never reaches the fixed cost before transparent exit under the maintained slack conditions, while Theorem 4 gives one post-exit crossing at t_D. Its proof establishes the active set at first adoption and the subsequent movement toward t_R and t_0. This

proves the delayed-build sequence.

If $K > S_0$, even the largest post-exit operating value cannot recover the fixed cost. Since $S_B < S_0$, coexistence value cannot recover it either. This proves nonadoption.

Restricting the ordered real-line phases to $t \geq 0$ removes every interval whose upper endpoint is negative. If the relevant build threshold is negative, $t = 0$ lies in a post-build interval, so the stationary architecture choice at the start of the policy domain is installed. If the threshold is nonnegative, it remains the first in-domain crossing. This proves the domain qualification in the statement.

At high t, $A_{\mathrm{T}} = v_{\mathrm{T}} - t \leq 0$. If $K \leq S_0$, the platform is installed under parts (A) or (B) and coverage equals a_0. No entry then requires $A_{\mathrm{O}} = u - \varphi a_0 \leq 0$. Since $a_0 \leq L = u/\varphi$, this inequality holds exactly when $a_0 = L$, and feasibility of that coverage requires $L \leq 1$. If $a_0 < L$, Theorem 1 gives opaque-only mass $\lambda w_{\mathrm{O}}(u - \varphi a_0)/(\lambda + w_{\mathrm{O}})$. If $K > S_0$, part (C) gives no platform and zero coverage, so the same routing theorem gives $\lambda w_{\mathrm{O}} u/(\lambda + w_{\mathrm{O}})$. These are the three cases in equation (259). □

At an early build threshold, the exact tactic-mass changes are

$$\Delta m_{\mathrm{T}} = \frac{w_{\mathrm{T}}}{\lambda + w_{\mathrm{T}}} m_K, \qquad -\Delta \mathcal{M} = \frac{\lambda}{\lambda + w_{\mathrm{T}}} m_K. \tag{264}$$

At delayed adoption, let pre-policy masses be $(0, \bar{m}_{\mathrm{O}})$. If chosen coverage equals r_K, post-policy transparent mass remains zero and opaque mass is the opaque-only mass at that coverage. If chosen coverage is $a_B > r_K$, the extra coverage $a_B - r_K$ operates in coexistence, producing

$$m_{\mathrm{T}}^{+} = \frac{p w_{\mathrm{T}} w_{\mathrm{O}}}{D}(a_B - r_K), \tag{265}$$

$$m_{\mathrm{O}}^{+} = \frac{\lambda w_{\mathrm{O}}}{\lambda + w_{\mathrm{O}}}(u - p r_K) - Q_B(a_B - r_K). \tag{266}$$

These expressions show the transparent reactivation and partial opaque suppression described in the main text.

When the aligned ordering fails, Lemma 12 remains necessary and sufficient at every t. In particular, $B_0 > 0 \geq B_B$ supports delayed-only adoption, $B_B > 0 \geq B_0$ supports value only during coexistence, $B_B > B_0 > 0$ can produce declining post-exit value and bounded stationary adoption, and nonpositive benefits in both regions rule out positive coverage.

Proof of Proposition 4. Part (i) follows from Theorem 3. Part (ii) follows from Theorem 4 and the fixed-cost partition in Theorem 5. The active set at first delayed adoption, the survival of opacity, and the terminal exception are proved in the final portion of Theorem 4's proof. These results establish every statement in Proposition 4. □

G Welfare Comparison with a Cancellation Rule

G.1 Equal-deterrence comparison

Compare a source policy and a cancellation rule that induce the same complete tactic-penalty vector. Theorem 1 then gives identical net attractiveness, routing masses, and manipulation harm.

Write the common penalty vector as $\mathbf{t} = (t_{\mathrm{T}}, t_{\mathrm{O}})$. For the source policy, tactic-specific coverage and bona fide burden are

$$a_{S,x}(\mathbf{t}, J) = \frac{t_x}{F\eta(J,b)}, \qquad p_{S,x}(\mathbf{t}, J) = \frac{f_x t_x}{F}\frac{\alpha(J,b)}{\eta(J,b)}, \tag{267}$$

provided each coverage is at most one. The complete source error term in tactic x is

$$D^{err}_{S,x} = \chi_{S,x} a_{S,x} \alpha \ell_x(p_{S,x}) + \chi_{10,S,x} a_{S,x} \eta n_{10,S,x}. \tag{268}$$

Counting shared platform cost once, total source implementation cost is

$$C_S(\mathbf{t}, J) = K(J) + \sum_x \left[c_{S,x} a_{S,x} + \frac{\kappa_{S,x}}{2} a^2_{S,x} + \{\mathcal{V}_x(0) - \mathcal{V}_x(p_{S,x})\} + D^{err}_{S,x} \right]. \tag{269}$$

For a cancellation rule calibrated to the same vector, let $p_{C,x}(\mathbf{t})$ be its burden on bona fide providers, $A_C(\mathbf{t})$ its total administrative cost, and $D_{C,x}(\mathbf{t})$ its tactic-specific adjudicative and market-design loss. Total implementation cost is

$$C_C(\mathbf{t}) = A_C(\mathbf{t}) + \sum_x \left[\mathcal{V}_x(0) - \mathcal{V}_x(p_{C,x}) + D_{C,x}(\mathbf{t}) \right]. \tag{270}$$

Lemma 13 (Exact equal-deterrence ranking)**.** *At a common feasible tactic-penalty vector, source enforcement has higher welfare exactly when*

$$C_S(\mathbf{t}, J) < C_C(\mathbf{t}). \tag{271}$$

A sufficient set of conditions for weak source dominance is

$$p_{S,x}(\mathbf{t}, J) \le p_{C,x}(\mathbf{t}) \qquad \textit{for every } x, \tag{272}$$

$$K + \sum_x \left(c_{S,x} a_{S,x} + \frac{\kappa_{S,x}}{2} a^2_{S,x} \right) \le A_C(\mathbf{t}), \tag{273}$$

$$\sum_x D^{err}_{S,x}(\mathbf{t}, J) \le \sum_x D_{C,x}(\mathbf{t}). \tag{274}$$

The ranking is strict if one of the last two inequalities is strict, or if the first comparison is strict in some cell with $p_{S,x} < s_x$. Reversing all three sets of comparisons gives the corresponding cancellation-rule result.

Proof. Equal penalty vectors give equal $A_x = v_x - \tau_x$. Uniqueness in Theorem 1 gives equal masses, so manipulation harm is equal. Welfare differs only by the implementation costs in equations (269) and (270). This proves the exact condition equation (271).

For $0 \le p < s_x$, differentiation of legitimate-sector value in cell x gives

$$\frac{\mathrm{d}\mathcal{V}_x(p)}{\mathrm{d}p} = v_x \left\{-(s_x - p) - q_x^L\right\} < 0. \tag{275}$$

For $p \ge s_x$, the positive parts make $\mathcal{V}_x(p) = 0$. Thus each $\mathcal{V}_x$ is weakly decreasing everywhere and strictly decreasing when its smaller argument lies below s_x. The first set of inequalities in equation (274) therefore makes every source chilling term weakly smaller, with the stated strictness condition. The second and third inequalities order all remaining implementation terms after summing across tactics. Adding the three comparisons proves weak source dominance and the strict cases. Reversing the comparisons proves the cancellation-rule conditions. □

If the source policy can implement every vector available to the cancellation rule and has weakly lower implementation cost at each common vector, maximizing over the common feasible set preserves weak source dominance. The same pointwise argument applies in reverse when the cancellation rule has the larger feasible set and weakly lower cost throughout.

G.2 Different penalty vectors and displacement

Suppose both tactics are active. Let policy $r \in \{S, C\}$ create infinitesimal penalty vector $(\mathrm{d}\tau_{\mathrm{T}}^r, \mathrm{d}\tau_{\mathrm{O}}^r)$ at implementation cost $\mathrm{d}C_r$. Lemma 9 gives

$$\mathrm{d}H = -G_{\mathrm{T}}\,\mathrm{d}\tau_{\mathrm{T}}^r - G_{\mathrm{O}}\,\mathrm{d}\tau_{\mathrm{O}}^r. \tag{276}$$

The policy's first-order welfare gain is therefore

$$\mathrm{d}W_r = G_{\mathrm{T}}\,\mathrm{d}\tau_{\mathrm{T}}^r + G_{\mathrm{O}}\,\mathrm{d}\tau_{\mathrm{O}}^r - \mathrm{d}C_r. \tag{277}$$

Lemma 14 (Displacement can reverse an implementation-cost advantage)**.** *Suppose* $\mathrm{d}C_C - \mathrm{d}C_S > 0$*, so source enforcement has lower implementation cost. The cancellation rule nevertheless has the larger first-order welfare gain exactly when*

$$G_{\mathrm{T}}(\mathrm{d}\tau_{\mathrm{T}}^C - \mathrm{d}\tau_{\mathrm{T}}^S) + G_{\mathrm{O}}(\mathrm{d}\tau_{\mathrm{O}}^C - \mathrm{d}\tau_{\mathrm{O}}^S)$$

$$> \mathrm{d}C_C - \mathrm{d}C_S. \tag{278}$$

A common cancellation penalty d$d > 0$ *on both tactics has gross harm gain*

$$\mathrm{d}d\,(G_\mathrm{T} + G_\mathrm{O}) = \mathrm{d}d\,\frac{\lambda(h_\mathrm{T}w_\mathrm{T} + h_\mathrm{O}w_\mathrm{O})}{D} > 0. \tag{279}$$

Proof. Subtract source gain from cancellation-rule gain using equation (277):

$$\mathrm{d}W_C - \mathrm{d}W_S = G_\mathrm{T}(\mathrm{d}\tau_\mathrm{T}^C - \mathrm{d}\tau_\mathrm{T}^S) + G_\mathrm{O}(\mathrm{d}\tau_\mathrm{O}^C - \mathrm{d}\tau_\mathrm{O}^S) \tag{280}$$

$$- (\mathrm{d}C_C - \mathrm{d}C_S). \tag{281}$$

This difference is positive exactly under equation (278).

To prove equation (279), use the general formula in equation (154) with $S = \{\mathrm{T}, \mathrm{O}\}$:

$$G_\mathrm{T} + G_\mathrm{O} = h_\mathrm{T}w_\mathrm{T} + h_\mathrm{O}w_\mathrm{O} - \frac{(w_\mathrm{T} + w_\mathrm{O})(h_\mathrm{T}w_\mathrm{T} + h_\mathrm{O}w_\mathrm{O})}{D} \tag{282}$$

$$= \left\{1 - \frac{w_\mathrm{T} + w_\mathrm{O}}{D}\right\}(h_\mathrm{T}w_\mathrm{T} + h_\mathrm{O}w_\mathrm{O}) \tag{283}$$

$$= \frac{\lambda(h_\mathrm{T}w_\mathrm{T} + h_\mathrm{O}w_\mathrm{O})}{D} > 0. \tag{284}$$

Multiplication by d$d > 0$ completes the proof. □

G.3 No unconditional ranking

Neither policy dominates for every admissible parameter vector. To obtain cancellation-rule dominance, hold its finite implementation cost fixed and raise source-platform cost K until equation (269) exceeds equation (270) at a common penalty vector. To obtain source dominance, take $J \to \infty$. Lemma 5 gives $\eta \to 1$ and $\alpha \to 0$, so for every tactic with positive target penalty,

$$a_{S,x} \to \frac{t_x}{F}, \qquad p_{S,x} \to 0. \tag{285}$$

With finite source administration cost, choose a cancellation rule whose bona fide burden reaches s_x in at least one valuable cell or whose total administration and legal-error costs are sufficiently large. Its legitimate-sector value in that cell then falls to zero or its remaining cost exceeds the source cost. These admissible sequences produce opposite strict rankings.

Proof of Proposition 5. Theorem 2 and Lemmas 13 and 14 establish existence, the equal-deterrence comparison, and the displacement comparison. The preceding parameter sequences rule out an unconditional ranking. These results prove Proposition 5. □

References

Aggarwal, R. K. and Wu, G. (2006). Stock market manipulations. *Journal of Business*, 79(4):1915–1953.

Allen, F. and Gale, D. (1992). Stock-price manipulation. *Review of Financial Studies*, 5(3):503–529.

Brogaard, J., Hendershott, T., and Riordan, R. (2019). Price discovery without trading: Evidence from limit orders. *Journal of Finance*, 74(4):1621–1658.

Brogaard, J., Li, D., and Yang, J. (2026). Spoofing and market quality. Working paper, revised February 2026.

Brunnermeier, M. K. and Pedersen, L. H. (2005). Predatory trading. *Journal of Finance*, 60(4):1825–1863.

Buchak, G., Matvos, G., Piskorski, T., and Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453–483.

Cartea, Á., Chang, P., and García-Arenas, G. (2025). Spoofing and manipulating order books with learning algorithms. Working paper, revised February 2025.

Cartea, Á., Jaimungal, S., and Wang, Y. (2020). Spoofing and price manipulation in order-driven markets. *Applied Mathematical Finance*, 27(1–2):67–98.

Christensen, H. B., Hail, L., and Leuz, C. (2016). Capital-market effects of securities regulation: Prior conditions, implementation, and enforcement. *Review of Financial Studies*, 29(11):2885–2924.

Comerton-Forde, C. and Putniņš, T. J. (2014). Stock price manipulation: Prevalence and determinants. *Review of Finance*, 18(1):23–66.

Commodity Futures Trading Commission (2013). Interpretive guidance and policy statement on disruptive practices. 78 Federal Register 31890.

Conrad, J., Wahal, S., and Xiang, J. (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics*, 116(2):271–291.

Cumming, D. J., Johan, S. A., and Li, D. (2011). Exchange trading rules and stock market liquidity. *Journal of Financial Economics*, 99(3):651–671.

Do, B. L. and Putniņš, T. J. (2023). Detecting layering and spoofing in markets. Working paper, revised November 2023.

Fischel, D. R. and Ross, D. J. (1991). Should the law prohibit "manipulation" in financial markets? *Harvard Law Review*, 105(2):503–553.

Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *Journal of Finance*, 66(1):1–33.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842.

Kaplow, L. (2011). On the optimal burden of proof. *Journal of Political Economy*, 119(6):1104–1140.

Kumar, P. and Seppi, D. J. (1992). Futures manipulation with "cash settlement". *Journal of Finance*, 47(4):1485–1502.

Kyle, A. S. and Viswanathan, S. (2008). How to define illegal price manipulation. *American Economic Review*, 98(2):274–279.

Lee, E. J., Eom, K. S., and Park, K. S. (2013). Microstructure-based manipulation: Strategic behavior and performance of spoofing traders. *Journal of Financial Markets*, 16(2):227–252.

Malik, A. S. (1990). Avoidance, screening and optimum enforcement. *RAND Journal of Economics*, 21(3):341–353.

Mookherjee, D. and Png, I. P. L. (1994). Marginal deterrence in enforcement of law. *Journal of Political Economy*, 102(5):1039–1066.

Plantin, G. (2015). Shadow banking and bank capital regulation. *Review of Financial Studies*, 28(1):146–175.

Stenfors, A., Doraghi, M., Soviany, C., Susai, M., and Vakili, K. (2023). Cross-market spoofing. *Journal of International Financial Markets, Institutions and Money*, 83:101735.

U.S. Securities and Exchange Commission (2012). Consolidated audit trail. Securities Exchange Act Release No. 34-67457, 77 Federal Register 45721, Rule 613 of Regulation NMS.

Williams, B. and Skrzypacz, A. (2025). Spoofing in equilibrium. *Journal of Finance*. Forthcoming.

Zhang, A. L. (2022). Competition and manipulation in derivative contract markets. *Journal of Financial Economics*, 144(2):396–413.