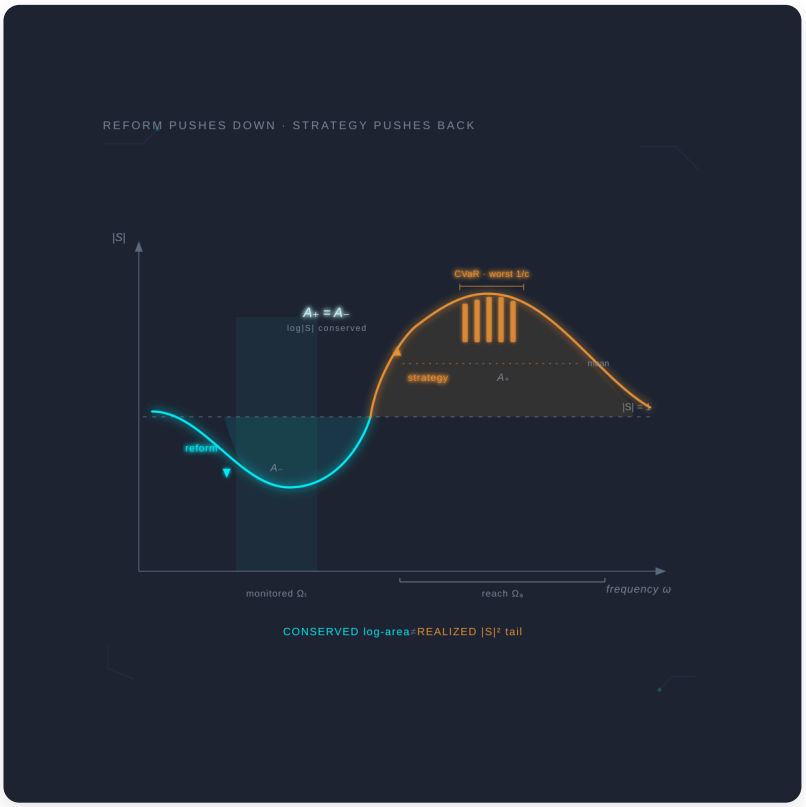

Where Reform Pushes Down, Strategy Pushes Back

Conserved sensitivity, an imported risk measure, and the geometry between them

Paper XXV in the Governance as Engineering series

Imports Bode's sensitivity integral and CVaR to separate the geometry of conserved amplification from the geometry of strategic loss. Shows that deeper proxy suppression raises the accessible exploitability floor, and that the gap between risk-blind and risk-aware design closes at maximal suppression. Paper XXV in the Governance as Engineering series.

Björn Kenneth Holmström

July 2026

`https://bjornkennethholmstrom.org/working-papers/where-reform-pushes-down`

Abstract

A recurring claim in the analysis of proxy-driven governance is that suppressing a monitored measure does not remove the underlying pressure but relocates it into channels the measure no longer sees. The claim is almost always made informally: it asserts that pressure migrates without supplying a budget, an accounting identity, or conditions under which the relocated pressure is actually exploitable. This paper supplies those three things by importing the strongest available formal analogue — Bode's sensitivity integral, a genuine conservation law for linear feedback — while being explicit about where the analogue stops.

We separate two objects that the informal argument runs together. The first is an allocation law describing how a strategic actor converts a fixed response profile into realized loss: a density-capped adversary with bounded budget, confined to a reachable set, realizes the upper-tail conditional-value-at-risk (CVaR) of the response over that set. This law is not new; it is the risk-envelope duality of CVaR / expected shortfall, equivalently a distributionally-robust worst case over a likelihood-ratio-bounded ambiguity set. The second object is Bode's integral, which constrains not the loss but the *profile*: for a stable loop of relative degree at least two, log-sensitivity conserves, so suppression in a monitored band is a lower bound on compensating amplification elsewhere **[R]**.

The paper's conceptual content is the relation between these two objects, and it is a mismatch of geometry. Conservation acts on the positive part of $\log|S|$; the adversary's loss functional acts on the upper tail of $|S|^2$. Because these are different functionals of the same profile, conserved pressure does not determine realized harm, and — more sharply — the accessible amplified *area* does not determine the accessible *loss* **[R]**. Exploitation is licensed by three separable capacities: concentration (can the actor stack pressure), accessibility (does the amplified region lie within reach), and discovery (is the amplified region legible to the actor). None is implied by conservation.

We then ask, constructively, how much freedom this geometry gap actually grants a designer, using a discrete-time convex synthesis. The minimum accessible exploitability achievable at a given proxy-suppression level — a value function over admissible controllers — is nondecreasing as suppression tightens by construction, and in the tested system rises strictly over the sampled range: buying deeper monitored-band suppression raises the exploitability floor. At a *fixed* suppression level a looser complementary-sensitivity allowance lowers that floor, because it enlarges the set of achievable profiles; but the advantage vanishes at the maximum attainable suppression, where the risk-blind and risk-aware optima coincide and design freedom has collapsed. The governance reading follows only in two explicitly separated layers, and we claim no domain-general theorem.

1. Introduction

The governance concern this paper addresses is familiar in the series. A system optimizes against a proxy — a monitored metric standing in for an objective it cannot observe directly — and the act of optimization erodes the correlation that made the proxy informative. Much of the series' diagnostic vocabulary is built around this Goodhart–Ashby mechanism. A particular corollary appears repeatedly and is rarely stated carefully: that tightening control over a monitored channel does not eliminate the pressure driving the system off-target but displaces it into channels that are not monitored, so that the ledger of trouble is conserved and merely relocated. Paper XVIII's laundering mechanism, in which local adaptation scrubs coupling evidence from precisely the residuals a registered index watched, is one instance; audit-evasion and regulatory arbitrage are others.

The trouble with the corollary is not that it is wrong but that, as usually stated, it is unfalsifiable. It asserts migration without a conserved quantity to migrate, without a budget bounding how much can move, and without any condition distinguishing the cases where relocation produces real harm from the cases where the pressure moves somewhere harmless. A claim of the form "pressure must reappear elsewhere" needs a conservation law to be more than a metaphor, and a claim that the reappearance is dangerous needs a licensing condition.

Control theory offers a genuine conservation law of exactly the required shape. Bode's sensitivity integral, in the Freudenberg–Looze form, states that for a stable feedback loop of relative degree at least two the integral of `log|S(jω)|` over frequency is zero: sensitivity suppressed in one band must be amplified in another, with the total log-area conserved. For a loop with unstable open-loop poles the conserved total is strictly positive, so the system must amplify more than it suppresses. This is the waterbed effect, and it is a theorem, not an analogy.

The temptation is to import it wholesale — to announce that governance reform is conserved, that every improvement in a monitored dimension is paid for by degradation in an unmonitored one, and to dress the claim in the integral's authority. The series has a standing commitment against exactly this move: borrowed formal authority, in which a heuristic claim is given the notation of a rigorous one and inherits an unearned certainty. We therefore proceed under two disciplines. First, every claim is tiered, and the conservation law is rigorous **[R]** only inside its licensing conditions — a linear time-invariant plant, a fixed controller, and non-strategic disturbances — none of which hold literally in governance, where the disturbances are strategic agents and the "plant" adapts. Second, we separate the imported mathematics from whatever is actually new, because most of what one needs here already exists and is better cited than reinvented.

That second discipline is worth stating plainly at the outset, because it disarms the obvious objection. The frequency-domain machinery in this paper is standard robust control. Weighted-sensitivity design already shapes $|S|$ to be small where disturbances are expected and lets it rise elsewhere; the worst-case band-limited disturbance concentrating at a sensitivity peak is the ordinary induced-norm picture; finite-frequency performance is the province of the generalized Kalman–Yakubovich–Popov lemma; the adversarial reading of disturbance rejection is the H_∞-as-game formulation of Başar and Bernhard; and the allocation law we use is CVaR duality from the risk literature. None of this is claimed as novel. What the paper offers is not a control result but a transfer discipline: a conserved quantity, an explicit decomposition of which agent controls what, and a checklist of conditions under which proxy suppression is actually dangerous rather than merely conserved. The one narrow technical question that appears genuinely open — design against the intermediate risk functional that interpolates between band-limited H_2 and finite-frequency H_∞ — is flagged where it arises and pursued in the synthesis section.

The remainder proceeds as follows. Section 2 states the imported allocation law and fixes notation. Section 3 introduces Bode's integral as a constraint on the achievable profile and establishes the paper's central structural point, that the conservation law and the harm functional inhabit different geometries. Section 4 gives the licensing conditions. Section 5 reports the constructive achievability result. Section 6 states the governance transfer in two separated layers, and Section 7 the limits.

---

2. The imported allocation law

Fix a scalar response profile $f(\omega) = |S(j\omega)|^2 \geq 0$ over a frequency variable ω, where S is the sensitivity function of a feedback loop to be specified in Section 3. A disturbance environment allocates energy across frequency; realized loss is the energy-weighted response,

$$J(D) = \int f(\omega)\, D(\omega)\, d\omega,$$

where D is a disturbance spectral density. A *passive* environment spreads a fixed budget uniformly over the reachable band; a *strategic* environment allocates the same budget to do as much damage as it can, subject to how concentrated it is permitted to be.

We model the strategic actor by three parameters. Its budget is a total energy E. Its reach is a set Ω_a — the frequencies at which it can place disturbance at all. Its concentration is a factor $c \geq 1$ bounding the density it may place at any single frequency to c times the uniform density over Ω_a; $c = 1$ forces the strategic actor to be indistinguishable from the passive one, and $c \to \infty$ permits it to place all of its budget at a single frequency. Writing μ_a for the normalized (uniform) measure on Ω_a, the strategic actor solves

$$J_{strat} = E \cdot \sup \left\{ \int f\, q\, d\mu_a : 0 \leq q \leq c, \int q\, d\mu_a = 1 \right\}.$$

The optimal q fills the highest-f frequencies it can reach up to the density cap, exhausting the budget on the worst reachable $1/c$ fraction of the band. The value of this program is exactly the conditional value-at-risk of f at level $\alpha = 1 - 1/c$:

$$J_{strat} / E = \mathrm{CVaR}_{1-1/c}(f ; \mu_a \text{ on } \Omega_a), \qquad \text{[imported]}$$

the mean of f over its worst $1/c$-measure fraction. The passive loss is the ordinary mean, $J_{pass} / E = E_{\mu_a}[f]$, and the strategic premium is the difference,

$$G / E = \mathrm{CVaR}_{1-1/c}(f) - E_{\mu_a}[f] = M \cdot U,$$

where $M = E_{\mu_a}[f]$ is the accessible mean level and $U = \mathrm{CVaR}/M - 1 \geq 0$ is a scale-invariant index of upper-tail heterogeneity. The concentration parameter interpolates two classical endpoints: at $c = 1$ the strategic loss is the accessible mean, a band-limited H_2-type quantity; as $c \to \infty$ it is the accessible essential supremum, the finite-frequency $H\infty$ gain.

This law is imported, not derived here. The supremum above is the risk-envelope (dual) representation of CVaR / expected shortfall: the worst-case expectation of `f` over all measures whose density with respect to the base measure is bounded by `1/α`. It is equivalently a distributionally-robust worst case over a likelihood-ratio-bounded ambiguity set. We therefore label it an *imported allocation lemma* (Rockafellar–Uryasev) and claim no novelty for it **[R, imported]**. Two qualifications travel with it. It is a *static* lemma: loss is linear in the allocated density and there is no temporal, causal, or transition cost to moving pressure between frequencies. And it silently presumes *discovery* — the value `CVaR` is attained only by an actor that knows where `f` is large and can therefore locate the upper tail; an actor with only noisy knowledge of `f` realizes strictly less, interpolating back toward the passive mean. We return to both qualifications in Section 4.

A note on verification, since the series treats it as a matter of discipline. The accompanying simulator computes `J_strat` two independent ways — an operational greedy water-filling allocator, and the analytic CVaR formula above — and confirms they agree to the order of 10^{-14}. This checks the discretization, not the theorem; the theorem is imported and the identity is exact.

---

3. Bode as an achievability constraint, and the geometry gap

The allocation law of Section 2 takes the profile `f` as given. What determines `f` is the controller, and this is where Bode's integral enters — not as a statement about loss, but as a constraint on which profiles a controller can present.

Let the loop transfer function be `L = PK` for plant `P` and controller `K`, with sensitivity `S = 1/(1+L)`. For a proper, internally stabilizing controller with no prohibited unstable cancellations, the Bode–Freudenberg–Looze integral gives

```
B_S = ∫₀^∞ log|S(jω)| dω = π · Σ Re(pₖ⁺) ≥ 0,          [R]
```

the sum taken over open-loop right-half-plane poles of `L`; the integral is zero when there are none and the loop has relative degree at least two. Decompose the log-sensitivity into its positive and negative areas, `A₊ = ∫ [log|S|]₊` and `A₋ = ∫ [log|S|]₋`, so that `B_S = A₊ − A₋`. Then `A₊ = A₋` when `B_S = 0`, and `A₊ > A₋` when uncancelled unstable poles make `B_S` positive; in either case `A₊ ≥ A₋`. Now let `A_m ≤ A₋` denote the log-sensitivity suppressed inside a monitored band `Ω_m` — the "proxy" whose performance the controller is optimizing. Because monitored suppression is a component of total suppression,

```
A₊ ≥ A₋ ≥ A_m,          [R, architecture-independent]
```

for *every* proper stabilizing controller. This is the honest, defensible core of the conservation intuition: a controller cannot suppress the monitored band for free, and the compensating amplification is at least as large as the suppression it bought. It holds regardless of controller architecture; it is an identity, not a feature of any particular design.

But the identity is weaker than it looks, and the reason is the paper's central structural point. Bode's integral constrains a functional of the profile — the positive part of `log|S|` — while the adversary of Section 2 evaluates a *different* functional of the same profile, the upper tail of `|S|²`. Conservation lives in log-sensitivity geometry; harm lives in squared-magnitude tail geometry. Neither functional determines the other absent additional support, peak, or achievability constraints — the counterexample below shows only that they are not identified in general — and the immediate consequence is: **the accessible amplified area does not, by itself, determine the accessible loss [R]**.

The point is sharp enough to warrant a bare counterexample. Consider two response profiles over a unit-measure reachable band. The first takes `|S|² = 4` on a quarter of the band and `1` on the rest; the second takes `|S|² = 2` on half the band and `1` on the rest. Both have identical accessible positive log-area, `A₊ = 0.1733`. Yet under a concentration factor `c = 2` the strategic losses differ — `J_strat/E = 2.50` for the first profile against `2.00` for the second — because the tail geometries differ even though the log-areas coincide. Equal conserved amplification, different realized harm.

This counterexample must be labelled carefully, because it is doing less than it might appear to. The two profiles are arbitrary measurable functions; they are not the sensitivity functions of any stabilizable feedback loop, and they take no account of the Bode constraint that couples suppression to amplification. The demonstration is therefore one of *functional non-identifiability* — that the two functionals are not related in general — and not a claim that a controller can realize an arbitrary gap between them. Whether the gap survives inside the manifold of *achievable* sensitivity profiles, once Bode's budget and engineering constraints are imposed, is a separate and more demanding question, and it is the subject of Section 5. We flag here only the direction of the answer: the gap is real inside the achievable manifold but modest, and it collapses as proxy pressure and robustness constraints bind.

The correct causal picture, then, is not a scalar chain running from monitored suppression to total amplification to loss. It is a two-stage structure in which the Bode budget constrains the achievable profile `f`, controller architecture shapes `f` within that budget, and only the triple `(f, Ω_a, c)` determines loss through the imported law. Conservation restricts the profile the defender may present; it does not restrict the tail functional the adversary applies to it. This separation — conservation constrains the profile, the harm metric is a different functional — is the canonical reading of fundamental limitations in Seron, Braslavsky and Goodwin, and we adopt it as the paper's organizing frame. The mechanism by which monitored-band suppression manufactures a compensating amplification peak, and the location of that peak relative to a reachable band, is shown in Figure 1.

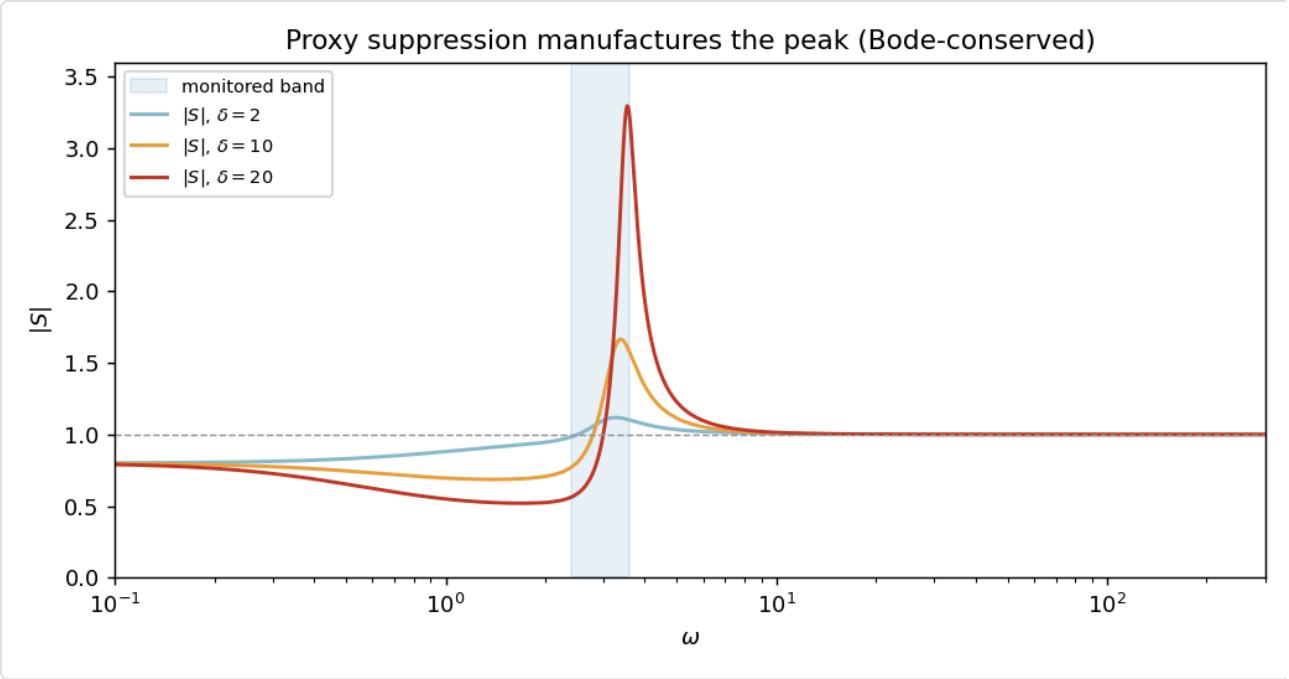

Figure 1. Monitored-band suppression manufactures a compensating amplification peak. Sensitivity `|S|` *for resonant controllers of increasing proxy gain* `δ`*; the monitored band* `Ω_m` *is shaded. Driving* `|S|` *below one inside* `Ω_m` *forces a taller peak just outside it, with the total log-sensitivity area conserved by Bode's integral. It is the accessible upper-tail distribution of* `|S|²`*, not the conserved log-area, that determines exploitability; location matters only because it changes that distribution (§3–§4).*

4. Licensing conditions

Section 3 established that conserved amplification does not determine realized loss; the imported law of Section 2 shows what does. Reading the two together yields the conditions under which a manufactured amplification becomes exploitable — the checklist the informal "pressure migrates" argument lacks. There are four capacities, and they are separable: none implies another, and each can be absent while the others hold.

Concentration. The strategic premium `G = J_strat − J_pass` is positive if and only if `c > 1` and the accessible profile `f` is non-constant almost everywhere **[R]**. This is immediate from the CVaR representation: `G/E = CVaR − mean` is the gap between the upper-tail mean and the ordinary mean, and that gap vanishes exactly when the distribution is degenerate or the actor cannot concentrate. An actor with no concentration capacity gains nothing from strategy, whatever the profile — it is reduced to the passive mean. What the condition does *not* require is any accessible frequency at which `|S| > 1`: a strategic actor extracts a premium over the passive baseline merely by sorting within its reach, even across a band that is entirely attenuated. The companion simulator exhibits exactly this — a reachable window on which every `|S|² < 1` still yields `G > 0` (Figure 2).

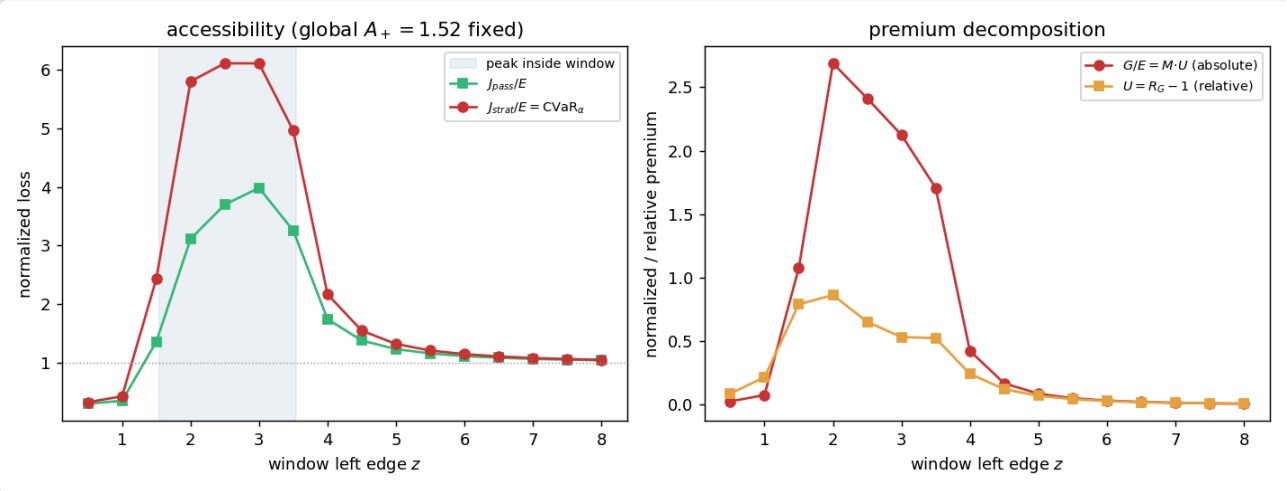

Figure 2. Left: passive and strategic loss as a fixed-width reachable window slides across frequency, with the global positive log-area `A_+` held fixed throughout; realized loss tracks whether the window contains the amplified region, not the conserved total — the accessibility condition. Right: the strategic premium decomposed as `G/E = M·U` (absolute) and `U = R_G − 1` (relative). The premium is positive wherever the accessible profile is heterogeneous, including windows on which `|S| < 1` everywhere — concentration does not require accessible amplification.

Realized amplification. The premium measures advantage over the passive baseline; it is a different question whether strategic loss exceeds the disturbance budget outright. That has its own condition: `J_strat > E` if and only if the cap-aware accessible upper-tail mean exceeds unity, `CVaR_{1−1/c}(f | Ω_a) > 1` **[R]**. This is where accessible `|S| > 1` genuinely matters — for absolute amplification, not for the premium. The distinction is worth stating plainly because conflating the two is an easy error: an actor can strictly out-perform a naive disturbance (`G > 0`) inside a band that produces no amplification at all (`J_strat < E`), and can equally suffer heavy amplification with no strategic premium if its accessible band is uniformly high.

Relocation. Both conditions above take the profile as fixed and ask whether the amplified region is reachable. The defender's countermove is to relocate the amplification out of reach, and whether this is possible is an achievability question that the continuous-time framing can make look settled when it is not. Bode's integral forces the positive log-sensitivity area to be strictly positive whenever suppression occurs, but it does not on its own force `|S| > 1` on an unbounded tail, nor fix where the amplification sits: on an unbounded frequency axis the integral supplies no controller-independent peak bound and no finite-localization constraint. (Whether `|S|` exceeds one at high frequency depends on the loop's relative degree and sign, a property of the chosen controller, not of the conservation law.) Whether a given controller class can realize remote or diffuse amplification — pushing the peak beyond any reachable band — is therefore a property of the achievable profile set, to be established by synthesis, not read off the integral. A discrete-time, sampled formulation makes the question well-posed by compactifying the axis to `[0, π/T_s]`, so relocation becomes redistribution within a closed budget; in the governance reading the sampling period is the institution's own bandwidth, its reporting or audit cycle. Even then, discrete time closes only export beyond the represented band, not relocation to weakly-monitored or near-Nyquist frequencies within it, so a complementary-sensitivity allowance remains indispensable — and, as Section 5 shows, that allowance is what governs how much relocation the achievable set actually permits **[R/IP]**.

Discovery. A fourth capacity is presumed silently by the imported law and deserves a name. The CVaR value is attained only by an actor that knows where `f` is large — that can locate the upper tail. Reach, concentration, and discovery are three separate things: an actor may be able to place pressure in the amplified band, stack it densely, and still fail to exploit it because it cannot find it. In the governance reading discovery is the legibility of the manufactured vulnerability to the strategic population; a peak no actor knows how to locate is not exploited, and an actor with only noisy knowledge of `f` realizes strictly less than CVaR, interpolating back toward the passive mean **[IP]**. The present model grants perfect discovery for free; relaxing it is the most decision-relevant extension.

The cost side: margin erosion. The conditions above describe when the adversary can exploit a manufactured peak. It is worth recording what manufacturing the peak costs the defender in classical terms, because it makes the trade concrete. The peak sensitivity `M_s = max_ω |S|` controls guaranteed stability margins through the standard bounds `GM ≥ M_s/(M_s−1)` and `PM ≥ 2·arcsin(1/(2M_s))`. In the companion simulator, driving monitored-band suppression to the level that produces `M_s ≈ 3.3` collapses the guaranteed margins from roughly 31 dB and 58° at negligible suppression to about 3.1 dB and 17.5° **[R]**. The proxy optimizer is not merely relocating sensitivity; it is spending the system's robustness margins to buy monitored-band performance, and the expenditure is measurable in the oldest vocabulary control has. Section 5 shows this is not incidental — robustness is the currency on both sides of the trade.

---

5. What architecture can and cannot buy: an achievability result

Section 3 left a question open. The geometry gap between conserved amplification and realized loss is real for arbitrary profiles, but the counterexample there used functions no feedback loop realizes. Inside the manifold of *achievable* sensitivity profiles — those a stabilizing controller can present under engineering constraints — how much of the gap survives, and can a risk-aware controller do materially better than a risk-blind one at the same proxy performance?

We study this by convex exploratory synthesis. Because `|S(Q)|² = |1 − PQ|²` is a convex quadratic in the Youla parameter `Q` (here a finite-impulse-response filter) and CVaR is convex and monotone, minimizing accessible CVaR subject to a proxy-suppression floor, an effort bound, and a complementary-sensitivity peak bound is a convex program. We solve it by multistart local optimization on a 400-point frequency grid rather than a certified conic solver, so the values below are treated as accurate exploratory minima, not proven global optima. We work in discrete time so the relocation question of Section 4 is well-posed, and we read the resulting frontier as the shape of the achievable Youla profile set under finite-band engineering constraints rather than attributing it to the continuous-time Bode identity. (The discrete-time sensitivity integral, `∫₀^π ln|S| dω = π Σ ln|pᵢ|` over unstable poles, is zero for the stable loop used here; we do not lean the result on it.) The risk-blind reference is the proxy-greedy controller, which minimizes monitored-band cost alone.

Two findings emerge from a three-by-three sweep over adversary concentration (`c ∈ {1.5, 2, 5}`) and complementary-sensitivity allowance (`TP ∈ {1.5, 2, 3}`), and both are weaker and more defensible than the earlier draft of this section claimed.

First, the achievable minimum exploitability is a *value function* of the suppression demanded, and it is nondecreasing by construction: tightening the suppression floor shrinks the feasible controller set, so its minimum can only rise. This much is not an empirical discovery. What is empirical is that the rise is strict over the sampled range and how it scales — the minimum accessible CVaR climbs monotonically with suppression in every one of the nine configurations, more steeply for a more concentrated adversary. What the experiment does *not* establish, and an earlier version of this section wrongly asserted, is that every admissible controller is ordered this way. It shows the optimal floor rises; it does not order arbitrary pairs of controllers **[R]**.

Second, the effect of the robustness allowance is genuine but modest, and it survives only in a properly matched comparison. At a *fixed* suppression level, loosening the complementary-sensitivity cap lowers the achievable CVaR floor: at the deepest common suppression and the strongest

adversary the minimum falls from 1.317 at `TP = 1.5` to 1.244 at `TP = 2` to 1.217 at `TP = 3`, because a looser allowance enlarges the achievable profile set and lets the optimizer move amplification a little further from reach **[R]**. But this advantage disappears at the *maximum* attainable suppression. Comparing the risk-aware optimum against the proxy-greedy controller at the greedy's own suppression, the two coincide to within 10^{-6} in all nine cells: once the controller is pushed to the deepest suppression its class and caps permit, the effort and complementary-sensitivity constraints saturate, design freedom collapses, and there is nothing left to trade. The proxy-greedy design is already CVaR-optimal there.

The corrected reading is undramatic but sound. Deeper proxy suppression raises the exploitability floor, and this holds for the optimal controller, not merely a naive one. A looser robustness allowance buys some exposure back at a fixed suppression, but the currency is the complementary-sensitivity budget, and at maximal suppression that budget is spent and the risk-aware and risk-blind designs are the same controller. The governor's room to place vulnerability cleverly lives in the interior of the suppression–robustness trade and vanishes at its corner **[IP, under the transfer conditions of Section 6]**.

The sweep is shown in Figure 3.

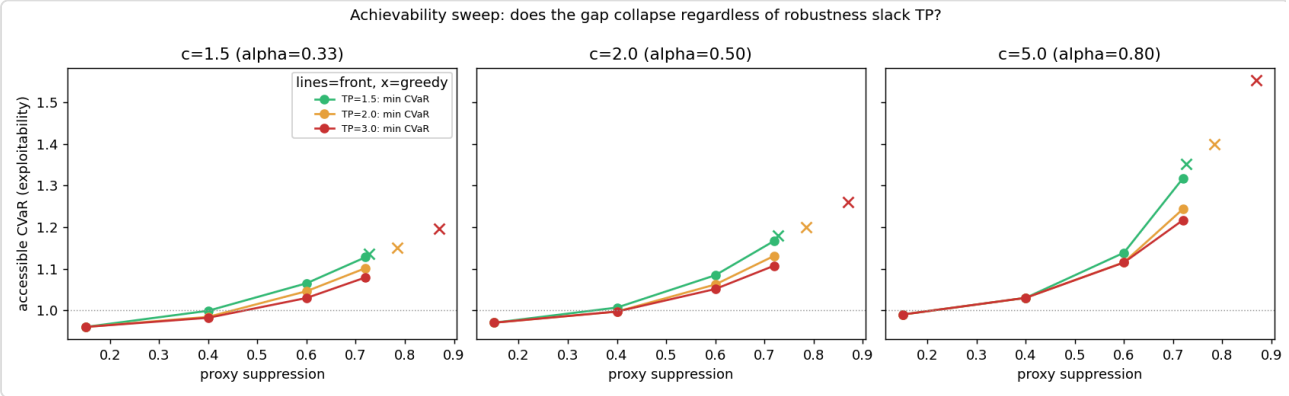

Figure 3. Achievability sweep. Each panel fixes adversary concentration `c`; curves give the minimum achievable accessible CVaR (a value function) against proxy suppression under complementary-sensitivity allowances `TP ∈ {1.5, 2, 3}`, crosses marking the risk-blind proxy-greedy controller at its own suppression. The floor rises monotonically with suppression in every cell; at fixed suppression a looser allowance sits lower; and each greedy cross lies on its own front — at maximal attainable suppression the risk-blind and risk-aware optima coincide (§5).

Stated plainly, the caveats: this is a single stable plant, one monitored band, an eight-tap controller, one family of caps, and an uncertified solver, so the magnitudes are specific to that setting and the optima are exploratory. What the sweep establishes is the *shape* — a value-function floor rising strictly with suppression, a modest matched robustness effect, and coincidence of the two designs at maximal suppression. Whether it survives a substantially different plant, in particular an unstable one where the Bode budget is strictly positive, is left to Section 7.

---

6. Governance transfer, in two separated layers

The results so far are theorems about a linear feedback loop. What, if anything, do they say about governance? The honest answer separates into two layers with very different transfer conditions, and the discipline of the section is to never let the stronger borrow the first's authority.

Layer A — the allocation logic, which transfers without Bode. The imported law of Section 2 is a statement about any constrained optimizer facing a fixed harm profile; it uses no conservation law. To export it, one identifies five objects in the governance domain: a location variable along which a strategic actor chooses where to act, a harm profile over those locations, the actor's reachable set, its concentration bound, and its budget — together with the discovery capacity of Section 4. Given these, and provided harm is linear and additive in the allocated density with costless static reallocation, the claim transfers as a static benchmark: strategic actors realize not the *average* harm over what they can reach but its *upper tail*, scaled by how densely they can concentrate and only to the extent they can discover it. Real actors face nonlinearity, congestion, switching costs, and endogenous response, so this is a benchmark model rather than a verbatim theorem about strategic behaviour in general. This formalizes the adversarial branch of the Goodhart taxonomy specifically — the regime where an agent responds to the metric and exploits residual structure — and supplies what the verbal statements of that literature lack: a budget, a licensing checklist, and a quantitative worst case **[IP]**. It does not require, and does not assert, that reform created the harm; it says only that whatever reachable, discoverable heterogeneity exists will be exploited to its tail rather than its mean.

Layer B — the manufacture claim, which does not transfer without a conserved burden. The stronger and more commonly intended claim is that suppressing the monitored channel does not merely expose pre-existing harm but *manufactures* it — that reform forces compensating vulnerability into existence. This is the Bode half, and it is exactly here that borrowed formal authority threatens. Bode's integral guarantees manufacture because it is a conservation law with proven licensing conditions. A governance system inherits that guarantee only if one can exhibit, empirically, an analogous conserved or lower-bounded quantity — a total burden that suppression in one place provably forces up elsewhere — together with the domain over which it is conserved, the mechanism by which the burden relocates, and the boundary beyond which conservation fails. The control case supplies all four for free; a governance domain supplies none of them by default. Absent an argument for all four, Layer B is unlicensed: to assert it is to dress a heuristic in the integral's authority, the precise move this series exists to refuse. We therefore state Layer B conditionally. Where a domain supports a conserved-burden argument, the manufacture claim and

the Section 5 trilemma — suppress the proxy hard, keep the system robust, avoid manufacturing exploitable exposure; one must give — transfer at **[IP]**. Where it does not, only Layer A transfers, and the honest sentence is the weaker "strategic actors exploit whatever reachable, discoverable heterogeneity exists," not "reform creates it." Paper XVIII's laundering mechanism is the closest the series has come to a domain instance of a Layer B burden, and even there the conservation is demonstrated in a model, not the field.

A licensed Layer B instance: public-service queues. Layer B is not merely a gate no case passes; at least one governance structure supplies a genuine conservation law of the required form. For a broad class of work-conserving multiclass single-server queues, Kleinrock's conservation law fixes a traffic-weighted sum of class waiting times, $\Sigma\ \rho_i\ E[W_i] = C$, with C independent of the scheduling discipline. Prioritising one class — lowering its mean wait — therefore forces weighted waiting onto the others; the burden is conserved and merely relocated. The mapping to governance is direct: the classes are urgent versus routine cases or permit types, the server is the shared judicial, clerical, inspection, or clinical capacity, the monitored proxy is the wait for a politically salient class, the conserved burden is traffic-weighted waiting time, the relocation mechanism is scheduling priority, and the model boundary is any change to capacity, arrivals, service requirements, admission, or the measurement clock. Courts have been modelled explicitly as resource-constrained case-management queues; and the documented gaming of English NHS waiting-time targets — holding patients in ambulances outside the emergency department, cancelling unmonitored procedures during measurement windows, reclassifying recorded times — is the relocation mechanism in the field, its most revealing move being not redistribution *within* the queue but export of patients *outside* the measured queue, the governance analogue of pushing amplification beyond the certified band **[IP]**. Audit and Stackelberg security allocation give a second family, combining both layers: a fixed inspection or protection budget lies in a resource simplex (Layer B), while an attacker chooses a reachable, discoverable, weakly-covered target (Layer A).

These instances also show what kind of claim Layer B is. It is a *subclass* result, not a general property of governance: it holds where a domain exhibits identifiable stock–flow, workload, material, or resource conservation, and it is an unlicensed heuristic everywhere else. The conserved objects differ by structure — traffic-weighted waiting time in a service queue, a coverage budget in audit allocation, a revenue identity in a fiscal system under fixed targets, physical mass in material-flow regulation — and each comes with its own escape boundary (added capacity, deterrence spillovers, growth, transformation or a shifted system boundary). The transfer discipline is to name the conserved quantity and its boundary before invoking manufacture, not after.

Two cautions close the section. First, high realized harm requires no strategic adversary at all. The premium `G` measures advantage over a *uniform* baseline; under a naturally colored environment — one whose disturbance incidence is already concentrated near the response peak — realized loss can be high while the strategic premium is modest, because the danger already lives in the base measure. Reading the model as "no adversary implies safety" inverts its content **[IP]**. Second, discovery is the governance-specific gate that most limits Layer A in practice, and it is itself a policy variable: a manufactured vulnerability is exploited only to the extent it is legible to the strategic population, so opacity can suppress exploitation of a peak that reach and concentration would otherwise find. This cuts both ways — legibility that helps a regulator can also arm an adversary — and neither direction is modeled here.

---

7. Limits and open problems

The most important limit is the scope of the achievability result. Section 5's sweep is a single stable plant, one monitored band, an eight-tap controller, and one family of caps; the claim it supports is the *shape* of the result — a monotone floor and a robustness-governed gap — not its magnitudes. The sharpest test is the unstable plant: when the loop has open-loop right-half-plane poles the Bode budget is strictly positive, `A_+ > A_-`, and the defender must amplify strictly more than it suppresses. Whether the robustness-governed collapse survives when the conserved total is positive rather than zero is the first extension we would run, and it is where the governance stakes are highest — the systems this transfer is aimed at, those with intrinsic positive-feedback dynamics such as wealth concentration or polarization, are exactly the unstable ones.

Three modeling assumptions bound the rest. The allocation lemma is static: it charges no cost for moving pressure between locations and grants the adversary perfect knowledge of the profile. A learning or co-adapting adversary breaks both, and in breaking them breaks the frame — once the disturbance allocation responds to the controller and the controller retunes in reply, the sensitivity function is no longer fixed, the single Bode integral is no longer the right invariant, and the CVaR lemma no longer describes the loss. That regime is a different paper; what this one establishes is the static baseline such a dynamics would depart from. Discovery, the third licensing capacity, is granted for free throughout; the natural relaxation — an adversary with noisy knowledge of `f` realizing strictly less than the CVaR, interpolating back toward the passive mean — is both tractable and the most decision-relevant addition for the governance reading.

Finally, the one technical question this paper flags but does not resolve. The concentration parameter `c` indexes a continuous family of performance functionals running from the accessible mean (`c = 1`, a band-limited H_2 quantity) to the accessible essential supremum (`c → ∞`, finite-frequency $H\infty$). The endpoints are classical; controller synthesis against the *intermediate* CVaR functional, with the interpolation parameter carrying the operational meaning of adversary capability, is a place we have not identified prior work on this exact frequency-indexed CVaR sensitivity objective — though CVaR-based and risk-sensitive LTI synthesis and mixed $H_2/H\infty$ design are mature, and absence from our search is not a proof of novelty. We have used the interpolant as a design objective in Section 5 but not studied the functional in its own right.

Appendix A. The allocation lemma

Let `Ω_a` be the reachable set with normalized base measure `μ_a`, and `f : Ω_a → [0,∞)` the response profile. A strategic actor with budget `E` allocates density `D = E·q`, `q ≥ 0`, `∫ q dμ_a = 1`, under a concentration cap `q ≤ c` (`c ≥ 1`); its loss is `J_strat = E ∫ f q dμ_a`. Then

```
sup { ∫ f q dμ_a : 0 ≤ q ≤ c, ∫ q dμ_a = 1 } = CVaR_{1−1/c}(f ; μ_a),
```

the mean of `f` over its highest-valued `1/c`-measure fraction. Equivalently this is the worst-case expectation of `f` over all probability measures whose density with respect to `μ_a` is bounded by `c` — the risk-envelope (dual) representation of conditional value-at-risk / expected shortfall, a distributionally-robust worst case over a likelihood-ratio-bounded ambiguity set. The optimizer is bang-bang: density `c` on the top `1/c`-fraction of `f`, zero elsewhere, with a fractional boundary cell. The proof is standard (Rockafellar–Uryasev); we import it. Endpoints: `c = 1` forces `q ≡ 1`, value `E_{μ_a}[f]`; `c → ∞` gives `ess sup f`.

Two qualifications travel with the lemma. It is **static** — loss is linear in the allocated density, with no temporal, causal, or transition cost — and it presumes **discovery**, since the supremum is attained only by a `q` supported on the true upper tail of `f`. Under a non-uniform base measure q_0 (a colored environment) the identical statement holds with `μ_a` replaced by the q_0-weighted measure and the cap read relative to q_0; the strategic premium is then measured against the q_0-baseline rather than the uniform one.

Appendix B. Model, simulators, and reproducibility

Plant and controllers. Continuous-time plant `P(s) = 1/((s+1)(s+2))` — stable, minimum-phase, relative degree two. The analysis stage (§2–§4) uses a resonant controller `K(s) = kp + δ·s/(s² + (ωm/Q)s + ωm²)`, `kp = 0.5`, `ωm = 3`, `Q = 2`, with `δ` the proxy-gain parameter and `Ω_m` a neighborhood of `ωm`. The synthesis stage (§5) discretizes `P` by zero-order hold at `Ts = 0.1` and parameterizes stabilizing controllers by a length-8 FIR Youla parameter `Q(z)`, so `S = 1 − PQ` is affine in the coefficients; caps `EQ` (effort) and `TP` (complementary-sensitivity peak). Seed 20260716 throughout. Simulators: `paper_xxv_simulator_bode_adversary.py` (§2–§4), `paper_xxv_simulator_bode_synthesis.py` (§5).

Bode anchor (R0). The integral is analytically zero (Freudenberg–Looze, relative degree two). The high-frequency expansion `log|S| = 1/(2ω²) + 693/(8ω⁴) + O(ω⁻⁶)` gives a truncated value `B_S(0,W) = −kp/W − 231/(8W³) + O(W⁻⁵)`; the residual after the leading tail matches the next analytic coefficient:

W	B_S + kp/W	−28.875/W³
100	−2.888e−5	−2.888e−5
300	−1.069e−6	−1.069e−6
1000	−2.888e−8	−2.887e−8
3000	−1.070e−9	−1.069e−9

Functional non-identifiability (§3). Two profiles on a unit-measure band: `(|S|² = 4` on measure `0.25`, `1` on `0.75)` and `(|S|² = 2` on `0.5`, `1` on `0.5)`. Both have accessible positive log-area `A_+ = 0.1733`; under `c = 2` their strategic losses are `2.50` and `2.00`. These are arbitrary measurable profiles, not achievable sensitivity functions.

Achievability sweep (§5). Minimum accessible CVaR obtained by multistart SLSQP over the FIR coefficients on a 400-point grid (convex program, uncertified solver); the risk-blind reference is the proxy-greedy controller (minimum monitored-band cost). Two comparisons matter. *Matched* — evaluating the risk-blind and risk-aware optima at the proxy-greedy controller's own suppression — gives a gap that vanishes to numerical precision in all nine cells, so at maximal attainable suppression the two designs coincide:

c	TP	greedy suppression	greedy CVaR	matched gap
1.5	1.5	0.727	1.137	3.5e−7
2.0	1.5	0.727	1.181	6.8e−7
5.0	1.5	0.727	1.352	2.2e−6
1.5	2.0	0.785	1.151	2.9e−7
2.0	2.0	0.785	1.199	5.1e−7
5.0	2.0	0.785	1.399	2.1e−6
1.5	3.0	0.870	1.197	6.0e−7
2.0	3.0	0.870	1.261	9.1e−7
5.0	3.0	0.870	1.554	2.6e−6

The genuine robustness effect appears only in the *fixed-suppression* comparison: at suppression 0.72, the minimum achievable CVaR falls as the complementary-sensitivity allowance `TP` loosens.

c	TP = 1.5	TP = 2.0	TP = 3.0
1.5	1.128	1.102	1.080
2.0	1.167	1.131	1.108
5.0	1.317	1.244	1.217

The value-function floor is monotone in suppression in all nine cells (nondecreasing by construction; strict here). An earlier version of this appendix reported a "matched gap" comparing controllers at unequal suppression levels; that comparison was mismatched and its "robustness reopens the gap" reading has been withdrawn.

Verification. Strategic loss is computed two independent ways — an operational greedy water-filling allocator and the analytic CVaR formula — agreeing to `~1e-14`; this checks the discretization, the theorem being imported.