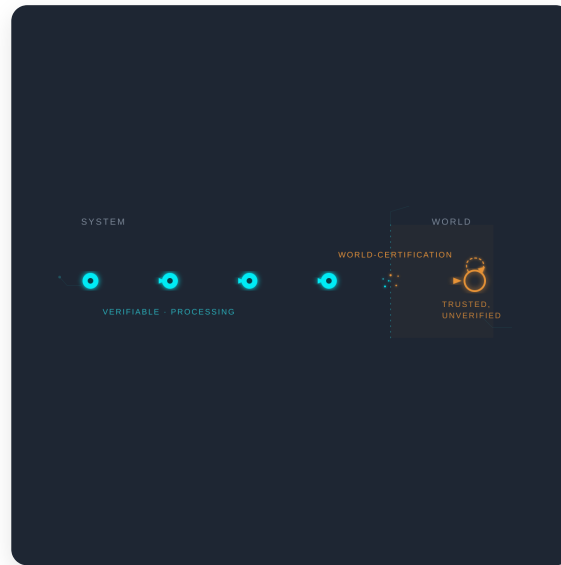




The Certification Floor

Why world-coupled coordination retains an irreducible link, and self-referential coordination does not

Paper XVII in the Governance as Engineering series

Isolates the structural reason that an adaptive coordination boundary retains an irreducible certification link. Derives a relocation invariant: automating a coordination boundary relocates the irreducible world-certification link upstream but never removes it. Paper XVII in the Governance as Engineering series.

Björn Kenneth Holmström

July 2026

Creative Commons Attribution-ShareAlike 4.0 International

<https://bjorkennethholmstrom.org/working-papers/certification-floor>

Abstract

Paper XVI established *source-term locality*: a quantity representing unused alternatives decays under optimization and persists only through a source term whose position — inside or outside the optimizer's control set — determines whether collapse is robust or contingent. This paper isolates the structural reason such a source term can be irreducible.

The core object is an asymmetry between two kinds of dependency. **Processing** — whether a system computes what it claims — can in principle be made arbitrarily verifiable. **Certification of reality** — whether the external fact a rule depends on actually obtained — cannot be made self-verifying, because a verifier that certifies a world-fact needs a certifier in turn, and the regress terminates only by accepting one anchor as trusted-unverified. Chain length (the number of editable links between an institution and its nearest fixed invariant) is a coordinate on this asymmetry, not a separate thing.

From the asymmetry follows a *relocation invariant*: automating a coordination boundary relocates its irreducible certification link upstream but does not remove it. The invariant is **scope-bounded**. It holds for *world-coupled coordination* — coordination that must answer to some fact outside itself — and demonstrably fails for *self-referential coordination* (pure convention), where the rule and the practice are the same thing and there is no external fact to certify. Governance is world-coupled by construction, so the invariant applies to it; pure convention escapes it.

The result is not a theory of institutional persistence and not a universal law of adaptive systems. It is one bounded structural regularity, its domain of validity stated, its evidence base and its limits recorded honestly. The evidence is concept-isomorphism across independently-formalized disciplines; it is strong on "these fields formalize the same shape" and weak on "the shape is true of the world," and the paper claims no more.

1. The tradeoff: chain length against adaptivity

Paper XVI asked what makes a boundary durable against an optimizer that gains by moving it. Begin with a distinction the answer turns on.

A coordination boundary can enforce a rule in two structurally different ways. It can be **frozen into a physical invariant** — a mountain range that makes invasion from one direction impossible enforces "no invasion here" with no sensor, no custodian, and no interface to suborn. Or it can be **maintained through representation** — a rule that some agent senses is being violated, verifies against a standard, and acts on.

Name the first regime **$L_c = 0$** : zero editable links between the enforced constraint and the physical fact that enforces it. It buys *absolute enforcement* at the cost of *zero adaptivity*. The mountain enforces its rule perfectly and forever, and is helpless the moment the environment routes around it — the instant someone invents the airplane, the constraint is void, and it cannot update, because it has no representation of its own rule to revise. It is a frozen Execute phase with no Sense and no Learn.

Each editable link added moves the system to **$L_c \geq 1$** and buys adaptivity — the capacity to detect violation, re-verify, and change the rule — at the cost of one more subornable interface. This yields the first result **[IP]**:

*The minimum chain length of any adaptive coordination boundary is **$L_c \geq 1$** , and this is the price of adaptivity, not a definitional fact about governance.*

The distinction matters because an earlier route to this claim was circular. "Governance coordinates agency; agency implies representation; therefore $L_c \geq 1$ " wins by defining governance to exclude the $L_c = 0$ cases (terrain, thermodynamic gradients, pure physical constraint), which do coordinate agency against counterfactuals with no representation. The non-circular route is the tradeoff: $L_c = 0$ coordination exists, but it cannot represent its own rule and therefore cannot adapt it. Governance is $L_c \geq 1$ because governance is the kind of coordination that must update its rules — and updating requires representation, and representation is editable.

This folds directly into Paper XVI's coupling rather than standing apart from it. Source-term locality concerned the capacity to *generate alternatives*; chain length concerns the capacity to *update the rule*; both are one coupling — **the property that makes a system adaptive (a source term inside the loop, a representation of its own rule) is the same property that makes it editable, erodible, or subornable.**

The in-between of that coupling — a rule editable in principle but costly enough to edit that ordinary pressure leaves it standing — is the protection class Paper XVI §6 leaves open and probes computationally. There the irreversible trap opens only when exit from the incumbent rule is nearly foreclosed, so a merely-costly $L_c \geq 1$ link is metastable, not absorbing: it decays back to contestability once the pressure lifts. Editability is therefore not binary. The certification floor fixes what cannot be removed at all; whether a removable-in-principle link behaves as removable depends on how close its switching cost sits to foreclosure. **[IP]**

$L_c = 0$ is the limiting case: un-editability purchased by surrendering adaptivity entirely, the geographic equivalent of a fully concentrated posterior that can no longer learn.

2. The deeper object: the certification / processing asymmetry

Chain length is a convenient measure, but it is a coordinate on something more fundamental, and the result is clearer with that object as the headline:

Processing can be made arbitrarily verifiable; certification of reality cannot be made self-verifying.

A processing step — does this code compute what it claims? — can in principle be checked, formally verified, attested. A certification step — did the external fact this rule depends on actually obtain? — cannot be verified without an external certifier, whose certification in turn requires a certifier, a regress that terminates only by accepting one anchor as trusted-unverified. The reason L_c cannot reach zero for adaptive coordination is that at least one of its links is always a certification step, not a processing step. Chain length is the measure; the certification asymmetry is the mechanism.

The same asymmetry surfaces wherever the question is asked in native terms: control theory (you verify the controller, but must trust the sensor still measures reality), distributed systems (the root of trust), economics (contractual incompleteness), science (experimental observation), law (witness credibility), machine learning (grounding). That breadth is why the object is worth isolating — and it is also the temptation the paper must refuse: **that the same asymmetry appears in six fields is evidence it is a real structural feature, not license to promote it into a law that explains all six.** The discipline of Paper XVI applies to this paper's own central object.

3. The relocation invariant

The strongest candidate for driving an adaptive boundary to $L_c = 0$ is automation: a system that updates its own rule with no human hand at the moment of execution. Consider the sharpest case — an immutable smart contract that executes an on-chain vote, or a "governance-minimized" DAO. The code is deterministic; no custodian touches the lever at execution.

Automation does not reach $L_c = 0$. It **relocates the irreducible link upstream rather than removing it**. The immutable contract closes the *execution* link but reopens the same vulnerability at the *specification* link: what counts as a valid vote (tokens standing for stakeholders), what a proposal means (its text standing for its intended effect), what quorum represents (a threshold standing for legitimate consensus). Attacking such a system requires touching none of the immutable code — only acquiring the representation (tokens), or exploiting the gap between a proposal's read meaning and its bytecode effect. That gap is the chain. This gives the invariant **[IP]**:

***The relocation invariant.** Automating a link relocates the irreducible representational dependency; it does not delete it. Representation can be made deterministic (code executes exactly as written) but not interpretation-free (what the code should say, and whether the tokens represent the right parties, remains editable).*

A note on naming. This regularity is not called a conservation law. A conservation law derives from a symmetry; none has been derived here. What is observed is that every attempt to remove a certification interface relocates it — an invariance under transformation, an empirical regularity across formalisms. "Relocation invariant" claims exactly that and no more.

4. The floor mechanism: irreducible world-certification

Why can the link not reach zero? The answer is not "representation is irreducible," which restates rather than explains. It comes from two disciplines that share no vocabulary and were asked the question in isolation.

In **trusted computing**, the trusted computing base shrinks under verification but floors at the root of trust plus the attestation verifier. Zero would require a self-verifying root — which introduces a Byzantine fault in the verification path unless some primitive is accepted as non-verifiable. The floor exists because verification cannot ground itself: the verifier needs a verifier, and the regress terminates only by trusting one primitive unverified.

In **contract theory**, the residual requiring ex-post interpretation falls with completeness but floors at facts not verifiable to third parties. The floor is the gap between the rule and whoever certifies that the rule's triggering fact obtained — a certification that cannot be folded into the contract.

These are one structure in two costumes. The root of trust and the third-party fact-certifier are the same object:

***The irreducible link is always a world-certification link** — the point where an external fact must enter the system and be trusted, because certification cannot self-ground. Automation can relocate which fact must be certified, and can raise the cost of suborning the certifier, but cannot eliminate the certification step.*

Automation cannot remove this link because it is not a processing step (verifiable) but a world-certification step (not verifiable without an external certifier, ad infinitum). This also closes a loop with Paper XVI's residual: the floor is, in each field, the system's inability to certify — or even enumerate — the part of the world it has not modeled. Decision theory's catch-all hypothesis, the contract's non-verifiable contingency, and the TCB's unverifiable root are one object seen three ways.

5. Not all irreducible links are equal: discrete vs. ambient

The design consequences turn on a second-order distinction. An $L_c = 1$ link can be **discrete** — a specific vote, a specific signing key, a specific attestation — or **ambient** — a standing condition present continuously rather than at identifiable decision points. Proof-of-work illustrates the latter: its irreducible dependency is not a single decision but the continuous social consensus that a given ledger *is* the canonical one, a fact reasserted every block.

The distinction matters because a discrete link can be isolated, audited, and hardened; an ambient link cannot. You can rotate a key and log its use; you cannot rotate or audit "what the community takes the ledger to mean." Part of the engineering problem, then, is not only to minimize chain length but to force the irreducible link to be *discrete* rather than ambient — to concentrate the unavoidable trust into an accountable act rather than diffusing it into an un-isolable standing condition.

6. The scope bound: world-coupled vs. self-referential coordination

The invariant is not universal, and establishing its boundary is as important as establishing the invariant. A counterexample probe asked two further disciplines — the theory of linguistic convention and the theory of coordination equilibria — whether a self-sustaining convention must refer to anything outside itself.

The answer bounds the claim. A convention can be **self-sustaining with no external referent**: slang shifting meaning by usage alone; a repeated coordination equilibrium held by mutual expectation alone. Neither certifies any fact about the world. But the certification link *reappears* precisely when the convention must connect to something outside itself — a world-fact its meaning must track ("water" must track H₂O), or a selection among multiple equilibria resolved by a focal point that lives outside the strategy profile. Two vocabularies, one boundary:

Self-referential coordination — where the rule and the practice are the same thing and there is no external fact to certify — is exempt from the relocation invariant.
World-coupled coordination — which must answer to some fact outside itself — carries the irreducible world-certification link. The invariant was never about coordination as such; it is about coordination that answers to something outside itself.

Governance is world-coupled by construction: it certifies breaches, eligibility, harms, entitlements — facts about the world. That is why the invariant bites on governance and does not bite on pure convention. The naïve universal — that *every* adaptive coordination architecture contains an irreducible certification link — is false, and stating its falsity is what makes the bounded claim trustworthy.

One seam is left open [H]. Two lenses independently raised a temporal crack in the exemption: a self-referential convention may be certification-free *statically* yet require external coupling *dynamically* — to resist drift, to stay salient across generational turnover — because the community must observe *that it still coheres*. This is the difference between needing no external referent to *exist* and possibly needing one to *persist*. The field is split; this paper does not resolve it, and resolving it toward "so it needs certification after all" would be the convenient move that quietly restores the universal. The evidence does not support that move.

7. Design implication

The result yields an engineering target rather than a counsel of despair. Because the world-certification link cannot be removed from world-coupled coordination, the design problem is not "eliminate trust" (impossible) nor "rebuild the boundary faster than optimizers erode it" (eternal expenditure), but:

*Minimize the chain to a single **discrete** world-certification link, and raise the cost of suborning that certifier toward thermodynamic levels.*

Three levers follow. **Shorten the chain** — reduce the number of editable links between the world-fact and the enforced rule. **Discretize the irreducible link** — force the unavoidable trust into an accountable, auditable act rather than an ambient standing condition (§5). **Raise the edit-cost** of suborning the certifier: rotation, multi-witness attestation, audit trails. A high-enough edit-cost is, for practical purposes, indistinguishable from a floor.

A caution that falls directly out of §4: hardening the *record* is not hardening the *certification*. An immutable ledger closes the execution link and leaves the world-fact entering the system exactly as soft as before. The most common design error this predicts is investing in tamper-proof storage while the input — the human or sensor certifying that the recorded fact obtained — remains the softest point in the system. The record is a processing artifact; the input is a certification. Only the second is the floor.

8. What this does not show

The discipline of the result is in its refusals.

It is **not** a general theory of institutional persistence — that is the high-reach unifying object Paper XVI exists to distrust. It is **not** a universal law of adaptive systems; §6 supplies the counterexample and the boundary. It does **not** define a measurable index; the certification floor is, in the cases that matter most, not cleanly observable. It does **not** claim the physical invariants themselves are editable — factoring hardness does not care about politics — but rather that no institution can *inherit* their fixity directly, because it accesses them only through a certification step that cannot self-ground. And it does **not** license "reach," "epistemic adaptivity," or a "contestability index" as primitives; each was tried during the investigation and each reintroduced a hidden global variable.

The permitted claim is one bounded regularity with its domain stated: within world-coupled coordination, adaptive boundaries retain an irreducible, relocatable-but-not-eliminable world-certification link, and the design lever is to minimize, discretize, and cost-harden it.

9. Method and confidence

The result was produced by an adversarial multi-model protocol: a structural hypothesis routed through several frontier language models used as deliberately decorrelated observers, with a human executing prompts and making structural decisions, and one model integrating. The method's discipline, and its documented failure, are part of the evidence.

The hypothesis narrowed under pressure across several rounds — from a universal coupling, through level-relative and process-relative forms, to the scope-bounded invariant stated here. That graceful degradation is itself evidence: a claim held contestable by genuinely decorrelated observers decays toward something real, and the decay works only while the observers do not share a frame. The protocol also *failed* mid-investigation — the observers converged, one model hallucinated a referent, and a decisive test was briefly resolved using the integrator's own vocabulary, which cannot distinguish a real invariant from a sufficiently flexible language. The response was to route the claim to observers decorrelated *by vocabulary*: disciplines asked the question in native terms, hypothesis withheld, cross-cutting vocabulary banned. Those probes reproduced the structure (§4) and bounded it (§6).

A literature check against the primary sources (Hart–Moore on contractual incompleteness; Lampson on trusted computing) was attempted through research-mode retrieval. It strengthened the contract-theory leg to secondary-literature consensus — non-verifiability to third parties is a named driver of incompleteness — and *weakened* one attribution: the accessible text of Lampson defines the trusted base as what security *depends on*, not as unverifiable-from-inside, so the trusted-computing floor is here credited to the verification-regress argument (logical) and to probe corroboration, not to a primary-source assertion. No primary text was reached; the [R] gate remains open on full-text reading by a reader with database access.

Confidence labels are reserved for the points where miscalibration would cost most:

- **[IP]** — the chain-length/adaptivity tradeoff (§1); the relocation invariant and the certification/processing asymmetry (§2–§4); the world-coupled/self-referential scope bound (§6). These rest on concept-isomorphism across independently-formalized disciplines and hypothesis-blind probes — strong on shared shape, weak on world-truth, because the observers remain correlated language models.
- **[H]** — the temporal seam (§6): whether self-referential coordination needs external coupling to *persist* though not to *exist*. The field is split; preserved, not resolved.

- **[R]** — none. Promotion requires primary-text reading (Hart–Moore 1988; Lampson §6) not yet performed. The correlated-observer caveat governs the whole result: it establishes that several disciplines formalize the same structure, not that the structure is confirmed against the world.

The most durable contribution the series can make is not a single overarching architecture but a collection of modest, well-scoped regularities connected by a disciplined method. This paper is one such regularity, offered at the confidence its evidence licenses and no higher.