Fixed Points, Nuclei, and the Convention Layer: Order-Theoretic Coherence for Agent–Tool Composition

John Komkov, Independent Researcher

July 2026

Abstract

Empirical work on agent–tool composition found that an LLM agent acts as its own convention translator: value corruption survives only on opaque passthroughs (paths, identifiers, cursors, markup), the failure unit is caller–tool rather than tool–tool, and convention agreement is a fact about a partial order of refinements, not a vector space. We develop the mathematics this relocation demands, in four parts. (1) A star-complement localization theorem: for an agent-star with opaque seams, the coboundary's cokernel is carried entirely by the opaque edges, with a unique opaque-supported representative per tension class. (2) Convergence theorems for refuse-and-cure protocols: inflationary rounds reach a fixed point within lattice height; monotone rounds reach the least fixed point; descent bounds rounds by the initial fee; gate PROCEED regions are up-sets. (3) A nucleus-theoretic account of disclosure: every nucleus on a complete Boolean algebra is closed (Johnstone's classification, formalized); sufficient disclosure sets of the witness matroid have minimal elements exactly the bases, admit a least element iff every element is a loop or coloop, and intersect in the coloop set. (4) A complete resolution of the degree-1 Hodge–Tarski conjecture of Riess: refuted as stated on a pendant edge; shown ill-posed for arbitrary orientations; repaired and proven on cyclically-oriented cycles for arbitrary complete-lattice stalks and arbitrary Galois connections; the natural "balanced orientation" repair refuted by a bowtie counterexample; and the exact class characterized: equality for all sheaves holds precisely on orientations with in-degree = out-degree = 1 at every non-isolated vertex, machine-checked as an iff on self-loop-free quivers. Every theorem in the development — the refutation, the cyclic repair, the class direction on arbitrary in-out-one quivers, and the characterization — is Lean-verified against a pinned Mathlib with statements pre-registered before proof; the orientation sweeps are machine-demonstrated within stated bounds; and the code-level identifications are property-tested on deterministic adapters against the shipped implementation.

Contents

1 1. Introduction: the relocation

Three empirical findings about LLM agent–tool composition motivate this paper. First, a pre-registered density test found that realistic multi-tool workflows corrupt values far more rarely than all-pairs schema analysis predicts, because the agent's reasoning layer re-generates every value it understands; corruption survives only on values the agent pipes verbatim — paths, identifiers, cursors, markup. Second, the recurring failures observed in deployed tooling are caller–tool, not tool–tool: the question that matters is whether the value reaching a tool honors that tool's convention. Third, the false positives that discredit schema-level analysis are order-theoretic facts: two tools that both use 1-based pagination agree, and no vector-space model of the seam can see it.

Together these relocate the mathematics. The correct base space is the caller-star — the agent at the center, tools at the leaves, plus the sparse opaque seams the agent does not mediate. The correct coefficients are refinement lattices — partial orders of convention knowledge, where meet is intersection of possibilities and bottom is contradiction. The correct dynamics are fixed-point — the shipped repair and refuse-and-cure protocols are monotone operators on disclosure states. And the correct notion of "fee" is a localization — the minimal forgetting that restores coherence — rather than a dimension count.

The prior, abelian layer of this program (the coherence fee as a rank gap of a totally unimodular coboundary; the witness matroid; witness geometry) is retained as the recall engine it was shown to be. This paper builds the order-theoretic layer on top of it: Section 2 explains, with a theorem, why the abelian layer kept reporting that obstructions were local; Sections 3 and 4 give the fixed-point and nucleus-theoretic accounts of cure and disclosure; Section 5 reports a pre-registered model-adequacy experiment separating the lattice model from the value-blind fee; Section 6 resolves an open conjecture of Riess's lattice-sheaf program, in both directions; Section 7 records a tropical postscript. Every theorem was stated and pre-registered before any proof was attempted; proofs were produced by the Aristotle prover against a pinned Mathlib and re-verified by local rebuild; statements were never modified after registration. The canonical record — theorem names, run stamps, axiom audits — is the package `papers/tarski-coherence/lean/` and its ledger, and we cite it rather than restate it.

A remark on honesty conventions, used throughout: verified means a sorry-free machine proof of the stated abstraction; demonstrated means an executed, pre-registered check on real artifacts; conjectured means exactly that; and the identifications between theorems and shipped code are labeled separately (Section 9), because a theorem about an abstraction never verifies an implementation.

2 2. Star-complement localization

Model an agent-mediated composition as a graph: a center vertex (the agent), one vertex per tool, one star edge per tool, and a finite set O of opaque edges between tools — the passthrough seams the agent does not re-generate. Over $\mathbb{Q}$, take the standard coboundary $\delta : C^0 \to C^1$ of this graph, and let $Z_{op} \subseteq C^1$ be the subspace of cochains vanishing on every star edge.

Theorem (star-complement localization; Lean: `StarComplement`). $\ker \delta$ is the constants; $C^1 = \operatorname{im} \delta \oplus Z_{op}$; the cokernel of δ has dimension exactly $|O|$; and every tension class has a unique opaque-supported representative.

The reading is the point. Star seams carry no cohomology: whatever tension a composition has lives, after a canonical change of representative, entirely on the opaque edges. This is the theorem behind both banked empirical facts. The density finding — corruption survives only on opaque passthroughs — is the localization read operationally. And the earlier program's persistent observation that empirical H^1 was "100% triangle-generated" is its homological dual: the fundamental cycles of a star-spanned graph are exactly the triangles tool–agent–tool, so obstructions supported on opaque edges appear as triangle classes. The mathematics is elementary (fundamental cycles with respect to a spanning tree); the application and the formalization are new.

3 3. Cure-loop convergence

The shipped repair protocol iterates rounds: probe undisclosed conventions, confirm some, disclose them, recompute the fee. Abstract the state as the set of disclosed items and each round as a map f on subsets; write φ for the fee. Four theorems, in increasing strength of hypothesis (Lean: `CureLoop`):

1. Inflationary termination. If rounds only add disclosures ($s \subseteq f(s)$), the orbit from $\emptyset$ is monotone and reaches a fixed point within $|A|$ rounds.
2. Least fixed point. If the round map is monotone, the orbit from $\emptyset$ lands on $\operatorname{lfp}(f)$ — the minimal disclosure closure, not merely some closure. This is a protocol-design theorem: it says what a monotone cure loop buys you. Whether the shipped greedy round is monotone is deliberately not claimed (its choice step is order-sensitive), so as applied this theorem is labeled conditional.
3. Fee descent. If every non-fixed round strictly decreases φ, a fixed point is reached within $\varphi(\emptyset)$ rounds — the honest formalization of the implementation's documented bound.

4. PROCEED regions are up-sets. If φ is antitone (disclosing more never raises the fee — the shipped submodularity), then $\{s \mid \varphi(s) \leq \theta\}$ is upward closed: once a gate would proceed, further disclosure never revokes it.

Section 9 reports the property tests that demonstrate the shipped implementation satisfies these hypotheses on real runs: rounds are inflationary, confirming rounds strictly descend, termination lands within the initial fee, and the loop's exit state is the fee-zero closure.

4 4. Nuclei and disclosure

A nucleus on a lattice is an inflationary, idempotent, meet-preserving operator — the order-theoretic form of "coherently forgetting distinctions." The natural first guess for connecting nuclei to the program's disclosure calculus — "the least nucleus whose sheafification admits a global section is the minimum disclosure set" — is ill-typed (order-least versus cardinality-minimum) and generically false. The true statements are sharper (Lean: `BooleanNuclei`, `DisclosureNuclei`).

Classification (Johnstone, Stone Spaces II.2, formalized). On a complete Boolean algebra B, every nucleus is closed: $j(x) = j(\bot) \vee x$. The nuclei of B are order-isomorphic to B itself. The pinned Mathlib has the `Nucleus` API but not this classification; the formalization fills that gap and is a candidate for upstreaming.

Disclosure as spanning. Say D is sufficient for the witness matroid M when the ground set lies in the closure of D: disclosing D drives the fee to zero. Then: sufficiency is an up-set; the minimal sufficient sets are exactly the bases of M — plural in general, which is precisely why "the" minimum disclosure set was a misnomer; a least sufficient set exists iff every ground element is a loop or a coloop, i.e. iff every leverage score is 0 or 1 in the witness geometry; and the intersection of all sufficient sets is the coloop set — the must-disclose core every cure shares. Through the classification, sufficient closed nuclei on the Boolean convention algebra correspond to sufficient sets, and their meet corresponds to the coloop core: the least common forgetting is the must-disclose core. This is the honest repair of the naive guess, and the least-element criterion is, to our knowledge, new.

5 5. The model-adequacy gate

The lattice model claims two structural advantages over the value-blind fee: agreement cannot false-positive (equal conventions meet without contradiction) and never-composed seams cannot flag (the base space contains only real handoffs). A pre-registered experiment tested whether an honest encoding delivers these on the program's own recorded corpus: six realistic workflows (two with real corruptions) plus four wild caller–tool issues from public trackers. The rubric — powerset stalks over fixed convention sets; mediation as a property of the handoff (verbatim-piped versus agent-regenerated); an evidence hierarchy with citation-per-assignment — was frozen before any encoding; the encoder was an agent that saw only the rubric and public documentation; the evaluator computes the Kleene least fixed point of meet-propagation over bitmask lattices.

All five pre-registered clauses passed with zero per-case tuning: the bottom-support equaled exactly the two real corruptions; the known agreement pair did not flag; all four wild issues flagged; and both negative controls (mediation-scramble sensitivity, stalk-swap discriminability) behaved. The separation is quantified: the value-blind fee flags all ten handoffs of the corpus; the lattice model flags exactly the six real ones, and the fee's four false alarms are precisely the agent-translated handoffs. Scope, honestly: this is a constructed corpus — authored encodings of previously observed traces — so it is a model-adequacy result about the encoding discipline, not new evidence about the world; one wild issue flags by assignment rather than by intersection (the deployed endpoint rejected even the documented-correct value), recorded verbatim in the artifact.

6 6. The degree-1 Hodge–Tarski conjecture, resolved

Riess's thesis develops sheaves of complete lattices on graphs with a Galois connection per incidence, proves the degree-0 Hodge–Tarski theorem — the suffix points of the Tarski Laplacian are exactly H^0 — and defines the degree-1 apparatus: cochains on unordered edges; coboundaries δ_-, δ_+ from an orientation, a choice of signed endpoints per edge; H^1 as the coequalizer of the pair in Sup, realized (by the thesis's own construction) as Q^{op} with $Q = \{y \mid \delta_-^*(y) = \delta_+^*(y)\}$; and the Tarski Helmholtzian $\mathbb{H}$, whose value on an edge joins the pulled-and-pushed values of the neighboring edges at both endpoints. The thesis then poses, explicitly as open: do the prefix points of $\mathbb{H}$ coincide with H^1?

The resolution has five parts, each machine-checked or machine-evidenced.

Refuted as stated. On the single edge K_2 with two-element stalks and identity connections, both neighbor joins are empty, so $\mathbb{H}$ is constantly $\bot$ and every cochain is a prefix point; meanwhile Q's condition, computed nodewise with the thesis's own empty-meet convention, top-forces both endpoints, so $Q = \{\top\}$. Two elements against one: not equal, not even equinumerous, so both the set reading and the isomorphism reading of "coincide" fail. The refutation is robust to the nearest repair — adding the same-edge composites, the formula appearing in the thesis's own commented-out proof attempt — because those composites are below the identity for every Galois connection. (Lean: `TarskiHarmonicDeg1`, seven theorems.)

Ill-posed for orientations. $\mathbb{H}$ is manifestly orientation-independent; Q is not. On the triangle C_3 with two-element stalks and identities, the cyclic orientation gives $Q = \{000, 111\}$ while any source/sink orientation gives $Q = \{111\}$ — the cardinality changes, so no convention of reading "coincide" up to isomorphism can rescue orientation-independence. The thesis introduces orientation as a choice and nowhere claims independence; the conjecture simply inherits a dependence its left-hand side does not have.

Repaired and proven. On a cyclically-oriented cycle C_n — every node one edge in, one edge out — the conjecture is a theorem, in full generality:

Theorem (cyclic Hodge–Tarski; Lean: `CyclicHodgeTarski`). For arbitrary complete-lattice stalks and arbitrary Galois connections on a cyclically-oriented C_n, $\mathrm{Pre}(\mathbb{H}) = Q$.

The proof is pure adjunction algebra: each joinand inequality of $\mathbb{H}y \leq y$ transposes through the Galois law into a one-sided comparison of upper-adjoint pulls at a node; the two families pair at the same node; antisymmetry yields exactly Q's matching condition, and the converse transposes back. The same transposition gives an orientation-free description on any graph: the prefix points are the cochains whose upper-adjoint pulls agree at every node of degree at least two — degree-one nodes impose nothing, which is precisely the pendant-edge refutation seen from above. Fidelity note: the thesis Helmholtzian indexes neighbors, so the two-term cyclic form requires $n \geq 3$; the digon is excluded and says so in the formalization.

The wrong repair, refuted. "Pendant-free" is necessary but not the right hypothesis, and neither is the natural strengthening "balanced" (in-degree equal to out-degree). On the balanced bowtie — two directed triangles sharing a hub with in-degree and out-degree two — Q constrains only the meet of the pulls at the hub while the prefix condition constrains each pull, and meets do not split: Q has four elements against $\mathrm{Pre}(\mathbb{H})$'s two. Note the failure direction flips — pendant edges give $\mathrm{Pre} \supsetneq Q$ by top-forcing, the bowtie gives $Q \supsetneq \mathrm{Pre}$ by meet-collapse — and a machine harness records this signed direction as a mechanism diagnostic.

The exact class, characterized (Lean: `QuiverHodgeTarskiClass`, `QuiverHodgeTarskiGadget`). The theorem: on a self-loop-free quiver, $\mathrm{Pre}(\mathbb{H}) = Q$ holds for all complete-lattice stalks and all Galois connections if and only if every non-isolated vertex has in-degree = out-degree = 1 — i.e. the orientation is a disjoint union of directed cycles (in a simple graph necessarily of length at least three, so the digon caveat is vacuous for the class). The class direction is proved with no cycle decomposition: in a cast-free incidence formulation, the in/out subtypes at each vertex are singletons, the meets and joins collapse, and the cyclic transposition pairs the unique in-edge with the unique out-edge locally at every vertex — so the theorem strictly exceeds the disjoint-cycles phrasing. The converse is a two-family gadget: uniform Prop stalks with connections killed off the bad hub localize the comparison, where sources and sinks separate by $\top$-forcing (the all-$\bot$ cochain is a prefix point that fails Q) and hubs with a doubled side separate by meet-collapse (a formula cochain lies in Q but not in the prefix set). No-self-loops is load-bearing, not hygienic: a self-loop vertex satisfies the degree conditions with the same edge while the Helmholtzian sees no other incidences there. The $\forall$-sheaf side of the iff is stated at `Type 0` (which contains `Prop`); the class direction is universe-polymorphic.

The empirical trail that preceded the proof is retained as the method's receipt. A seeded, control-hardened sweep (six frozen oracle fixtures, the proven degree-0 theorem as a positive control on every instance, an adjunction-law check on every enumerated connection, a cross-derivation of δ_-^* against the generic adjoint) evaluated 2.7 million instance–orientation rows with zero in-class failures, and a correctly-quantified aggregation confirmed the converse half exhaustively at two-element stalks — before the gadget was formalized. One methodological note is recorded rather than hidden: the pre-registered outcome surface initially tested the per-row existential reading, which accidental equalities under degenerate connections satisfy anywhere; the claim is universal in the sheaf, and the surface was re-aggregated at the correct quantifier before any theorem was stated.

The class result carries a protective consequence for the program that motivated it. Caller-star topologies are maximally outside the class — any hub with in-degree or out-degree at least two meet-collapses — so the Tarski prefix set cannot serve as a lattice-valued fee on agent graphs, and the apparent mismatch between Section 1's base space and this section's positive regime is the point rather than an irony. Like the earlier bounded-frustration gate in the abelian layer, the theorem's negative half is load-bearing: it forecloses a tempting construction before anything is built on it.

One enumeration lesson from building that harness deserves record, because it is an instance of the program's gate discipline catching its own author: the standard shortcut "join-preserving maps out of L are monotone maps on join-irreducibles" is false for non-distributive L — $M_3 \to C_2$ has five join-preserving maps, not eight — so the harness enumerates monotone maps and filters by binary-join preservation, with the count as a frozen unit test.

7 7. Tropical postscript

A side conjecture from the abelian layer — that total unimodularity tames the gap between tropical and Barvinok rank under the literal $\{-1, 0, 1\}$-entry encoding — is refuted by the smallest star: the signed incidence of S_3 has tropical rank 2 and Barvinok rank 3, by hand on both sides (all four 3×3 minors have tied permanent minimizers; a rank-one term covering a symmetric off-diagonal pair would undercut a diagonal cell). Total unimodularity does not constrain the min-plus world under that encoding. The support encoding behaved differently: on every enumerated signed-incidence instance the two ranks agreed, and for $0/\infty$ patterns tropical nonsingularity is unique perfect matching while Barvinok factorization is biclique cover — so the surviving conjecture is a matching-versus-covering identity on incidence patterns, recorded as conjectured with the enumeration's stated bounds and budget-outs.

8 8. Related work

Ghrist and Riess introduced cellular sheaves of lattices and the Tarski Laplacian and proved the degree-0 Hodge–Tarski theorem; Riess's thesis defines the degree-1 apparatus and poses the conjecture resolved in Section 6. Ghrist, Lopez, North, and Riess develop quantale-weighted diffusion with a degree-0 Hodge–Lawvere theorem (suffix points are weighted global sections); no degree-1 statement appears there, and the graded/agent-composition applications remain open. Tenório, Arndt, and Mariano extend Čech cohomology to quantale bases with Set- and Ab-valued presheaves and adjoint-functor sheafification — orthogonal machinery to the congruence-and-nucleus route here. Grandis's semiexact framework, in which lattices-with-Galois-connections form a homological category, is the foundation we cite for the ambient algebra and deliberately do not claim. Johnstone's classification of nuclei on Boolean locales is the classical content of Section 4's first theorem. Young's sheaf-cohomological program analysis is the nearest neighbor in applications: lattice-valued presheaves over a program site with Galois connections and Kleene iteration, and a minimum-independent-fixes reading of H^1 rank that parallels this program's fee identity; it has no Tarski/fixed-point operator on the coefficient lattice, no caller-star topology, and no multi-agent instantiation, which is exactly the ground this paper occupies. The tropical rank inequalities and separating examples are due to Develin, Santos, and Sturmfels, with hardness results by Kim and Roush and the 0/1 support-pattern characterization by Gunawardena.

9 9. Honest-labeling appendix

Verified (sorry-free Aristotle solution + local rebuild, standard axioms, statements byte-identical to pre-registration): the twenty-five theorems of the first round (`StarComplement` 5, `CureLoop` 5, `BooleanNuclei` 5, `DisclosureNuclei` 6, `TarskiHarmonicDeg1` 7, counted with their inline lemmas per the ledger) and the four of `CyclicHodgeTarski`, including the cyclic Hodge–Tarski theorem; run stamps and axiom audits in `lean/aristotle_runs.json` and `lean/LEAN-FILE-CONVENTION.md`.

Verified negation: the seven-theorem refutation of the conjecture as stated (`TarskiHarmonicDeg1`).

Demonstrated: the model-adequacy gate (five pre-registered clauses, constructed corpus, encoder/evaluator separation); the prober's control battery and per-orientation table (seeded, stated bounds); the tropical enumeration within its bounds; and the code-binding property tests — `bulla/tests/test_lean_bindings.py` demonstrates, on deterministic adapters against the shipped round operator, that the operator is inflationary, strictly descends on confirming rounds, terminates within the initial fee, and that the shipped greedy disclosure set is sufficient and drop-minimal (a matroid basis). These tests upgrade the theorem–code identifications from asserted to demonstrated; they do not make the implementation verified, and nothing in this paper claims otherwise.

Verified (sprint 3): the class direction on arbitrary self-loop-free in-out-one quivers (`QuiverHodgeTarskiClass`, six theorems) and the full characterization with the two-family gadget (`QuiverHodgeTarskiGadget`, seven theorems, axiom audit clean on both headline results); the ∀-sheaf quantifier of the iff is pinned at `Type 0`, the class direction universe-polymorphic — stated as such.

Conjectured: the support-encoding rank identity (matching versus covering, enumerated bounds); the matroid-flats extension of the nucleus correspondence beyond spanning sets;

the invariant question for the pull-agreement object (Section 6's orientation-free $\mathrm{Pre}(\mathbb{H})$: the machine probes show it is not recoverable from any orientation-quantified family of Q's — the bowtie separates even the source/sink-free intersection — while the thesis's coequalizer realization audits clean on the zoo and a vertex-amalgamation fiber-product identity holds on all tested gluings; whether it is the right degree-1 invariant, with a lattice Mayer–Vietoris, is the program's next target).

Known-math attributions: Section 2 is standard linear algebra newly applied; Section 3 is Knaster–Tarski-adjacent folklore made precise; Section 4's classification is Johnstone's, formalized; Section 6's degree-0 control is Riess's theorem, used as an instrument.