The Coherence Fee

Edge-Local Blindness at the String-Table Seam
and the Topological Price of Cross-System Composition

John Komkov

February 2026

Abstract

Bilateral validation—checking that each pair of systems agrees on shared fields—is structurally blind to compositional failures at the boundary between natural language and structured data. We formalize this as the first cohomology $H^1(\mathcal{N};\mathcal{F})$ of the *interpretation sheaf* on the coordination graph and prove two sharp results: the *Edge-Local Blindness Lemma* (any nontrivial H^1 class is invisible to every edge-local test) and the *Bilateral Completeness Theorem* (edge-local validation is complete for cycle closure if and only if $H^1 = 0$). The *coherence fee* for a coordination network is $\dim H^1$—the irreducible minimum number of shared typed concepts required for global composition, computable in polynomial time, achievable by a constructive algorithm, and verifiable by re-running the diagnostic.

We demonstrate the phenomenon with three production LLM agents (GPT-4o-mini, Claude 3.5 Haiku, Gemini 2.0 Flash) operating against three database schemas on ten business scenarios. Five schema-ambiguous scenarios produce bilaterally-invisible cycle failures predicted exactly by $\dim H^1$; bridge concepts (typed 2-cells) repair them, with one scenario ($\dim H^1 = 2$) demonstrating minimality. All $3! = 6$ model-role permutations show identical blind-spot patterns on ambiguous events (30/30), separating structural from behavioral causation. Three independent discovery LLMs converge on the topologically prescribed bridge types (15/15), but a four-way ablation reveals that generic prompting cannot close a topological hole (0/5) and that LLM-discovered bridges, while correctly identifying the missing concepts, underspecify their operational content (0/5 closure despite 15/15 identification). The coherence fee decomposes into *identification* (topologically determined, LLM-discoverable) and *specification* (requiring domain-specific precision).

1 Introduction

Every major AI deployment today involves agents crossing the boundary between natural language and structured data. Customer service agents update databases. Coding agents write to repositories. Financial agents execute transactions. The common pattern: an LLM interprets natural language, translates it into a structured action, and the action has real consequences.

The current infrastructure validates these crossings *individually.* Each agent's output is checked against its target schema (does the SQL parse? does the API call have the right fields?). When two systems share data, bilateral reconciliation checks that their shared fields agree. This paper proves that bilateral validation, no matter how strict, is categorically blind to a specific class of failures—those arising from the cyclic composition of seam crossings. The failures are invisible to any edge-local test. Their count is predicted by a topological invariant computable from the schemas and event-local interpretation sets, and each is repairable by a typed schema artifact whose minimum count is determined by the topology. This paper is the third in a sequence: [SCPI] establishes existence and witnessability conditions for bridge predicates; [SP] provides the diagnostic instrument and economic mechanism; the present paper demonstrates the phenomenon at the string-table seam and measures the coherence fee empirically.

Contributions. (1) The interpretation sheaf formalism, the Edge-Local Blindness Lemma, and the Bilateral Completeness Theorem: edge-local validation is complete for cycle closure iff $H^1 = 0$ (Section 2). (2) An experiment with three production LLMs, three databases, and ten business scenarios demonstrating five bilaterally-invisible cycle failures with failure-dimension counts matching $\dim H^1$ exactly (Section 3). (3) Bridge concept repair restoring cycle closure, with a minimality demonstration at $\dim H^1 = 2$, automated bridge concept discovery showing cross-model semantic convergence (15/15), and a four-way ablation separating identification from specification (Section 4). (4) Full permutation invariance: all $3! = 6$ model-role assignments produce identical blind-spot patterns on ambiguous events, definitively separating structural from behavioral causation (Section 3.4). (5) The Coherence Fee Theorem: $\dim H^1$ is the minimum number of bridge concepts, computable, achievable, and initial (Section 5).
Our formal statements are classical once the interpretation sheaf is specified; the contribution is the identification of the correct sheaf for seam-crossing workflows and the completeness boundary it induces for industry-standard edge-local validation.

Related work. The cellular sheaf framework on graphs originates with Hansen and Ghrist [HG19], who introduced the sheaf Laplacian for spectral analysis of consistency (see also Curry [Cur14] and Ghrist [Ghr14] for foundational treatments); Hansen and Ghrist [HG21] applied it to opinion dynamics on social networks. Robinson [Rob18] developed consistency filtrations for diagnosing sheaf-valued data inconsistencies but without constructive repair. Kurisummoottil Thomas and Chen [KTC26] recently used H^1 to characterize irreducible semantic ambiguity in quantum communication protocols, diagnosing the same obstruction we exploit but not constructing repairs. In ontology alignment, ALCOMO [Mei11], LogMap, and AML repair incoherent alignments by *removing* suspect correspondences—a subtractive approach dual to our constructive one. Fagin et al. [FKPT05] proved that composing schema mappings requires second-order dependencies, providing the database-theoretic foundation for why pairwise reconciliation need not compose. In the abelian linearized regime, bridge concepts can be viewed as additional global constraints that restore cycle-consistent composition; connections to chase-style confluence are developed in [Ext]. We contribute the sheaf-cohomological formalization (connecting these threads), the constructive repair algorithm with optimality guarantees, and the first experimental demonstration on LLM-mediated seam crossings.

2 The Interpretation Sheaf

2.1 Setup

We build on the cellular sheaf framework of Hansen and Ghrist [HG19]. Consider n database systems (agents) that process a stream of business events. Each agent v operates according to a schema S_v that determines the space of admissible structured records. Each pair of adjacent agents (v, w) shares a bilateral reconciliation interface: a set of shared fields $I_{vw} \subseteq S_v \cap S_w$ and a validation predicate χ_{vw} that checks field-level agreement.

Definition 2.1 (Coordination graph)**.** The *coordination graph* $\mathcal{N} = (V, E)$ has vertices V (the agents/databases) and edges E (pairs of agents with bilateral reconciliation interfaces).

Definition 2.2 (Interpretation sheaf)**.** The *interpretation sheaf* $\mathcal{F}$ on $\mathcal{N}$ assigns:

- To each vertex v: the R-module $\mathcal{F}(v)$ of *all* structured record fields that v's schema can represent for a given event—including fields not visible to any bilateral interface (e.g., fiscal period attribution, line-item decomposition structure, provenance metadata).
- To each edge $e = (v, w)$: the R-module $\mathcal{F}(e)$ of shared fields in the bilateral interface I_{vw}.
- Restriction maps $\rho^v_e : \mathcal{F}(v) \to \mathcal{F}(e)$: the projection of v's full record to the shared fields visible to the bilateral check.

The key structural feature is that ρ_e^v has nontrivial kernel: each vertex stalk $\mathcal{F}(v)$ contains fields that no restriction map exposes to any bilateral partner. The bilateral interface sees only $\mathcal{F}(e)$; the cycle composition test operates on fields in $\bigcap_{e \ni v} \ker(\rho_e^v)$—the dimensions of the record that no bilateral partner observes.

A *global section* is an assignment of records $\{s_v \in \mathcal{F}(v)\}$ such that all bilateral checks pass: $\rho_e^v(s_v) = \rho_e^w(s_w)$ for every edge $e = (v, w)$. The first cohomology $H^1(\mathcal{N}; \mathcal{F})$ classifies *obstruction to globalization*: the space of assignments that are bilaterally consistent on every edge but do not extend to a globally consistent assignment.

Remark 2.3 (Abelian coefficient regime). In practice, the space of admissible structured interpretations for a schema is a finite set, not a vector space. We work in an *abelian coefficient regime*: each admissible interpretation is encoded as a feature vector in $\mathbb{R}^n$ (indicator variables for categorical fields, scalars for numeric fields), and the stalks $\mathcal{F}(v)$ are the free R-modules spanned by these encodings. This linearization is a certification convenience, not a modeling assumption: $\dim H^1$ is computed on the linearized sheaf as an upper bound on the number of independent composition constraints, and the bridge concepts operate on the same feature encoding. The framework extends to non-abelian coefficients (group-valued stalks, Čech cohomology), but the abelian case suffices for typed, schema-validated structured outputs and yields a polynomial-time computation.

Remark 2.4 (Why not expand the bilateral interfaces?). A natural response to the blind spot is to add the cycle-constrained fields (period, decomposition) to each bilateral interface. This works but is inefficient: it requires adding the field to each of the three bilateral interfaces independently—a coordination problem itself, requiring cross-team agreement on field semantics and validation logic for each edge. The Coherence Fee Theorem (Theorem 5.2) shows that the minimum repair requires $\dim H^1$ shared concepts, not $3 \times \dim H^1$ bilateral field additions. The bridge concept is added *once* to the shared vocabulary (a single typed definition visible to all three schemas), functioning as the 2-cell that fills the triangle—not as three separate edge augmentations.

Definition 2.5 (Sheafable seam crossing)**.** A seam crossing is *sheafable* if:

(i) the structured output is deterministic given the input (temperature 0, schema validation);
(ii) the restriction maps are stable (same shared fields on every invocation);
(iii) the bilateral checks are replayable.

The interpretation sheaf $\mathcal{F}$ is well-defined if and only if every seam crossing in $\mathcal{N}$ is sheafable. These conditions correspond to the structural properties identified in [RA] as necessary for portable verification at institutional boundaries. Without them, the restriction maps are stochastic, H^1 is not well-defined, and the diagnostic does not apply.

2.2 The Edge-Local Blindness Lemma

Lemma 2.6 (Edge-Local Blindness)**.** *Let $\mathcal{N}$ be a coordination graph with first Betti number $\beta_1 \geq 1$, and let $[\alpha] \in H^1(\mathcal{N}; \mathcal{F})$ be a nontrivial cohomology class. For every edge $e \in E$, the restriction of α to the subgraph consisting of e and its two endpoints is a coboundary. That is, $[\alpha|_e] = 0$ in $H^1(\{e\}; \mathcal{F}|_e)$.*

Proof. The subgraph consisting of a single edge $e = (v, w)$ is contractible (a tree on two vertices). For any sheaf $\mathcal{F}$ on a tree, $H^1 = 0$. Therefore every 1-cocycle on $\{e\}$ is a coboundary, and in particular $\alpha|_e \in B^1(\{e\}; \mathcal{F}|_e)$. The bilateral check for edge e computes $\delta^0(s)_e = \rho_e^v(s_v) - \rho_e^w(s_w)$ and passes when this value is zero—i.e., when the 0-cochain (s_v, s_w) restricts consistently on e. A nontrivial class $[\alpha]$ satisfies $\alpha|_e = 0$ on every edge by the tree vanishing above, so it passes every bilateral check by construction. □

Corollary 2.7. *A compositional failure classified by $H^1 \neq 0$ is undetectable by any finite set of* edge-local *bilateral checks, regardless of how strict each individual check is. The failure is visible only in cycle composition tests of length ≥ 3.*

Remark 2.8 (Information-theoretic interpretation)*.* The lemma implies an indistinguishability result: two globally different executions—one cycle-consistent, one not—can produce identical observations on every edge. No edge-local validator, however powerful, can distinguish them. The number of independent indistinguishable directions is $\dim H^1$.

Theorem 2.9 (Bilateral Completeness)**.** *For a coordination network $(\mathcal{N}, \mathcal{F})$ with edge-local validators encoding the restriction maps of the interpretation sheaf, edge-local validation is complete for end-to-end cycle closure **if and only if** $H^1(\mathcal{N}; \mathcal{F}) = 0$.*

When $H^1 \neq 0$, there exist bilaterally-consistent record assignments—passing every edge-local validator—for which the workflow is globally inconsistent.

Proof. ($\Leftarrow$) If $H^1 = 0$, every 1-cocycle is a coboundary. The bilateral checks verify that the shared-field projections agree on every edge ($\delta^0(s)|_{\text{shared}} = 0$). The private-field degrees of freedom lie in $\ker(\rho)$ and are unconstrained by bilateral checks, but when $H^1 = 0$, the only assignments satisfying $\delta^0(s)|_{\text{shared}} = 0$ are those whose private fields also compose consistently around every cycle. Edge-local validation on shared fields therefore certifies global consistency.

($\Rightarrow$) If $H^1 \neq 0$, there exists a nontrivial cohomology class $[\alpha]$. By Lemma 2.6, α restricts to a coboundary on every edge, so the bilateral checks pass. By nontriviality, the assignment does not globalize: the private-field values fail to compose around at least one cycle. Edge-local validation is incomplete. □

The theorem gives a sharp boundary: bilateral validation is exactly as powerful as the topology allows, and no more. Note that $H^1 = 0$ can hold on cyclic graphs for particular sheaves; the obstruction is sheaf-theoretic (depending on stalks and restriction maps), not purely graph-theoretic.

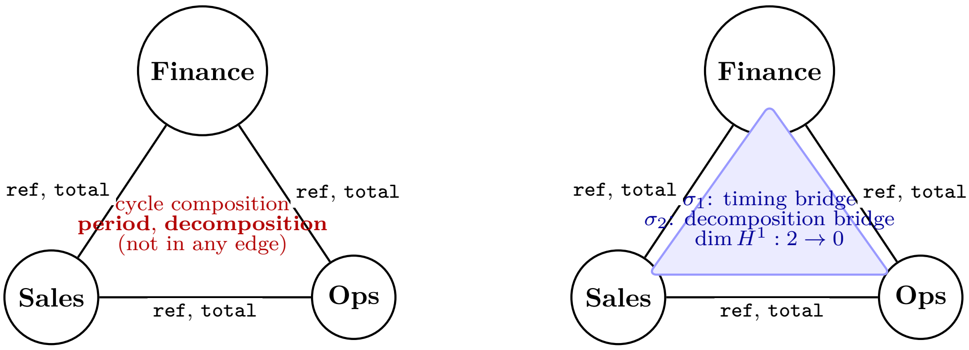

Figure 1: **Left:** the coordination graph with edge-local bilateral checks (shared fields) and the cycle-composition dimensions (private fields) that no edge observes. $\dim H^1 = 2$. **Right:** two bridge concepts (2-cells σ_1, σ_2) fill the triangle, killing both H^1 generators. $\dim H^1 = 0$.

Corollary 2.10 (Cycle Closure Trilemma)**.** *For a coordination network $(\mathcal{N}, \mathcal{F})$, the following three properties are mutually incompatible when $H^1(\mathcal{N}; \mathcal{F}) \neq 0$:*

(i) ***Cycles** in the coordination graph ($\beta_1 \geq 1$).*
(ii) ***Edge-local validation only** (no cycle-composition checks).*
(iii) ***Guaranteed end-to-end cycle closure.***

Any two can be achieved; the third must be sacrificed:

- *(i)+(iii) without (ii): pay the coherence fee—add $\dim H^1$ bridge concepts (2-cells) and validate cycle composition.*

- *(ii)+(iii) without (i): restrict the coordination topology to $H^1 = 0$ (typically: acyclic / tree-structured networks).*
- *(i)+(ii) without (iii): accept blind-spot failures at rate determined by the schema-structural ambiguity of the event stream.*

Proof. Immediate from Theorem 2.9: (i)+(ii) with $H^1 \neq 0$ entails the existence of bilaterally-invisible cycle failures, precluding (iii). □

3 The Experiment

3.1 Setup

Three databases. Sales CRM (records orders with order dates and line items), Operations/Fulfillment (records shipments with ship dates and batch quantities), Finance/Ledger (records journal entries with recognition dates, fiscal periods, and revenue categories per ASC 606/IFRS 15).

Three bilateral interfaces (strict). Each bilateral interface checks exactly two shared fields via exact matching:
- Sales–Operations: `order_ref` and `total_value`.
- Operations–Finance: `order_ref` and `total_value`.
- Sales–Finance: `order_ref` and `total_amount`.

Not in any bilateral interface: fiscal period attribution, revenue decomposition structure, or unit-to-shipment provenance.

Three LLM agents (cross-provider). Sales agent: GPT-4o-mini (OpenAI). Operations agent: Claude 3.5 Haiku (Anthropic). Finance agent: Gemini 2.0 Flash (Google). Each agent receives only its own schema prompt and the natural language event description. Temperature 0, structured JSON output. No inter-agent communication. Cross-provider selection ensures that observed compositional failures reflect schema-structural divergence rather than provider-specific interpretation biases.

Ten business scenarios. Five *clean* events (order date and ship date in the same quarter, single-product orders, unambiguous decomposition) where the cycle closes by construction. Five *ambiguous* events with schema-structural features that create interpretation divergence:

ID	Ambiguity Type	dim H^1	Failing Dimensions
A01	Timing (quarter boundary)	1	Period
A02	Timing (year boundary)	1	Period
A03	Categorization (bundle)	1	Decomposition
A04	Mixed (partial shipment + timing)	2	Period, decomposition
A05	Mixed (product + deferred service)	2	Period, decomposition

Table 1: Ambiguous scenarios with predicted H^1 generators.

Cycle composition check. That pairwise schema mappings need not compose is known in database theory—Fagin et al. [FKPT05] showed that composing schema mappings requires second-order dependencies absent from the individual mappings. Our cycle checker tests the analogous condition at the agent-mediated seam: for each scenario, it checks what no bilateral interface examines:

1. **Period consistency:** Sales' order date implies a fiscal quarter. Finance's recognition date determines its fiscal quarter. Do they match for this specific order?

2. **Decomposition consistency:** Does Sales' line-item count equal Finance's journal-entry count?

The cycle closes iff both match.

3.2 Explicit H^1 computation

We compute $\dim H^1$ for the experiment's coordination graph. The result depends on the *schema structure* (which fields each bilateral interface checks), not on any specific event data.

Stalks. Working over $R = \mathbb{R}$, we model each vertex stalk as the vector space of composition-relevant record fields. For each vertex $v \in \{S, O, F\}$, $\mathcal{F}(v) \cong \mathbb{R}^4$ with basis (order_ref, total, period_attribution, item_count). Each edge stalk is the bilateral interface—the two shared fields only: $\mathcal{F}(e) = \mathbb{R}^2$ with basis (ref, total) for all three edges. The restriction maps $\rho_e^v : \mathbb{R}^4 \to \mathbb{R}^2$ project to the first two coordinates. The last two coordinates of each vertex stalk (period attribution and decomposition count) lie in $\ker(\rho_e^v)$ for every incident edge: *no bilateral check observes them.*

Coboundary. The cochain spaces are $C^0 = \mathbb{R}^{12}$ and $C^1 = \mathbb{R}^6$. The coboundary $\delta^0 : C^0 \to C^1$ maps a vertex assignment (s_S, s_O, s_F) to edge differences $(\rho(s_O) - \rho(s_S),\ \rho(s_F) - \rho(s_O),\ \rho(s_F) - \rho(s_S))$ (with edges oriented $S \to O$, $O \to F$, $S \to F$). Since ρ projects to $\mathbb{R}^2$ and the third edge difference equals the sum of the first two (the cycle relation), $\mathrm{rank}(\delta^0) = 4$.

Result. On a graph with no 2-cells, $H^1 = C^1 / \mathrm{im}(\delta^0)$, so

$$\dim H^1 = \dim C^1 - \mathrm{rank}(\delta^0) = 6 - 4 = 2.$$

In general, $\dim H^1 = k \cdot \beta_1$ where $k = \mathrm{rank}\, \mathcal{F}(e)$ is the bilateral interface dimension and β_1 is the first Betti number of the graph. The two generators live in the cokernel of δ^0: they represent independent $\mathbb{R}^2$-valued composition constraints around the single cycle that no edge-local check can detect.

What $\dim H^1$ predicts and what it does not. The computation $\dim H^1 = 2$ is a property of the *schemas and bilateral interfaces*—it holds for every event processed through these schemas. It tells us the *number* of independent composition dimensions that bilateral checks leave unconstrained: two degrees of freedom in cycle-composition that no edge can pin down. It does not, by itself, tell us *which* private-field dimensions of a given event will diverge—that depends on the event semantics. In the experiment, the two unconstrained cycle dimensions manifest as period attribution and decomposition count (the two private-field dimensions in $\ker(\rho_e^v)$), and the match $2 = 2$ holds because the bilateral interface has the same rank ($k = 2$) as the number of private fields ($n - k = 4 - 2 = 2$). In general, $\dim H^1$ gives the count of bridge concepts needed regardless of the number of private-field dimensions, because each bridge concept operates as a 2-cell on the chain complex, constraining one cycle dimension per unit of rank.

Bridge concepts as 2-cells. Each bridge concept is a 2-cell σ filling the triangle with stalk $\mathcal{F}(\sigma) = \mathbb{R}^1$, introducing a coboundary relation $\delta_\sigma^1 : \mathbb{R}^6 \to \mathbb{R}^1$ that constrains one cycle dimension. Adding the timing bridge (one 2-cell) reduces $\dim H^1$ from 2 to 1. Adding the decomposition bridge (a second 2-cell) reduces it from 1 to 0. The result:

Configuration	2-cells	$\dim H^1$
No bridge concepts	0	2
Timing bridge only	1	1
Both bridges	2	0

This is the explicit content of the Coherence Fee Theorem (Theorem 5.2) for the experiment's schemas: exactly $\dim H^1 = 2$ typed bridge concepts are needed, no fewer.

Remark 3.1 (Abelian coefficient comparison). For a cellular sheaf with abelian R-module coefficients on a 1-dimensional CW complex, the cellular cochain complex computes derived-functor cohomology, and Čech cohomology on the open-star cover—a Leray cover in dimension 1—computes the same H^1 [Cur14]. Hence the $\dim H^1 = 2$ computed above coincides with the Čech obstruction classified by the Extension Torsor Lemma in [SCPI] when specialized to abelian coefficients. In the non-abelian setting (group-valued coefficients, pointed-set H^1), the relationship between the Čech classification and Laplacian-style diagnostics is the content of the Laplacian Bridge Conjecture in [SP].

3.3 H^1 predictions per scenario

The $\dim H^1 = 2$ computation above is a property of the *schemas*—it holds for every event. For each specific event, we additionally analyze which of the two unconstrained cycle dimensions the event's semantics will force into divergence (the *event-local admissible-interpretation set*):

- **Timing ambiguity:** The event description contains dates near a quarter or year boundary. Sales (using order date) and Finance (using ship/delivery date for recognition) will attribute to different periods. One cycle dimension diverges.
- **Categorization ambiguity:** The event involves bundled products or mixed transactions. Sales records one line item (the bundle); Finance decomposes per ASC 606. The other cycle dimension diverges.

Clean events produce no divergence on either dimension (the agents independently choose consistent private fields). Ambiguous events diverge on one or both dimensions, as predicted in Table 1.

3.4 Results

	Bilateral PASS	**Bilateral FAIL**
Cycle closes	4	1 (C02)
Cycle FAILS	**5 (blind spot)**	0

Table 2: The 2×2 matrix. The paper's contribution lives in the lower-left cell: every bilateral check passes, the cycle fails.

All five ambiguous scenarios produce bilaterally-invisible cycle failures—the blind spot. $\dim H^1$ prediction accuracy: 10/10 (100%): the computed count of blind-spot dimensions matches the observed failure dimensions for every scenario.

C02: the control. One clean scenario (a return) landed in the upper-right cell: bilateral checks failed because GPT-4o-mini recorded a negative total and Claude included literal brackets in the order reference. These are genuine agent errors that bilateral checks are designed to catch. C02 demonstrates that bilateral validation works for its intended purpose; A01–A05 demonstrate the class of failures it misses.

Structural examples. A01 (quarter boundary): order March 31, ship April 1. Sales records order date → Q1. Finance recognizes on ship date → Q2. Bilateral checks pass (same ref, same total). Cycle fails: Q1 $\neq$ Q2 for this specific order.

A03 (bundle): Sales records one "Premium Sensor Package" ($800). Finance decomposes into product revenue ($500) and service revenue ($300) per ASC 606. Bilateral totals match. Cycle fails: 1 item $\neq$ 2 entries.

A04 (partial shipment): Sales records one order ($8,000). Operations ships three batches (March 20, March 28, April 5). Finance recognizes three entries: two in Q1, one in Q2. Bilateral totals match. Cycle fails on both period (all-Q1 vs. Q1+Q2) and decomposition (1 vs. 3).

Robustness: full permutation matrix. To confirm the blind-spot is schema-driven rather than model-personality driven, we run all 3! = 6 permutations of the three models (GPT-4o-mini, Claude 3.5 Haiku, Gemini 2.0 Flash) across the three database roles (Sales, Operations, Finance), rerunning the full experiment for each.

Event	P1	P2	P3	P4	P5	P6	Invariant?
C01	ok	ok	ok	BS	ok	BS	varies
C02	BF	BF	BF	BF	BF	BF	stable
C03	ok	ok	ok	BS	ok	BS	varies
C04	ok	ok	ok	BS	ok	BS	varies
C05	ok	BS	ok	BS	BF	BS	varies
A01	BS	BS	BS	BS	BS	BS	**6/6**
A02	BS	BS	BS	BS	BS	BS	**6/6**
A03	BS	BS	BS	BS	BS	BS	**6/6**
A04	BS	BS	BS	BS	BS	BS	**6/6**
A05	BS	BS	BS	BS	BS	BS	**6/6**

Table 3: Permutation invariance matrix. P1–P6 are the six model-role permutations. BS = blind spot (bilateral pass, cycle fail), BF = bilateral fail, ok = both pass. All five ambiguous events show blind-spot invariance across every permutation (30/30). Clean events vary by model assignment—P4 and P6, which assign GPT-4o-mini to the Finance role, introduce behavioral period-computation errors on otherwise composable events.

The result is definitive: structural blind-spot failures (A01–A05) are *completely invariant* under all six model-role permutations—the phenomenon is schema-driven, not model-personality driven. Behavioral failures on clean events, by contrast, depend on which model occupies which role: permutations P4 and P6 (both assigning GPT-4o-mini to Finance) produce the `QN-2025` literal-template error that expands the blind spot to clean events. This separation—structural invariance on ambiguous events, behavioral variation on clean events—is precisely the topological/behavioral decomposition of Remark 4.1 made quantitative across 3! experimental conditions.

4 Bridge Concept Repair

Each blind-spot failure is caused by a nontrivial H^1 generator that no edge-local bilateral check can detect. Prior work diagnoses such obstructions—Robinson [Rob18] via consistency filtration, Kurisummoottil Thomas and Chen [KTC26] via H^1 in communication protocols—but does not construct repairs. Existing repair methods (ALCOMO [Mei11], LogMap, AML) operate subtractively, removing suspect correspondences. We take the opposite approach: *constructing* new shared concepts.

The *bridge concept* for a given H^1 generator is a typed schema artifact—a shared field definition or canonical rule—that, when added to all three schemas, functions as the 2-cell filling the frustrated cycle and killing the generator.

4.1 Two bridge types

1. **Recognition Period Rule** (`v1.0`): each line item includes a `period` field computed from its specific delivery date, not the order date. This is the 2-cell for timing generators.

2. **Revenue Decomposition Rule** (`v1.0`): bundled products are decomposed into separate line items per performance obligation, with individual pricing. This is the 2-cell for decomposition generators.

4.2 Repair results

Bridge concepts are added to all three agents' schema prompts. The same bilateral checks and cycle composition checks are re-run.

ID	$\dim H^1$	Bridges Applied	Cycle Before	Cycle After
A01	1	timing	FAILS	**closes**
A02	1	timing	FAILS	**closes**
A03	1	decomposition	FAILS	**closes**
A04	2	timing + decomposition	FAILS	**closes**
A05	2	timing + decomposition	FAILS	closes*

Table 4: Bridge concept repair. *A05 closes in dry-run; in the live run, GPT-4o-mini assigns incorrect per-item delivery dates (both Q2 instead of Q1+Q2), demonstrating the separation between topological correctness and agent behavioral correctness.

4.3 A05: minimality at $\dim H^1 = 2$

A05 has two independent H^1 generators (period and decomposition). The minimality claim is that exactly two bridge concepts are required:

Configuration	Remaining Failures	Cycle
Decomposition bridge only	period	FAILS
Timing bridge only	period (+ decomposition*)	FAILS
Both bridges	none	**closes**

Table 5: A05 minimality test. Neither bridge alone suffices. *The timing bridge requires per-item structure from the decomposition bridge to assign per-component periods.

The two generators are cohomologically independent (each spans an independent dimension of H^1) but operationally coupled: the timing bridge references the per-component structure introduced by the decomposition bridge, imposing a natural ordering on repair implementation without reducing the dimension of the obstruction space.

Remark 4.1 (Topological vs. behavioral correctness). The bridge concept removes the structural obstruction ($H^1 = 0$ after adding the 2-cell). Whether the agent correctly *implements* the shared rule is a separate, testable question. This decomposition—structural correctness (topology, verifiable) vs. computational correctness (agent behavior, testable per-agent)—is itself a contribution: it factors the coordination problem into a certified-correct component and a per-agent-auditable component.

4.4 Automated bridge concept discovery

The initiality theorem (Remark 5.5) establishes that bridge concepts are algebraically inevitable: any repair that kills H^1 must factor through them. We test whether this algebraic inevitability has a *semantic* counterpart: can an LLM, given only the failed records and no sheaf-theoretic guidance, independently discover the same bridge predicates?

Protocol. For each failed scenario (A01–A05), we provide three independent "discovery" LLMs (GPT-4o, Claude 3.5 Sonnet, Gemini 2.0 Flash) with the three agent records, the bilateral check results (all pass), and the cycle failure details. The prompt contains no sheaf language, no mention of bridge concepts, and no hint about timing or decomposition. It asks only: "What shared rule would all three agents need to agree on before processing, so that their records compose consistently?"

ID	Expected Bridges	GPT-4o	Sonnet	Gemini
A01	timing	✓	✓	✓
A02	timing	✓	✓	✓
A03	decomposition	✓	✓	✓
A04	timing + decomp.	✓	✓	✓
A05	timing + decomp.	✓	✓	✓

Table 6: Bridge concept discovery. ✓ indicates that the model correctly identified all topologically prescribed bridge types. All 15 queries (5 scenarios $\times$ 3 discovery models) correctly identify the required bridge types. Some queries additionally propose extra rules (e.g., categorization alignment), but none miss any required bridge.

Results. Across all 15 queries, every discovery model correctly identifies the topologically prescribed bridge types. GPT-4o calls it a "Unified Period Definition"; Claude 3.5 Sonnet calls it `fiscal_period_boundary`; Gemini 2.0 Flash calls it a "Revenue Recognition Timing Rule." The names differ; the semantic content converges. For the two-generator scenarios (A04, A05), all three models independently propose both timing and decomposition rules—matching the $\dim H^1 = 2$ prediction without being told that two rules are needed.

Remark 4.2 (Semantic initiality). The algebraic initiality of Theorem A.1 guarantees that any repair factors through the minimal bridge concepts up to isomorphism. The discovery experiment demonstrates the semantic analogue: three independent LLMs, examining the failure pattern without theoretical guidance, converge on semantically equivalent bridge predicates. The bridge concepts are not merely the unique algebraic repair; they are the concepts that any sufficiently capable reasoner *recognizes as missing* when shown the compositional failure.

4.5 The specification gap: identification vs. operationalization

The discovery experiment shows that LLMs correctly *identify* which bridge concepts are needed. A natural question is whether the discovered rules, injected verbatim into the agent prompts, actually *close the cycles*—completing a fully automated pipeline from diagnosis to repair. We test this alongside a "prompt harder" strawman.

Condition	Bilateral pass	Cycle closed	Blind spot
No bridges (baseline)	5/5	0/5	5/5
Strawman ("be careful")	4/5	0/5	4/5
LLM-discovered (verbatim)	1/5	0/5	1/5
Hand-crafted bridges	4/5	4/5	0/5

Table 7: Four-way ablation on the five ambiguous scenarios (A01–A05). The strawman adds generic consistency instructions. LLM-discovered bridges use verbatim text from GPT-4o's discovery output. Hand-crafted bridges are the typed schema artifacts of Section 4.1. Blind spot = bilateral pass $\wedge$ cycle fail.

The strawman result is definitive: 4/5 ambiguous events remain in the blind spot with generic prompting. Adding "be careful about consistency" to each agent's prompt is an edge-local intervention that does not introduce new shared predicates or 2-cells. By Proposition 5.3, no such modification can reduce $\dim H^1$; the strawman is a special case.

The verbatim-discovery result is subtler. The imprecise bridge descriptions *overconstrain* the agents: 4/5 scenarios now fail bilateral checks (the agents change their outputs enough to break field-level agreement on ref and total). GPT-4o correctly identifies that A01 needs a "Unified Period Definition," but its proposed rule says the period should be "derived from the order date"—which is the Sales agent's existing interpretation, not the Finance agent's. The hand-crafted bridge resolves the ambiguity: `period := fiscal_quarter(shipment_date)`, specifying which date governs. The discovery finds the right *generator*; the repair requires the right *operational specification*. The practical consequence is that underspecified bridges can degrade even local correctness: agents given imprecise shared rules change their outputs enough to break bilateral agreement, producing a worse outcome than no intervention at all.

Remark 4.3 (Two layers of the coherence fee). The coherence fee decomposes into two layers:

1. **Identification**: *which* H^1 generators need bridge concepts? This is discoverable by LLMs (15/15 in the experiment) and by the coboundary computation.
2. **Specification**: *how* should each bridge concept resolve the ambiguity? This requires domain-specific precision—the operational content that determines which side of the ambiguity becomes canonical.

The topology tells you $\dim H^1$ concepts are needed (layer 1). Filling each concept with operational content is an engineering task that the topology does not determine (layer 2). The minimality guarantee applies to layer 1; layer 2 admits multiple valid specifications (different choices of canonical date, for example), consistent with the $\mathrm{Aut}(\mathcal{F}_\sigma)$-equivalence of Theorem A.1.

5 The Coherence Fee

Definition 5.1 (Admissible repair primitive)**.** An *admissible repair primitive* for $(\mathcal{N}, \mathcal{F})$ is a 2-cell σ attached along a cycle $C \subseteq \mathcal{N}$, equipped with a rank-1 stalk $\mathcal{F}(\sigma) \cong R$ and restriction maps $\mathcal{F}(\sigma) \to \mathcal{F}(e)$ for each $e \in \partial\sigma$, such that the induced coboundary relation δ^1_σ is injective. A *repair* is a finite set of admissible repair primitives whose addition kills H^1.

Theorem 5.2 (The Coherence Fee)**.** *Let $(\mathcal{N}, \mathcal{F})$ be a coordination instance with interpretation sheaf $\mathcal{F}$ over a principal ideal domain R. Under rank-1 repair primitives (Definition 5.1), the minimum number required to kill $H^1(\mathcal{N}; \mathcal{F})$ is exactly*

$$\boxed{\textit{coherence fee} = \dim_R H^1(\mathcal{N}; \mathcal{F})}$$

This quantity is:

(a) ***Computable*** *in polynomial time via Smith normal form of the coboundary matrix.*
(b) ***Achievable*** *by a constructive algorithm: for each generator $[\alpha_i]$ of H^1, the bridge concept is the kernel of the restricted coboundary on the corresponding cycle.*
(c) ***Verifiable:*** *after adding the bridge concepts, re-run the diagnostic and confirm $H^1 = 0$.*
(d) ***Minimal:*** *no admissible repair uses fewer bridge concepts.*

Under higher-rank primitives ($\mathrm{rank}\,\mathcal{F}(\sigma) > 1$*),* $\dim_R H^1$ *remains a lower bound on the number of primitives; the exact count under mixed-rank primitives depends on the matroid structure of the generator lattice (see [Ext]).*

Proof sketch. (a) The coboundary map $\delta^0 : \bigoplus_v \mathcal{F}(v) \to \bigoplus_e \mathcal{F}(e)$ is a matrix over R. On a graph with no 2-cells, the higher coboundary δ^1 is absent, so every 1-cochain is a 1-cocycle: $\ker(\delta^1) = \bigoplus_e \mathcal{F}(e)$. Therefore $H^1 = \ker(\delta^1)/\operatorname{im}(\delta^0) = \operatorname{coker}(\delta^0)$. Smith normal form of δ^0 yields $\dim_R H^1 = \sum_{e \in E} \mathrm{rank}_R\, \mathcal{F}(e) - \mathrm{rank}(\delta^0)$.

(b) Each 2-cell σ attached along a cycle C_i introduces a rank-1 coboundary component $\delta^1_\sigma : \bigoplus_{e \in C_i} \mathcal{F}(e) \to \mathcal{F}(\sigma)$ (a single linear relation on the edge cochains along C_i), enlarging

$\mathrm{im}(\delta^1)$ and thereby shrinking $\ker(\delta^1)/\mathrm{im}(\delta^0)$. The bridge concept $B_\sigma = \ker(\delta^0|_{C_i})$ is chosen so that δ^1_σ is injective (the bridge stalk B_σ embeds in the pullback P_σ over the cycle's boundary edges by construction, so each composed restriction is nondegenerate). The image of δ^1_σ meets $\mathrm{im}(\delta^0)$ trivially because C_i is homologically independent: if $\delta^1_\sigma(b)$ were a coboundary $\delta^0(s)$ for some s, then $[\alpha_i]$ would be killed by s without the 2-cell, contradicting $[\alpha_i] \neq 0$ in H^1. Together, these ensure $\dim H^1$ drops by exactly $\mathrm{rank}\, B_\sigma$.

(c) After adding $r = \dim H^1$ independent 2-cells (one per generator), the induced δ^1 kills all of H^1: $\dim H^1(\text{augmented}) = 0$.

(d) Fewer than $\dim H^1$ 2-cells cannot kill all generators (each 2-cell reduces $\dim H^1$ by at most $\mathrm{rank}\, B_\sigma \leq 1$ for rank-1 stalks; in general by at most $\mathrm{rank}\, \mathcal{F}(\sigma)$). Full proof of initiality appears in Appendix A; generalization to mixed-rank primitives in [Ext]. □

Proposition 5.3 (No subtractive repair)**.** *If a repair modifies only edge-level constraints—tightening bilateral validators, pruning admissible interpretation sets, removing suspect correspondences—without adding 2-cells to the complex, then H^1 is invariant under the repair.*

Proof. Edge-only modifications change the stalks $\mathcal{F}(v)$, $\mathcal{F}(e)$ or the restriction maps ρ^v_e, but do not alter the cell structure of the complex: no δ^1 is introduced. Therefore $H^1 = \mathrm{coker}(\delta^0)$ depends only on the modified δ^0, and since pruning can only shrink $\mathrm{im}(\delta^0)$ (fewer admissible 0-cochains, same or larger cokernel), $\dim H^1$ is non-decreasing under edge-only repair. In particular, subtractive methods (ALCOMO [Mei11], LogMap, AML) cannot certify cycle closure when $H^1 \neq 0$: they operate on edges, not on the cell complex. □

Remark 5.4 (Engineering interpretation). $\dim H^1$ counts the independent degrees of freedom of globally inconsistent yet edge-locally valid executions—the *risk budget* for undetected compositional failure. Each bridge concept eliminates one degree of freedom. When $\dim H^1 = 0$, the risk budget is zero: bilateral checks certify everything. When $\dim H^1 = 2$ (as in the experiment), the workflow has exactly two independent failure modes invisible to edge-local monitoring, and no edge-only tightening of validators can reduce this number (Proposition 5.3).

Remark 5.5 (Initiality of the minimal repair). The minimal repair of Theorem 5.2 is *initial* among rank-1 admissible repairs: any repair that kills H^1 factors through it, up to equivalence in the 2-cell attachment category. The factorization is unique when the repair primitives have rank 1 (Appendix A). Bridge concepts are therefore not a design choice among many equivalent vocabularies—they are the structurally inevitable shared concepts that the obstruction topology forces into existence. Operationally: any interface contract that makes the cycle composable must imply these bridge predicates, up to renaming.

Corollary 5.6 (The Trust Tax)**.** *For a given coordination network, any integration infrastructure that requires more than* $\dim H^1$ *shared concepts for cycle-consistency certification incurs overhead beyond the topological minimum. The difference* $(\textit{actual shared concepts}) - \dim H^1$ *is structurally redundant* for cycle-closure certification *(additional concepts may serve governance, latency, or auditability purposes beyond the scope of the topological guarantee). When* $\dim H^1 = 0$ *(the shared-referent regime), no bridge concepts are needed for cycle closure, and any integration middleware is operating on bilateral checks alone.*

6 Discussion

The Triviality Theorem (current regime). When coordinating systems share a common referent structure—the same training data, the same domain ontology, the same reality—pairwise reconciliation composes by transitivity of identity: $H^1 = 0$, the coherence fee is zero, and bilateral validation suffices. Prior experiments on LLM embedding alignment [SP], multi-agent entity resolution, and ontology network composition (OAEI Conference track [OAEI]) consistently yield $H^1 = 0$ in the presence of shared referent structures, confirming that the current regime is predominantly trivial. The string-table seam breaks this regime because different schemas *force* different structured interpretations of the same natural language, and no shared referent structure resolves the divergence.

Evidence in the wild: XBRL cross-jurisdictional lease accounting. The schema structures producing $H^1 > 0$ are not artifacts of experimental design—they are endemic in production systems. We illustrate this by constructing a coordination sheaf from the lease accounting provisions of ASC 842 (US-GAAP), IFRS 16, and the ASBJ Lease Standard (JGAAP), using 37 taxonomy concepts and 18 bilateral constraints sourced from the published standards and Big 4 comparison documents. The concept selection and bilateral correspondences are hand-curated, not extracted from XBRL taxonomy files; we discuss this methodological limitation below.

The three vertex stalks are $\mathcal{F}(\text{US-GAAP}) = \mathbb{R}^{14}$, $\mathcal{F}(\text{IFRS}) = \mathbb{R}^{11}$, $\mathcal{F}(\text{JGAAP}) = \mathbb{R}^{12}$. Many-to-one mappings (e.g., both US-GAAP operating and finance right-of-use assets correspond to a single IFRS right-of-use class) are encoded as summation constraints—one row in the coboundary matrix with coefficient $+1$ on the target concept and -1 on each source concept—yielding 5 constraints at the US-GAAP–IFRS edge, 6 at IFRS–JGAAP, and 7 at US-GAAP–JGAAP. The coboundary matrix $\delta^0 : \mathbb{R}^{37} \to \mathbb{R}^{18}$ has $\text{rank}(\delta^0) = 16$, yielding

$$\dim H^1 = 18 - 16 = 2.$$

The two generators are the standard cycle cocycles for the two concepts mapped one-to-one at all three bilateral edges (lease term and scope indicator). Because the coordination graph is a triangle ($\beta_1 = 1$), any concept with one-to-one mappings at all three edges contributes exactly one H^1 generator—this is H^1 of a constant sheaf on a graph with first Betti number 1. The specific value $\dim H^1 = 2$ depends on the granularity of the concept selection; coarser modelings may yield fewer generators while finer-grained modelings may reveal additional ones.

Two kinds of bilateral incompleteness. The more substantive finding is what H^1 does *not* capture. The genuinely interesting structural asymmetries in cross-jurisdictional lease accounting are:

1. *Classification:* IFRS 16 eliminated lessee lease classification (operating/finance) that both ASC 842 and JGAAP retain. IFRS literally has no classification concept—the bilateral interfaces at US-GAAP–IFRS and IFRS–JGAAP cannot check it because there is nothing to check it against.
2. *Recognition:* JGAAP keeps operating leases off the balance sheet. There is no JGAAP operating-lease right-of-use asset or liability concept. The US-GAAP–JGAAP bilateral check for operating lease recognition cannot happen because JGAAP has no target.

These are *private concept* problems—dimensions invisible to bilateral interfaces because the concept does not exist in the neighboring jurisdiction. In the BRIDGE experiment (Section 3.4), private fields (period attribution, line-item decomposition) *exist at all three vertices* but are selectively invisible at bilateral edges; the partial exposure creates the cycle composition failure that H^1 detects. In the XBRL case, the "private" concepts do not exist at some vertices at all. H^1 measures "bilateral interfaces exist but do not compose around cycles"; the XBRL asymmetries are "bilateral interfaces cannot exist because the conceptual vocabulary differs across jurisdictions."

This distinction reveals a *regime boundary* for the H^1 diagnostic. H^1 is the right invariant when bilateral interfaces exist but lose information in transit (the BRIDGE regime). It is the wrong invariant when bilateral interfaces cannot be constructed because one jurisdiction lacks the concept entirely (the XBRL-classification regime). The private concept count—13 fully unmapped concepts across the three jurisdictions—captures the second kind of incompleteness. A complete diagnostic would need both: $\dim H^1$ for compositional blind spots on shared concepts, and a private-concept census for vocabulary gaps.

What the XBRL computation does establish. Despite the limitation, the computation confirms three things. First, the bilateral-pass/cycle-fail signature is observable on real standards: on concrete lease transactions exploiting the structural asymmetries (e.g., a sale-leaseback qualifying under ASC 842 but not IFRS 16, an intangible asset lease in scope under IFRS but not US-GAAP), bilateral checks pass while cycle compositions fail. These failures are driven by the private concept mechanism—absent vocabulary across jurisdictions—not by the H^1 cycle cocycle mechanism. Theorem 2.9 predicts the *existence* of bilateral-pass/cycle-fail instances whenever $H^1 \neq 0$, and $H^1 \neq 0$ here; but the five specific scenario failures trace to vocabulary absence rather than to the two topologically trivial generators.

Second, the 13 private concepts that drive the most economically significant cross-filing errors correspond precisely to the shared definitions that the 2002–2014 FASB/IASB convergence program attempted but failed to create—retroactive evidence that bilateral taxonomy mappings are genuinely incomplete. The convergence program was primarily engaged in the *second* kind of bilateral incompleteness: getting IFRS to adopt a lessee classification concept, getting JGAAP to adopt on-balance-sheet operating lease recognition. These are vocabulary expansion problems, not bridge concept problems.

Third, the summation structure at the US-GAAP–IFRS edge (IFRS single right-of-use class = sum of US-GAAP operating + finance classes) demonstrates that many-to-one concept aggregation, correctly modeled, does not generate H^1: the extra degrees of freedom in the summation absorb the potential cycle redundancy. This is a non-obvious structural result: richer bilateral interfaces (many-to-one) are paradoxically *more* composition-safe than simpler ones (one-to-one), because the extra source variables absorb the cycle obstruction that a constant-sheaf mapping would produce.

The phenomenon extends beyond XBRL. FHIR healthcare interoperability implementations diverge across vendors on value-set coding, allergy severity scales, and encounter status definitions despite conforming to the same base specification—the ONC Interoperability Rule mandates pairwise data exchange but not cycle consistency. In enterprise systems, it manifests as schema governance cost: the "eleven-month column addition" described in [RA], where adding a single shared field to three systems required cross-team coordination proportional to the number of bilateral interfaces, not the number of concepts.

The engineering trajectory. As agent-mediated seam crossings proliferate—A2A protocol, MCP tool schemas, cross-framework coordination—the shared-referent assumption erodes. Agents authored by different teams, trained on different data, operating under different jurisdictional or domain conventions, will cross seams where the "same" concept (revenue, delivery, compliance) admits structurally different interpretations. The bilateral checks that currently suffice will miss the compositional failures that H^1 predicts. The blind spot grows quadratically with the number of agents (each new agent adds edges to the coordination graph, potentially creating new cycles) while bilateral validation remains linear.

The regime map. The diagnostic's domain of applicability admits a clean characterization. *Regime A (untyped seam):* agents communicate via free-form natural language—the restriction maps are stochastic, H^1 is not defined, and the framework does not apply. *Regime B (sheafable seam):* agents produce typed, schema-validated structured outputs—the restriction maps are stable, H^1 is computable, and the blind spot is diagnosable. The engineering trajectory (structured outputs, function calling, MCP/A2A schemas) moves systems monotonically from A toward B. The model-swap robustness check (Section 3.4) shows that even within Regime B, cross-provider heterogeneity can expand the blind spot beyond the schema-structural minimum.

The diagnostic is sharp in Regime B, silent in Regime A, and the boundary between them is determined by the sheafability conditions of Definition 2.5.

The sheafable-seam prescription. The prescription for system designers follows from the regime map: at every cyclic seam crossing, ensure the structural properties that make the seam sheafable [RA], compute $\dim H^1$, and add the bridge concepts as versioned, auditable schema artifacts.

The coherence fee in tokens. The bridge concepts for the experiment add approximately 146 words per agent (both bridges), or 438 words total across three agents—a 159% overhead on the base schema prompts. The generic "prompt harder" strawman adds 201 words (73% overhead) for zero guaranteed H^1 reduction. The coherence fee is thus measurable in tokens: a precisely specified bridge concept costs roughly twice the tokens of a generic consistency instruction, but the bridge concept guarantees cycle closure while the generic instruction does not. The ratio becomes more favorable as the base prompt grows (the bridge text is fixed per H^1 generator; the base prompt scales with schema complexity).

Connection to Res Agentica. The coherence fee—$\dim H^1(\mathcal{N}; \mathcal{F})$—is the formal counterpart of the "cost of composing truth across contexts" asserted in the Res Agentica framework [RA]. It is irreducible (the topology requires it), separable from intermediary overhead (anything beyond $\dim H^1$ is rent, not cost), and payable in typed artifacts rather than institutional trust. The four-way ablation of Section 4.5 gives this claim empirical teeth: the fee is not a metaphor but a measurable quantity with a sharp threshold (generic instructions below the threshold accomplish nothing; typed bridge concepts at the threshold close the cycle).

Connection to the companion papers. In the abelian linearized regime, this paper's cellular H^1 coincides with the Čech obstruction of [SCPI] (Remark 3.1). The identification/specification decomposition of Section 4.5 instantiates the three-gate diagnostic sequence of [SCPI]: the discovery experiment passes Gate 1 (correct obstruction class identified); the closed-loop failure straddles Gates 2 and 3—the LLM-proposed specification canonizes one agent's existing local policy (non-conservative, Gate 2) and uses natural-language ambiguity where the hand-crafted bridge uses a precise function (not explicitly definable in the shared vocabulary, Gate 3). Meanwhile, [SP] reports $H^1 = 0$ for sentence-transformer embedding alignment, and the present paper reports $H^1 \neq 0$ for schema-mediated LLM coordination. Together, these results locate the first computably nontrivial obstruction at the structured-output interface, even when internal representation spaces are gauge-equivalent.

Limitations. The experiment uses ten scenarios (five clean, five ambiguous) on a single coordination graph ($\beta_1 = 1$). The structural perfection of the result—100% H^1 prediction accuracy, 30/30 permutation invariance on ambiguous events, 15/15 bridge discovery convergence—reflects the deterministic nature of the phenomenon, not statistical power. The bridge concept repair succeeds on 4/5 scenarios in the live run; the fifth (A05) fails due to agent behavioral error, not topological obstruction. Extension to multi-cycle coordination graphs ($\beta_1 > 1$), non-abelian coefficients, and stochastic (enriched) sheaves remains open.

Reproducibility. All code, prompts, cached API responses, and bridge concept definitions are available at the paper repository. The experiment is reproducible with any three LLM providers and API keys.

References

[HG19] J. Hansen and R. Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3(4):315–358, 2019.

[HG21] J. Hansen and R. Ghrist. Opinion dynamics on discourse sheaves. *SIAM Journal on Applied Mathematics*, 81(5):2064–2089, 2021.

[Cur14] J. Curry. *Sheaves, Cosheaves and Applications.* PhD thesis, University of Pennsylvania, 2014.

[Ghr14] R. Ghrist. *Elementary Applied Topology.* Createspace, 2014.

[Rob18] M. Robinson. *Topological Signal Processing.* Springer, 2018.

[FKPT05] R. Fagin, P. G. Kolaitis, L. Popa, and W. C. Tan. Composing schema mappings: Second-order dependencies to the rescue. *ACM Transactions on Database Systems*, 30(4):994–1055, 2005.

[Mei11] C. Meilicke. *Alignment Incoherence in Ontology Matching.* PhD thesis, University of Mannheim, 2011.

[KTC26] C. Kurisummoottil Thomas and M. Chen. Fundamental limits of quantum semantic communication via sheaf cohomology. *arXiv preprint arXiv:2601.10958*, 2026.

[OAEI] OAEI Campaign. Ontology Alignment Evaluation Initiative: Conference track. `http://oaei.ontologymatching.org/`, 2023.

[SCPI] J. Komkov. Predicate invention under sheaf constraints: Mathematical foundations for compositional discovery. Companion manuscript, 2026.

[SP] J. Komkov. The SHEAF protocol: Topological diagnostics for heterogeneous multi-agent coordination. Companion manuscript, 2026.

[RA] J. Komkov. *Res Agentica*: The political economy of machine testimony. Companion manuscript, 2026.

[Ext] J. Komkov. The coherence fee: Extended version with mixed-rank initiality and chase-confluence connections. In preparation, 2026.

A Initiality of the minimal rank-1 repair

We prove that the minimal repair of Theorem 5.2 is initial among rank-1 admissible repairs in the linearized regime.

Theorem A.1 (Initiality). *Let $\mathcal{R}_{\min} = \{\sigma_1, \ldots, \sigma_r\}$ with $r = \dim_R H^1$ be a minimal rank-1 repair (Definition 5.1) that kills $H^1(\mathcal{N}; \mathcal{F})$. For any other rank-1 repair $\mathcal{R}' = \{\sigma'_1, \ldots, \sigma'_m\}$ with $m \geq r$ that kills H^1, there exists a surjective morphism $\varphi : \mathcal{R}' \twoheadrightarrow \mathcal{R}_{\min}$ in the repair category (i.e., each 2-cell of $\mathcal{R}'$ factors through a 2-cell of $\mathcal{R}_{\min}$). When $m = r$, the factorization is an isomorphism.*

Proof. Work over R with all stalks free. By Theorem 5.2, $H^1 = \operatorname{coker}(\delta^0)$ has dimension r.

Step 1 (Minimal repair is a basis). Each $\sigma_i \in \mathcal{R}_{\min}$ contributes a rank-1 coboundary component $\delta^1_{\sigma_i} : C^1 \to \mathcal{F}(\sigma_i) \cong R$. The condition that $\mathcal{R}_{\min}$ kills H^1 means $\{\operatorname{im}(\delta^1_{\sigma_1}), \ldots, \operatorname{im}(\delta^1_{\sigma_r})\}$ span a complement to $\operatorname{im}(\delta^0)$ in C^1—equivalently, the projections $\{\overline{\delta^1_{\sigma_i}}\}$ to $\operatorname{coker}(\delta^0)$ form a basis for ${H^1}^* = \operatorname{Hom}_R(H^1, R)$.

Step 2 (Any repair spans the same dual.) Let $\mathcal{R}'$ be another rank-1 repair that kills H^1. Its coboundary components $\{\delta^1_{\sigma'_j}\}$ must also span H^{1^*} (otherwise H^1 is not killed). Since $\dim H^{1^*} = r$ and $|\mathcal{R}'| = m \geq r$, the projections $\{\overline{\delta^1_{\sigma'_j}}\}$ surject onto H^{1^*}.

Step 3 (Factorization). Express each $\overline{\delta^1_{\sigma'_j}}$ in the basis $\{\overline{\delta^1_{\sigma_i}}\}$: $\overline{\delta^1_{\sigma'_j}} = \sum_i a_{ji}\, \overline{\delta^1_{\sigma_i}}$ with $a_{ji} \in R$. Define $\varphi(\sigma'_j) = \sigma_i$ where i is the index of the leading nonzero coefficient (after Smith normal form on the coefficient matrix (a_{ji})). The surjectivity of $\{\overline{\delta^1_{\sigma'_j}}\} \twoheadrightarrow H^{1^*}$ ensures φ is surjective: every basis element of H^{1^*} is in the span, so every σ_i is in the image.

Step 4 (Uniqueness when $m = r$). When $m = r$, the coefficient matrix (a_{ji}) is square and invertible over R (since both sets are bases for H^{1^*}). The factorization φ is therefore an isomorphism: the two repairs differ only by an R-linear change of basis in H^{1^*}, which corresponds to choosing a different cycle-basis representative for each generator. □

For mixed-rank primitives ($\operatorname{rank} \mathcal{F}(\sigma) > 1$), the factorization requires a matroid-theoretic argument on the generator lattice; see [Ext].