The Limits of Structural Consistency: Technical Supplement

Wuming Pan
Sichuan University

This supplement supports the short methods and negative-results article. It records mathematical controls, stage provenance, model settings, reference verification, all final paired count replies and the status of exploratory case explanations. It does not append the earlier paper's general identification theory or pool its experiments into a new benchmark. This revision makes no new model calls.

S1 Mathematical scope and constructed controls

All arguments concern finite carriers and one fixed semantic contract. The implementation handles an ordered binary relation. The Token-map formulation (Pan, 2026) also permits fixed relation markers for other arities, but the model evidence here does not validate a general natural-language semantic parser. A source variable not occurring in any record remains in its carrier. Its images are counted, including isolated target entities. Two occurrences of the same record do not create two facts.

For an arbitrary source map, its kernel partition is unique. Quotienting identifies exactly variables with the same image, retains fixed markers and deduplicates resulting records. The induced quotient map is injective. Conversely, composing an injective quotient map with the quotient projection has the prescribed kernel. This bijection proves the classical sum identity (Lovász, 1967), with multiplicity when different partitions give isomorphic quotients. Source isomorphism produces a bijection on maps into every target, justifying reuse of one model-facing coordinate. These are structural derivations, not weights fitted to observed errors.

S1.1 Three levels are genuinely different

For the loopless outstar on six entities, $(3, 1, 2)$ has zero residual but cannot be realized: $I = \sum_v d_v(d_v - 1)$ is always even. Thus passing one row need not imply realizability. This constructed vector is deliberately caught by a basic parity check; it is not presented as defeating every semantic baseline. The zero-residual wrong-world examples occupy a different set: they are exactly realizable vectors belonging to another target and pass every universal constraint that is applied to that target's exact answers.

The coherent-alternative proposition applies only to vectors exactly generated by one admissible structure. If a model uses differing facts, mapping conventions or arithmetic across questions, it can leave that set and become detectable. Constancy across repetitions does not establish joint realizability. Likewise, a few counts matching one candidate world do not establish that all of a model's answers arise from that world.

S1.2 Outstar quotient and pairwise witnesses

For source records $(0, 1), (0, 2)$, the five variable partitions are

$$0|1|2, \quad 0|12, \quad 01|2, \quad 02|1, \quad 012.$$

The last three introduce loops and vanish under the loopless contract. The first two give the injective outstar and edge, each with coefficient one. For the symmetric reported vector $(6, 0, 4)$, pairwise realizability is witnessed by the following simple graphs, with isolated vertices added to reach six:

Matched pair	Undirected graph	Full (H, I, e)
(H, I)	Matching of size three	$(6, 0, 6)$
(H, e)	Two-edge path	$(6, 2, 4)$
(I, e)	Matching of size two	$(4, 0, 4)$

The proposed full vector has residual two and no realization. For directed four-entity graphs, $(4, 0, 2)$ similarly has witnesses given by a directed four-cycle, two outgoing edges from a common center, and a directed two-edge path. These are software and mathematical controls, not native model cases.

S2 Evidence provenance and units

Table 1 maps the paper's stage labels to their original versions. Later manuscript versions copied earlier data; copies are not new runs. In particular, V5 implemented and validated the general compiler and specified the later pilot, but produced no new native model responses. The present article is a retrospective synthesis and offline audit. None of its newly generated tables is a new confirmation sample.

Table 1: Stage provenance. These units are neither pooled nor treated as a factorial comparison. Earlier exploratory runs and unused confirmation payloads remain separately archived.

Stage	Original archive	Units and scope used in this article
I	V2	Three Qwen models, each with 1824 calls including 112 domain controls. Structure pairs and targets, not calls, are scientific units.
II	V3	Calibrated main matrices of 704 questions per admitted model; sixteen consistency targets and 32 separation pairs per interface. Rendering repeats are not independent targets.
III	V4	32 main targets, plus eight separately reported stress targets. The main/stress/budget protocol has 848 responses; separate lifting pilots are not included in this article's comparisons.
IV	V6	Sixteen synthetic development targets and eight knowledge contracts per model; 440 and 480 responses per layer/model. The knowledge contracts use four underlying entity domains.
V	V7	Two 72-call counting gates; only Kimi proceeds to 120 screen replies and 90 paired replies, totalling 354. Five paired domains, one enriched and four unscreened.

The main article foregrounds stages III and V because they directly compare decision utility and matched information conditions. Stages I, II and IV explain calibration and endpoint limitations. The choice to organize them this way was made after their results were known. The individual protocols' prospective status is not transferred to the present cross-stage interpretation. Domain sizes, models, output budgets and predicates changed across stages; a difference cannot be attributed solely to one change.

S2.1 Model and output settings

Stage I uses Qwen2.5-Instruct 0.5B, 1.5B and 7B with greedy decoding, a 16-token answer limit, a 2048-token context budget and a strict integer parser. The 1.5B model was designated primary before its formal run. All 5472 formal replies were parseable, but all three domain-control scores were zero. Its failure is conditional on this direct-answer interface.

Stages II–V use hosted deployment names and preserve returned service identifiers. Astra's relevant returned identifier is `gpt-6-astra-2026-09-03`. Kimi requests use `Kimi-K3.1-Pro`, with `Kimi-K3.1` returned in the final study. DeepSeek uses `DeepSeek-V4-Pro`. No immutable provider-weight hash is available. Stage-specific full request bodies are retained; model names do not justify pooling observations collected under differing interfaces or budgets.

Stage II sets Astra's reasoning effort to medium and its completion limit to 8192; Kimi and DeepSeek use an 8192 limit without a reasoning override. Stage III keeps Astra's medium setting and 8192 limit, except for a separately specified 4096-token comparison on eight fixed main targets. No temperature or sampling

seed is sent in that protocol. These shorter-budget calls are independent scheduled requests, not retries of unsuccessful long-budget answers. Stage IV uses an 8192 limit without an explicit temperature or reasoning override. Its synthetic calibration scores are 32/32 for Kimi and 18/32 for DeepSeek; both remain diagnostic strata in that design. This differs from stage V's counting-admission gate and is another reason not to pool their outcomes.

The final study permits up to 8192 completion tokens without an explicit temperature or reasoning override. There are no browsing, execution or counting tools in its model calls. Each coordinate is a separate stateless request with the same three fictitious counting examples. Identical prompts are deliberately repeated, but independent factual samples or independently distributed errors are not assumed. Invalidity includes truncation and is distinct from an incorrect numerical answer. Strict parsing requires a normal stop and ASCII digits after stripping whitespace; the fact screen instead accepts exactly YES, NO or UNKNOWN.

Stage IV retained the twelve-main-repeat predicate C, coordinate-triple stability S, and a first-vector flag D. The truth-labelled endpoint additionally required a wrong repeated main answer W. In its knowledge layer the primary predicate was $C \cap S \cap D$. Stage V prospectively uses $S \cap D$, without the twelve-repeat requirement, and changes the source family and factual domains. Its result cannot be represented as another observation of stage IV's identical endpoint.

S3 Reference relation, selection and completeness

The archived 2023 and 2024 national Census files define the county-adjacency relation (U.S. Census Bureau, 2023; 2024). An independent parser agrees with the generator on all 864 ordered pair memberships across the 24 domains. The files include self-pairs, which are explicitly removed from the experimental relation. Every remaining domain graph is symmetric. Counties outside the declared six-entity domain cannot be assigned to a variable, even if they neighbor an included county. Census adjacency includes point contacts and boundaries within water; a land-only convention is not substituted.

The deterministic domain rule begins with a county and three within-state neighbors and adds two second-neighborhood counties not adjacent to the center. It requires branching, both positive and negative pairs, stable names and pair memberships across releases, and selects one domain from each of 24 states. Four controls are designated before any calls. Each other domain has one predetermined true and one false pair, not a pair chosen after observing model answers. Qualification requires three identical valid wrong replies to at least one of these pairs. Selection order is fixed, and insufficient qualifiers are not replenished. It uses no aggregate counts or residuals, but the fact and aggregate questions are logically related, so the enriched stratum is not an unbiased factual-error sample.

All 624 potential prompts were checked against their source and contract: 72 calibration, 120 screening and 432 candidate A/B requests. Only selected A/B payloads were dispatched. Independent direct enumeration and degree formulas agree on 96 reference count coordinates across the 24 factual and eight calibration domains. These are input checks, not 624 observed model answers.

The final phase totals are retained in full:

Phase	Returned	Valid	Exact	Truncated
Kimi counting gate	72	72	72	0
DeepSeek counting gate	72	72	16	0
Kimi factual screen	120	118	115	2
Kimi A/B	90	85	69	5

DeepSeek's later phases were not executed; they are not zero-yield observations. The final study's five A and five B first vectors are complete, although three A domains contain later truncations. The A domains with truncated replies remain in the primary denominator. Coordinatewise modes can be computed from available valid replies without satisfying the stronger stability predicate.

S4 Implementation, verification and preservation

The supplied `verification/` directory contains byte-identical copies of the general binary quotient compiler and two outstar modules. General certificates reconstruct canonical sources, quotient weights and coefficient maps. Outstar certificates reconstruct the five kernel partitions and the loopless reduction. Neither verifier reads a target relation table or evaluation counts. A contract hash records the stated interpretation, not proof of model understanding or authentication of provider weights.

Earlier deterministic validation used 579 targets, 9264 independently counted coordinates, 1737 ungrouped identities, 9264 single-error injections and 512 coherent alternative-target controls. The outstar preflight covered all 4166 loopless directed targets on zero through four entities and 12498 single-coordinate injections. These are preserved software-validation records; This article does not report them as newly run model experiments. Its offline audit independently reparses all 354 final-stage raw responses, recomputes screening eligibility and paired predicates, verifies the two displayed native certificates, recounts the reference and alternative G01 graphs, and checks the constructed pairwise witnesses.

Operational records. The original archives retain all input freezes, request bodies, raw returns, transport attempts and operational amendments. Local freezes are not public preregistration. No returned wrong, malformed or truncated answer is replaced by a new answer. Recovery is restricted to never-returned transport failures. Accounting proxies are not invoices, and no cost or budget history is used as scientific evidence. No operational record is removed from the original archive merely because the short main article omits it.

Offline reproducibility. Run `python3 audit_evidence.py` to validate the bound snapshots and regenerate the full paired table. It needs only the Python standard library and makes no network requests. The returned-response audit distinguishes read-only historical analysis checks from raw reanalysis of the final stage. The artifact also includes the complete G01 transcript and compiled joint system. The accompanying README specifies the included evidence and reconstruction commands; historical archive labels are provenance, not additional observations.

S5 All frozen factual domains

The 24 six-entity domains were fixed before the final study. Asterisks mark the four unscreened controls. Only G01 qualified through factual screening; G01 and the four controls received A/B count questions. Other domains are not A/B observations. Counts are reference values, not model answers. All 144 entities are distinct; thirteen unlabelled graph types occur, so named-domain distinctness does not imply distinct shapes.

ID	Declared entities	H	I	e
G01	Bullitt County, KY; Breckinridge County, KY; Marion County, KY; Jefferson County, KY; Spencer County, KY; Hardin County, KY	32	20	12
G02	Howard County, MD; Anne Arundel County, MD; Washington County, MD; Montgomery County, MD; Baltimore city, MD; Carroll County, MD	16	8	8
G03	Iberia Parish, LA; Vermilion Parish, LA; St. Mary Parish, LA; Ascension Parish, LA; Lafourche Parish, LA; Iberville Parish, LA	16	8	8
G04	White Pine County, NV; Churchill County, NV; Humboldt County, NV; Elko County, NV; Eureka County, NV; Lincoln County, NV	24	14	10
G05*	Navajo County, AZ; Apache County, AZ; Cochise County, AZ; Pinal County, AZ; Coconino County, AZ; Gila County, AZ	24	14	10
G06*	Androscoggin County, ME; Lincoln County, ME; Sagadahoc County, ME; Somerset County, ME; Cumberland County, ME; Kennebec County, ME	50	34	16
G07	Vinton County, OH; Highland County, OH; Fairfield County, OH; Hocking County, OH; Meigs County, OH; Jackson County, OH	16	8	8
G08	Clay County, WV; Lincoln County, WV; Fayette County, WV; Nicholas County, WV; Kanawha County, WV; Roane County, WV	52	36	16

ID	Declared entities	H	I	e
G09*	Lamoille County, VT; Addison County, VT; Orleans County, VT; Chittenden County, VT; Essex County, VT; Washington County, VT	36	22	14
G10	Humphreys County, MS; Coahoma County, MS; Carroll County, MS; Sharkey County, MS; Washington County, MS; Yazoo County, MS	26	16	10
G11	Clark County, IL; Shelby County, IL; Douglas County, IL; Coles County, IL; Jasper County, IL; Cumberland County, IL	50	34	16
G12	Chester County, TN; Wayne County, TN; Hardeman County, TN; Hardin County, TN; Crockett County, TN; Henderson County, TN	24	14	10
G13	Wheatland County, MT; Chouteau County, MT; Sweet Grass County, MT; Blaine County, MT; Meagher County, MT; Judith Basin County, MT	36	22	14
G14	Richland County, SC; Lancaster County, SC; Lexington County, SC; Greenwood County, SC; Newberry County, SC; Kershaw County, SC	28	16	12
G15	Vilas County, WI; Taylor County, WI; Ashland County, WI; Oneida County, WI; Forest County, WI; Iron County, WI	22	12	10
G16	Nez Perce County, ID; Clearwater County, ID; Valley County, ID; Benewah County, ID; Latah County, ID; Lewis County, ID	32	20	12
G17*	King County, WA; Mason County, WA; Kittitas County, WA; Island County, WA; Pierce County, WA; Kitsap County, WA	40	26	14
G18	Woodson County, KS; Labette County, KS; Linn County, KS; Wilson County, KS; Greenwood County, KS; Coffey County, KS	26	16	10
G19	Ventura County, CA; San Luis Obispo County, CA; Santa Barbara County, CA; Tulare County, CA; Kern County, CA; Los Angeles County, CA	52	36	16
G20	Mower County, MN; Wabasha County, MN; Dodge County, MN; Houston County, MN; Freeborn County, MN; Fillmore County, MN	16	8	8
G21	Essex County, NJ; Bergen County, NJ; Morris County, NJ; Union County, NJ; Sussex County, NJ; Somerset County, NJ	40	26	14
G22	LaPorte County, IN; Elkhart County, IN; Starke County, IN; Lake County, IN; Porter County, IN; Jasper County, IN	54	38	16
G23	Livingston County, NY; Genesee County, NY; Erie County, NY; Allegany County, NY; Monroe County, NY; Niagara County, NY	38	24	14
G24	Davison County, SD; Hutchinson County, SD; Kingsbury County, SD; Brule County, SD; Sanborn County, SD; Aurora County, SD	30	18	12

S6 Complete final paired count replies

The following table contains every scheduled count response from the five paired domains. A dash denotes a truncated, non-valid reply, not zero. Fixed repeat index does not identify network completion order. The reference column is used for evaluation and was absent from condition A. Each B condition has nine exact replies.

Domain	Condition	Coordinate	Reply 1	Reply 2	Reply 3	Reference
G01	A	H	22	22	22	32
G01	A	I	8	12	12	20
G01	A	E	10	10	10	12
G01	B	H	32	32	32	32
G01	B	I	20	20	20	20
G01	B	E	12	12	12	12
G05	A	H	24	24	–	24
G05	A	I	14	–	14	14
G05	A	E	10	10	10	10

Domain	Condition	Coordinate	Reply 1	Reply 2	Reply 3	Reference
G05	B	H	24	24	24	24
G05	B	I	14	14	14	14
G05	B	E	10	10	10	10
G06	A	H	50	50	–	50
G06	A	I	34	34	34	34
G06	A	E	16	16	16	16
G06	B	H	50	50	50	50
G06	B	I	34	34	34	34
G06	B	E	16	16	16	16
G09	A	H	36	36	36	36
G09	A	I	22	22	22	22
G09	A	E	14	14	14	14
G09	B	H	36	36	36	36
G09	B	I	22	22	22	22
G09	B	E	14	14	14	14
G17	A	H	50	50	–	40
G17	A	I	34	34	34	26
G17	A	E	16	–	16	14
G17	B	H	40	40	40	40
G17	B	I	26	26	26	26
G17	B	E	14	14	14	14

S7 Case provenance and interpretations

S7.1 G01: a screened fact error and compatible counts

The archived entity order is Bullitt, Breckinridge, Marion, Jefferson, Spencer, Hardin. The undirected reference edge set, with zero-based indices, is

$$\{\{0,3\},\{0,4\},\{0,5\},\{1,5\},\{3,4\},\{3,5\}\}.$$

Each edge contributes both ordered records. Marion is isolated but remains an allowed variable image. The positive screening question concerns indices $(3,5)$: Jefferson and Hardin. All three answers are NO although the reference answer is YES. The second predetermined question concerns Bullitt and Breckinridge; all three NO replies to that negative pair are correct.

Removing $\{3,5\}$ gives degree sequence $(3,1,0,2,2,2)$ and $(H,I,e)=(22,12,10)$. The independently enumerated assignments agree with the degree formulas. The exploratory candidate record was created on 7 November 2026 at 16:45:52 UTC, after screening and after the A/B dispatch began, before the A/B answers were inspected for this analysis. It was not part of the initial frozen protocol and does not change any endpoint, selection, request or denominator. Its exact match to repeat indices 2 and 3, and to the coordinate modes, is reported as an exploratory compatibility explanation.

The first vector $(22,8,10)$ is bound to its original indexed responses and has residual four. Its numerical contradiction is real but redundant with the two-coordinate projection check described in the main article. Neither that contradiction nor the matching alternative counts determine an internal model fact set. A cross-question inconsistency may arise from interpretation, changing factual assumptions or computation; the transcript does not isolate one cause.

G17, the unscreened Washington domain, has A mode $(50,34,16)$ instead of reference $(40,26,14)$. Its first vector and all available valid replies satisfy the same row, but two later replies are truncated. This case is retained even though it was not selected for the main narrative. A/B data are too limited to estimate the frequency of coherent factual errors or the prevalence of stable incremental contradictions.

S7.2 A native positive certificate is not a stable-error result

The earlier DeepSeek synthetic edge vector $(7, 4, 0)$ has residual three under

$$\hom(E_1, G) = \mathrm{inj}(E_1, G) + \hom(L, G).$$

On its four-entity domain, the individual upper bounds are $16, 12, 4$, so the values are individually in range. The certificate verifies without target facts. It does not meet the prior study's repeated stability criterion; its existence is not substituted for a zero numerator at that endpoint. The companion Kimi knowledge example had correct main count 3888 and an incorrect auxiliary count 248 instead of 128, yielding residual -600. That example illustrates why rejecting a main answer from any bundle contradiction can reject a correct main answer.

S7.3 A passing vector need not identify a world

The stage-II Kimi cancellation example has reference $(7; 2, 3, 1, 0, 1)$ and report $(7; 2, 2, 1, 0, 2)$. The semicolon separates the ordinary count from five injective quotient coordinates. The reported auxiliaries still sum to seven. This establishes cancellation in the tested row only; unlike the G01 construction, we do not claim a particular alternative realization for this full coordinate list. Both cases are useful precisely because row satisfaction, realizability and reference accuracy remain separate assertions.

S8 An exact model-facing prompt

The following system message and A user message are reproduced from the original G01 ordinary-count request, with line wrapping changed only for typesetting. B adds the complete relation table and a statement that unlisted pairs are false. The artifact retains the original messages for all 354 final-stage responses; the G01 transcript separately collects its six screening and eighteen paired responses.

System message.

> Answer the question under its explicitly stated finite-domain and mapping rules. You have no browsing or execution tool. Reason as needed, but return only the requested answer format.

User message.

> Counting conventions: assignments are ordered. For example, on three fictitious entities p,q,r with only the directed pairs (p,q),(p,r), the number of ordered triples satisfying R(x,y) AND R(x,z) is 4 when variables may coincide, and 2 when x,y,z must be pairwise distinct. The number of ordered related pairs is 2. These fictitious examples are unrelated to the target domain.
>
> Use the U.S. Census Bureau 2024 County Adjacency File definition: two DIFFERENT counties (including county equivalents such as parishes and independent cities) are related if their boundaries share an edge or a point, including an official boundary within a body of water. Adjacency is symmetric. Exclude self-pairs even though the source file includes them. Only the listed six counties are in the assignment domain; a neighbor outside this domain cannot be assigned to a variable.
>
> DOMAIN:
>
> 1. Bullitt County, KY
> 2. Breckinridge County, KY
> 3. Marion County, KY
> 4. Jefferson County, KY
> 5. Spencer County, KY
> 6. Hardin County, KY
>
> No relation table is supplied. Use your factual knowledge under the definition above.
>
> How many ordered triples (x,y,z) from this domain satisfy R(x,y) AND R(x,z)? Variables may take the same entity; there is NO pairwise-distinctness requirement. Return only the nonnegative integer, in ASCII digits.

References

László Lovász. Operations with structures. *Acta Mathematica Academiae Scientiarum Hungaricae*, 18: 321–328, 1967.

Wuming Pan. Token Space: A category theory framework for AI computations, 2026. URL https://arxiv.org/abs/2404.11624v4. arXiv:2404.11624v4.

U.S. Census Bureau. County adjacency file, 2023. URL https://www.census.gov/geographies/reference-files/2023/geo/county-adjacency.html. Accessed 8 November 2026.

U.S. Census Bureau. County adjacency file, 2024. URL https://www.census.gov/geographies/reference-files/2024/geo/county-adjacency.html. Accessed 8 November 2026.