Step-Addressed Parts: Where a Checkable Mind Stops, Errs and Grows

Ben Horn, Prometheus7
4 November 2026

Abstract

Oracle7 is a mind that answers from a store through steps anyone can read, and declines when it cannot. It contains no language model. Four limits remain in its current build: it does not write open-ended text, it declines most fact questions, its growth has not been shown to compound, and a residue of its answers are misreadings. We argue that the four share one cause. Oracle7 grows by *parts* attached to whole readings, but it fails at *steps*, and a step has no address. We test the premise on Oracle7 itself. On 1,000 KQA Pro validation questions, the reader's readings were walked step by step against the labelled readings. Of the questions that formed no reading, only 2 of 509 lit nothing; 85% stopped at two checks, and 65% at one: the judge refusing every shape the lighting formed. Of 52 wrong readings that leave the labelled reading, 35 do so in the first two steps. Below the score threshold, 317 of 379 readings are right. We propose *step-addressed parts*. The judge's refusal names the step it refused; parts attach to step addresses, each examined at its step alone; a reading is stated when every step passes its own check; and a document is a shape whose steps are discourse moves. Relational geometry proposes candidate fills at a step, and the exact store checks them. We pre-register the test that decides the proposal.

1. Introduction

Oracle7 is built to receive (Constitution, article 1). When it meets a question it lights the question's words in its store, lets the light run along the relations the words name, and reads a program of steps off the shape where the lights meet. It runs the program exactly, judges whether every word of the question was used, and answers, holds the reading back, or declines. Every answer carries its receipt: the steps, the store records they touched, and the checks they passed. A question it cannot answer becomes a *hole*, a fact a person tells it becomes a *claim*, and each night it *sleeps*: it goes back over the day's holes, searches its stores, builds candidate repairs and examines each one before keeping it. A kept repair is a *part*: a signed file with its own exam, admitted only if it passes questions it has never seen with no new wrong answer.

The measure of such a mind is what it does with what it knows (article 6). Measured that way, the current build is strong where it answers: on the gate of 5,128 items its 11.0 head lost no right answer and added no wrong one against the released 10.3.2 (a later change, made 4 November, holds back 15 passage answers that were right guesses); on 10.3.2's last KQA Pro measurement it answered 1,330 of 3,000 questions with 16 wrong. Four limits remain.

- **L1, writing.** It composes from cited readings and does not write open-ended text.
- **L2, declines.** It declines most fact questions it is asked; 92 of the 100 reference questions in the sleep trials.
- **L3, compounding.** Seven trial nights admitted 6 to 22 parts while the reference stayed at 5 of 100. Growth was not shown to compound.

- **L4, misreadings.** A residue of answers are wrong: on the gate, 305 judged wrong, the largest classes in unit and formula answers, multi-step physics and follow-up questions asked without their conversation.

This paper asks whether one mechanism can address all four, explains why each exists, and tests the shared premise on Oracle7.

Contributions. (i) A diagnosis that places the four limits at one missing element, the step address (section 4). (ii) A deterministic credit measurement on Oracle7: the first step at which each reading leaves the labelled reading, and the check at which each decline stops (section 5). (iii) The step-addressed part, its four consequences, and the role of relational geometry in it (section 6). (iv) A pre-registered test (section 7).

2. The line of work this builds on

The design draws on the author's earlier technical work, summarised here because each piece supplies a component.

The construction kernel and the hyper-object thesis (August 2026). A chart over typed constructions compiles English into clause programs, proved by forward chaining; its lambda form makes inference beta-application. Fold passes raised its parse rate on an everyday corpus from 0% to 98.2% while every earlier battery stayed green, and sealed prospective tests passed 55/55. The thesis behind it: a language's redundancy can be enumerated as a finite rule set, and the unparsed *residue* is the acquisition signal. That is the first statement of the loop this paper extends: residue names what is missing.

Relational geometry (August 2026). Holographic reduced representations (Plate, 1995) bind and superpose vectors by circular convolution. Two results set their role. First, comparing vectors in the projected space without inverting the projection had produced five false negatives; with cleanup back to the dictionary, pairwise discrimination reached 94.4% of the exact space's ceiling at 230x compression (AUC 0.9586 against 0.9856), and did not improve with dimension: the missing operation, not the dimension, was the bottleneck (cf. Johnson and Lindenstrauss, 1984). Second, *identity wants random codes and navigation wants relational codes, and one vector cannot do both*. Superposed and unbound once, HRR recovered 4.65 items per operation at 16 subjects against 1.0 for exact lookup; relational codes made cleanup navigable, 12 points better than random at 32 subjects. The law: geometry is a proposer of high throughput and low precision, and must feed an exact verifier.

The fold line. Foldtoys solves 131 optimisation problems across 13 domains with the same four operations (bind, unbind, permute, cleanup); Qum shows a four-operation algebra on 256-dimensional complex vectors reaching a fixed point of its own parameters; ProseFold measures a voice with nine parameters and, inside Oracle7, searches faithful compositions toward it (205/205 faithful on 20 topics).

Place. The spatial corpus treats every entity as placed content (position, form, neighbours, bindings), so that place, not token, is the primitive. The Tree of Life scaffold gives Oracle7 its addressing, with bounded cleanup.

The K-line diagnosis (August 2026). Every frozen-evaluation failure of the phase-3 runtime was at ingress: "ingress geometry is advisory, not authoritative". Geometry informed the route but a recogniser of wording committed it, and paraphrase is invisible to a recogniser in principle.

The specialist network (September 2026). Specialists are nodes; checked paths between them are kept; with K nodes the reachable paths grow faster than K. Sealed: 28 right, 0 wrong, 3 missed of 31, against 6 right and 6 wrong for the best single specialist. Compounding is reach from combinations (the author, 23 October 2026).

Two laws from measurement. *Reach is not answer* (September 2026): a join that reached the gold answer on 18 of 60 declined questions would have answered 34 of them wrongly. And the *core failure locator* (September 2026): of 600 open questions, 69% stopped before any search of the store.

Composition. The evidence-bound speaker showed that restating a claim by swapping its frame is unsafe (0 of 100), so writing must be composition of cited readings. The essay-residue protocol compared a reference essay with Oracle7's own on the same outline, over discourse moves: Oracle7 led on warrant density and lacked seven moves: works engaged, objections handled, definitions offered, generalisation, cross-references, return of the arc, external thinkers.

3. The four limits, measured

	Limit	Measurement on the current build
L1	Writing	ProseFold: 205/205 compositions faithful, beats a random composition 20/20 on voice distance. Quality never blind-scored; automatic metrics once rose while blind scores fell (2.13 to 1.07)
L2	Declines	Reference set: 7 right, 0 wrong of 100 on a first ask. A trial day of 140: 10 answered, 26 held back, 41 withheld, 63 asked for clarification
L3	Compounding	Seven trial nights: parts 6 to 22, reference flat; sleep's store search filled 0 of 934 holes. Reading rules keyed by step sequence: 0 of 80 admitted
L4	Misreadings	Gate, 5,128 items: 1,172 right, 305 wrong, 31 false yes (judged against each item's expected answer). Passage reader: 1 wrong of 290 after 11-03's fixes

4. Diagnosis: the step has no address

The design review was run as a structured panel of eight positions, each derived from one part of the work in section 2 and from Oracle7's measurements: the architecture's own terms; construction and meaning; relational geometry; place; credit and exams; composition; the network; and an examiner. The full record is in `PANEL.md`. Its round-one findings, why each limit exists:

L2. A reading is accepted or refused whole. The panel first placed this at the door, by analogy with the K-line diagnosis; section 5 moves it to the judge.

L4. Steps run with no check on their own output; only the final answer is judged. A step that reads a relation backwards, or takes the clause a subject belongs to for another, still yields a well-formed answer. In the terms of section 2: the head has a proposer at every step and a verifier only at the end.

L3. Parts attach to the end of a reading (a formula, a table) or to a whole shape (a reading rule). A part at the end combines with nothing inside the reader. A rule over a whole shape must be right about every step at once, and one

score cannot say which step was wrong; that is why 0 of 80 were admitted.

L1. A document has no steps. The moves the essay-residue run found missing are exactly the steps a document's shape would have.

What the four share. Each is the absence of a step-level place: a step with no part to widen it (L2), a step with no check of its output (L4), a part with no step to attach to (L3), a document with no steps (L1). Oracle7's unit of growth is attached to whole readings; its unit of failure has no address.

5. Experiment E1: credit at the step

Question. Do failures concentrate at a few steps and checks? If they don't, step addresses buy little.

Method. 1,000 questions from KQA Pro validation (Cao et al., 2022), shuffled with seed 2026. KQA Pro labels each question with the program of steps that answers it; we use the labels only to grade. For each question the reader forms its own reading with the score threshold lifted, as the head's held-back path does. A reading is walked against the labelled program step by step; the first step at which the function or its inputs differ is the step the outcome is credited to. A question with no reading is credited to the check whose stop counter moved. The tree is the 11.0 development head; the hash seed is fixed. Code: `step_credit.py`.

Results.

Outcome	n
Reading formed	491 (421 right, 70 wrong)
… stated at the threshold	112 (104 right, 8 wrong)
… below the threshold	379 (317 right, 62 wrong)
No reading	509

Where the 509 stopped	n	share
The judge refused every shape the lighting formed	332	65.2%
The asked kind does not hold	100	19.6%
The opening's kind vetoed it	32	6.3%
Shapes disagreed	30	5.9%
Nothing lit	2	0.4%
No counter moved	13	2.6%

The 70 wrong readings	n
Same steps as the label, wrong answer	18
A different function at a step	31

The 70 wrong readings	n
Different inputs at a step	20
A missing step	1
First divergence at step 0 or 1	35 of 52

Most frequent first divergences among wrong readings: the wrong thing found (`Find` inputs, 9), an attribute queried where the label filters on it (8), one thing found where the label finds all of a kind (5). Median score: right 0.584, wrong 0.494.

What E1 shows.

1. *Declines are not at the door.* Nothing lit for only 2 of 509 (13 more moved no stop counter). 85% stop at two checks; 65% at one, the judge. The K-line pattern has moved: the lighting is now authoritative, and the recogniser that commits is the judge, which accepts or refuses a whole shape and says only "refused".
2. *Wrong readings leave early.* 67% of the diverging ones do so at the first two steps: finding the question's thing and choosing the first relation or attribute. These step kinds are shared by almost every shape.
3. *The threshold is the wrong instrument.* Below it, 5 in 6 readings are right. A single score separates right from wrong readings on average but not at any floor inside a shape (L3's 0 of 80).
4. *Not every wrong answer is a misreading.* 18 run the labelled steps and still answer wrongly. Their failure is in the store or the comparison, and step addresses do not reach them.

6. Step-addressed parts

Definition. A *step address* is a step kind (find, relate, query an attribute, filter, compare, count, verify, and the moves of a document) placed on the Tree of Life scaffold, so that addresses of the same kind are neighbours. A *step-addressed part* is a part that attaches to one address. It carries an exam for that step alone, drawn from readings whose other steps are already confirmed, and it is checked at that step by the store.

Four consequences.

1. *The judge names the step* (L2). The judge already aligns each question word to a step. When it refuses a shape, the refusal records which step's words were not covered. The decline becomes a hole with an address, and the next turn or the next night can fill that address. 332 of 509 declines in E1 are of this kind.
2. *Parts at addresses combine* (L3). A part at the find address (an alias that links a name to its thing) or the relate address (a word that reads as a relation) serves every shape that uses the address. Reach then grows with combinations of step parts across shapes, which is the compounding of the specialist network carried inside the reader. Sleep fills addresses from two sources it already has: confirmed held-back readings (5 in 6 right) and the fix loop's receipts.
3. *Every step is checked where it runs* (L4, and L2 through the threshold). A reading is stated when each of its steps passes its own address's check, not when one score clears a floor. The checks are the store's: a found

thing must be the thing the words name; a relation must hold in the direction read; a filter's value must be of the attribute's kind. The 35 early divergences of E1 are where these checks would act first.

4. *A document is a shape of moves* (L1). The moves the essay-residue run found missing become addresses: define, engage a work, raise an objection and answer it, generalise, cross-refer, return. Each move is filled by cited readings and composed under ProseFold. Quality is claimed only from blind scoring.

The role of geometry. At a step address, candidate fills (which thing, which relation) are proposed together by one superposed unbind over relationally encoded neighbours, which section 2 found navigable, and every proposal is checked by the exact store. Identity checks use the random codes; navigation uses the relational ones. This is the proposer-verifier law used where the architecture placed it, and it bounds cost: one unbind per step instead of a search per candidate, within the scaffold's bounded cleanup.

What it does not change. Nothing about questions is written by hand: addresses are the reader's own step kinds, refusals come from the judge's alignment, exams come from confirmed readings. Every part goes through the fix loop: it is kept only if it removes a wrong or adds a right answer and loses no right one and adds no false yes on the gate (*reach is not answer*).

7. Pre-registered test

To be anchored before the run, and run once.

Id	Hypothesis	Test	Bar
S1	Refusals carry addresses	Share of judge refusals that name a step, on fresh questions	at least 90%
S2	Reach rises with step parts	Reference set reach before and after sleep admits step parts, paired, exact McNemar	$p < 0.01$
S3	No new wrong	Wrong answers on the reference set and the gate	0 new
S4	No new false yes	Must-decline controls	0 new
S5	Combination	Reach per step part, counted across distinct shapes	rises with the number of step parts
S6	Exams hold	Planted bad step parts	all refused

The 18 same-steps-wrong readings are traced separately and reported. Writing (L1) is outside S1 to S6; its test is the blind packet.

8. Threats and limits

- **The label is one reading among several.** 114 right readings differ from their label, so "first divergence" overstates divergence for some wrong readings. Credit is reported per step kind, where this matters less.
- **KQA Pro's step language is borrowed.** Oracle7 maps its own step names one to one onto it (11,797 of 11,797 programs give the same answer); the address set is Oracle7's.

- **Simulated panel.** The panel's positions are derived from written work and measurements, not independent reviewers. The test in section 7 is what decides.
- **Writing is the weakest claim.** No number here supports L1 beyond faithfulness.
- **Box.** All measurements ran on one four-core server with the hash seed fixed; repeated questions in one process change some answers, so every measurement is a first ask in a fresh process.

9. Conclusion

The four limits are one: Oracle7's parts and checks live at the level of whole readings, and its failures happen at steps. E1 shows the failures concentrated where a step address would act: two thirds of declines at one check, two thirds of diverging misreadings in the first two steps, five in six held-back readings right. Step-addressed parts put the hole, the part, the exam and the check at the same place, which is also where the network's compounding can happen inside the reader.

References

- Cao, S. et al. (2022). KQA Pro: A dataset with explicit compositional programs for complex question answering over knowledge base. ACL 2022.
- Huddleston, R. and Pullum, G. K. (2002). The Cambridge Grammar of the English Language. Cambridge University Press.
- Johnson, W. B. and Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. Contemporary Mathematics 26.
- Kanerva, P. (2009). Hyperdimensional computing. Cognitive Computation 1(2).
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions. Psychometrika 12(2).
- Montague, R. (1970). Universal grammar. Theoria 36.
- Plate, T. A. (1995). Holographic reduced representations. IEEE Transactions on Neural Networks 6(3).
- Prometheus7 (2026). The Oracle7 build constitution, version 2. prometheus7.com/whitepaper.
- Prometheus7 technical notes (2026): construction kernel M13e2-e4; HRR original and projected form; relational geometry; K-line geometric ingress; specialist network; reach is not answer; core failure locator; evidence-bound speaker; essay residue; ProseFold; coverage 10.2; fix loop. Available in the Oracle7 repository.