Typed Inference Dynamics: Auditable State Transitions for Mathematical Reasoning

Behnood Mashayekhi [1]

Abstract

Hard mathematical problem solving evolves through persistent partial states: paths are pruned or reactivated, constraints are deposited, subgoals close, and frames change what the next solve can see. Final answers, linear rationales, and scalar process scores flatten this motion. We introduce *typed inference dynamics* (TID), a state-transition substrate in which a helper proposes typed operations over candidate paths, constraints, subgoals, frames, and determined values; a deterministic reducer accepts, rejects, and logs the operation trace; and the accepted state is rendered to a frozen solver or distilled into adapters. In the original loop, full MATH-500 Level 5 ($n = 134$) within-loop cumulative coverage improves from $35/134$ to $53/134$ at 7B ($+13.4\,\mathrm{pp}$, 95% CI $[+7.5, +19.4]$) and from $36/134$ to $59/134$ at 14B-AWQ ($+17.2\,\mathrm{pp}$, $[+11.2, +23.9]$); on OlympiadBench, the 14B-AWQ carrier improves from $4/29$ to $11/29$, while AIME is null at the reported T2 endpoint. These cumulative endpoints are pass-by-budget coverage inside the original loop, not evidence that typing alone caused the lift; fresh fixed-horizon controls show that repeated attempts explain much of the discovery gain, detect no net any-turn advantage of full TID over a schema-lite state updater at $n = 134$, and expose a discovery-versus-retention gap that terminal and frozen-selector endpoints only partly realize. In the fresh cost profile, D uses 488,512 total tokens, about $1.98\times$ repeated sampling (B) and $1.05\times$ schema-lite updating (C*), while a 5,000-example transition corpus lifts a 7B LoRA from $6/90$ to $33/90$ passes on a hard harness. The evidence supports typed inference dynamics as an auditable, trainable transition layer between informal rationales and formal proof objects, with retention and state-authority calibration as the next bottlenecks.

[1]Amesha. Correspondence to: Behnood Mashayekhi <behnood@amesha.co>.

The 3rd *AI for Math Workshop* at the 43rd International Conference on Machine Learning (ICML), Seoul, South Korea, 2026.
Copyright 2026 by the author(s).

1. Introduction

Mathematical reasoning systems are usually observed through objects that are too flat for the process they try to supervise. A final answer says whether a path ended correctly, but not which subpath should be retired. A chain-of-thought gives a line of text, but not a persistent field of active and inactive candidates. A process reward model scores steps (Lightman et al., 2023; Wang et al., 2024), but a scalar score does not say what state should survive into the next solve. Tree search expands a frontier (Yao et al., 2023a), while formal proof systems provide validity only after reasoning has been translated into a proof language (Wu et al., 2022; Jiang et al., 2023). Between these objects lies an AI4Math-relevant substrate: a structured state whose transitions can be inspected, ablated, credited, and learned.

This paper studies that substrate as *typed inference dynamics*. The state $\mathcal{F}_t$ contains candidate solution paths, active constraints, open subgoals, the current frame, determined values, and audit metadata. A helper proposes typed operations over that state: retire a path, suspend it, reactivate it under a condition, merge paths, transform a construct, reframe the field, deposit a constraint, open or close a subgoal, or record a warning. A deterministic reducer accepts only locally valid operations and records rejected operations. The accepted state $\mathcal{F}_{t+1}$ is then rendered into a concise natural-language view for a frozen solver, and the solver outcome supplies downstream empirical credit.

The transition is the unit of study:

$$(\mathcal{F}_t, \tau) \mapsto (e, r, \Delta\mathcal{F}), \qquad \mathcal{F}_{t+1} = \mathcal{F}_t \oplus \Delta\mathcal{F}.$$

Here τ is a candidate path, e is an event type, r is the reason, and $\Delta\mathcal{F}$ is the proposed field-state change. This notation is deliberately modest: it is not a proof certificate and not a new model architecture. Its value is that path-status changes, accepted constraints, rejected operations, rendered observations, and downstream outcomes become

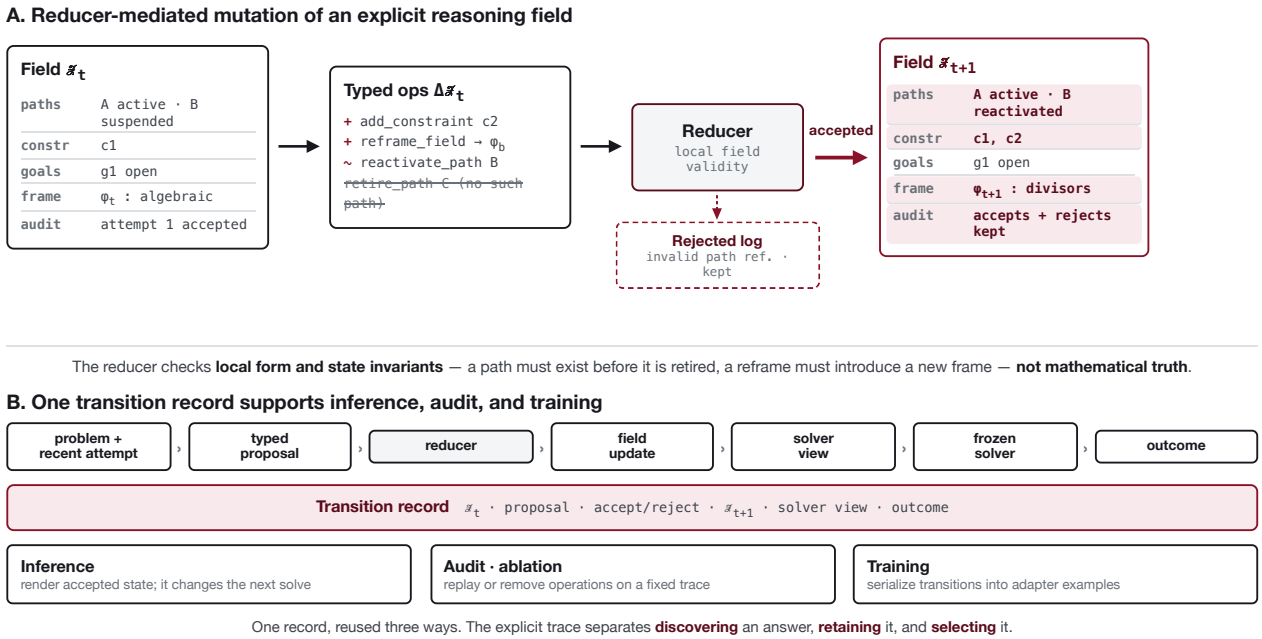

Figure 1. **Typed inference dynamics as persistent, reducer-mediated state mutation.** (A) Illustrative representation schematic: a typed field stores candidate paths, constraints, subgoals, frame, determined values, and provenance. Helper-proposed operations mutate this field through a deterministic reducer that preserves rejected proposals but validates local state motion rather than mathematical truth; the operation audit reports which operations the evaluated implementation actually realized. (B) Accepted state is used at inference time through solver-facing rendering, while the transition record itself is reused for audit, ablation, and adapter training. The explicit trace separates discovery, retention, and selection.

one auditable experimental object.

The evidence stack is correspondingly multi-layered. The original loop shows positive pass-by-budget coverage on full MATH-500 Level 5 (Hendrycks et al., 2021) for two frozen carriers at $n = 134$. On OlympiadBench (He et al., 2024), the stronger carrier transfers with a $+24.1$ pp T2 lift, while AIME (Mathematical Association of America, 2025) exposes a null endpoint. Fresh fixed-horizon controls separate repeated attempts, schema-lite state update, and typed reducer-mediated update. R1/R2/R3 controls test initial-field visibility, direct answer-string leakage, and cross-item structural scaffolding. A supervised adapter sweep shows that transition traces are learnable, not just inspectable. A supplementary proof corpus shows that the same operation alphabet can express proof-like motion across 200 locally checked trajectories.

Contributions. First, we define typed inference dynamics as a persistent reducer-mediated state-transition layer for mathematical reasoning. Second, we evaluate it as an inference-time intervention and audit object across MATH-500 Level 5, OlympiadBench, AIME, and fresh A/B/C*/D controls. Third, we show that the same transition record supports ablation and adapter training through a 5,000-example transition corpus and hard-harness sweep. Fourth, we connect the substrate to AI4Math proof and self-improvement themes while keeping proof validity, verifier- or evidence-calibrated authority, and RL-policy training outside the present claim.

2. Typed Inference Dynamics

The substrate separates three objects that are often conflated. The *field* is persistent state. The *operation trace* is a typed proposal for how the field should change. The *rendering* is the solver-facing observation derived from accepted state. The solver ultimately receives text, but the experimental record exists before that text: a typed transition trace over a persistent state, with rejected operations preserved rather than silently discarded.

State. A field stores beliefs, constraints, paths, frame, subgoals, and determined quantities. Paths carry status: active, suspended, soft-retired, hard-retired, merged, or reactivated. A status is not decorative metadata. It controls what later rendering can recommend, what obstruction is visible, and whether a previously abandoned route can return after a new constraint appears.

Actions and reducer. In the evaluated implementation, the helper receives the problem and a bounded recent solver attempt; it does not directly consume the full persistent field.

The policy is therefore attempt-conditioned typed patch generation, followed by deterministic reduction. The reducer checks local form and state invariants: a path must exist before it can be retired; a suspended route must name a reactivation condition; a reframe must introduce a new frame id; a subgoal closure must point to an existing subgoal; and malformed operations cannot mutate the field. This guarantee is intentionally local. The reducer can reject malformed state motion; it cannot certify that a mathematical predicate is true.

Rendering and credit. The accepted field is rendered as established facts, live or retired paths, recommended frame, open subgoals, and determined values. This choice makes the interface usable by a frozen solver, but it also creates a real confound: rendering can help by imposing structure, and unverified helper constraints can be promoted to authoritative language. We therefore measure this confound through controls instead of treating it as a nuisance. The credit signal in this paper is the downstream solved/unsolved outcome plus the operation/rejection log.

Concrete solver-facing view. One MATH-500 Level-5 rescue illustrates the interface. The helper proposed four operations: retire a path, add the constraint "sum of proper divisors of 220 is 284," reframe toward summing divisors of 284, and reactivate the path. The reducer accepted the constraint and reframe, but rejected the path operations because the referenced path did not exist. The frozen solver then saw this compact rendering:

```
Established (already derived, you may rely on
these):
- sum of proper divisors of 220 is 284
Recommended approach:  sum the proper divisors
of 284.
Using the above, solve the original problem.
```

This trace is useful precisely because it exposes both sides of the mechanism: accepted/rejected operations are auditable, while the current renderer validates operation form rather than mathematical truth. The word "Established" is therefore stronger than the reducer's guarantee; it is an empirical interface choice, not a proof certificate.

Not just structured prompting. A structured prompt has no necessary memory of which path was retired, why it was retired, which operation was rejected, or how a later solve should credit a path transition. Typed inference dynamics makes those events explicit. In Figure 1, the helper proposes both useful and malformed operations. A prose rationale would blur them into one fluent explanation. The reducer separates accepted constraints and frame transitions from invalid path references; the downstream solve then tests whether the accepted state helped.

Not process supervision. Process-reward models evaluate step quality (Lightman et al., 2023; Wang et al., 2024); here the question is what state change should survive. A scalar step score may influence search, but it usually does not preserve a reactivation condition, merge path fragments, or provide a replayable audit log of accepted and rejected state motion. This distinction is especially important for mathematical reasoning, where a failed path can still deposit useful information into the field.

Why path status matters. Mathematical work often advances through negative information. A route may fail because it assumes an impossible divisibility condition, but that failure can still reveal a conserved quantity or a useful substitution. A route may be suspended because a missing lemma is not yet available, then reactivated after a reframe makes that lemma natural. Linear rationales can describe these events after the fact, but they do not require the events to be typed, replayable, or available as a later training signal. In the runs reported here, however, the path-fate part of this ontology is only weakly exercised: the MATH setup initializes a single effective path, and the realized field changes are dominated by constraint accretion and reframing. The present setup therefore does not evaluate garbage collection or context growth under a large retired-path population. We treat full path-population management as future architecture rather than as an evaluated implementation claim.

From trace to training record. Each accepted/rejected transition can be serialized as a training object: pre-state, candidate path, operation sequence, reducer decision, rendered state, carrier, turn, and downstream outcome. This is why the adapter arm is scientifically connected to the inference-time arm. The adapter is not trained on arbitrary solution prose alone; it is trained on examples whose useful unit is a field transition. The present experiments do not prove that every learned behavior is internally represented as the same algebra, but they do test whether the transition format carries learnable signal.

3. Evidence Design

All headline solver evaluations use frozen Qwen2.5-family carriers (Qwen Team, 2025) served with vLLM (Kwon et al., 2023). C2 and C3 are carrier settings: C2 is a 7B solver and C3 is a 14B-AWQ solver. T0, T1, T2, and T3 are loop turns. The reported main loop endpoint is cumulative T2 coverage: an item counts if it is solved at T0 or any turn through T2. Paired item-level bootstrap intervals use seed 20260517 unless the source run is an aggregate adapter evaluation.

We organize the evidence as five layers. **Frozen-loop effects** measure downstream behavior under the original typed loop. **Dissociation** tests whether field rendering has directional-

ity rather than uniform benefit. **Controls** test initial-field visibility, direct answer-string leakage, repeated attempts, and structural scaffolding. **Trainability** tests whether many transition traces can be distilled into adapters. **Proof-like representation** tests whether the same operation alphabet can encode proof construction trajectories.

Transition-credit interface. The state is $\mathcal{F}_t$, a structured inference state over candidate paths, constraints, subgoals, frame, and determined values. The action is the helper's typed operation sequence. The transition is deterministic reduction to $\mathcal{F}_{t+1} = \mathcal{F}_t \oplus \Delta\mathcal{F}$, with rejected operations logged. The observation is the rendered accepted field given to the solver. The credit signal is the downstream outcome and the trace that produced it. This paper does not train an RL policy. It exposes a state-action-transition-credit interface that can support future credit assignment or reinforcement learning over mathematical reasoning trajectories.

Run provenance and denominators. Main loop and control run identifiers are recorded in the supplementary artifacts rather than printed as long code strings in the main text. OlympiadBench C3 began from 30 submitted items; one item had an execution/normalization exclusion, giving effective $n = 29$. Adapter results are aggregate hard-harness evaluations over 30 problems and three attempts per problem.

What counts as evidence. For loop runs, the statistical unit is the benchmark item. We do not count multiple operations on the same problem as independent evidence. T0 is the frozen solver's initial result; T2 cumulative coverage asks whether the item has been solved by any turn through the reported endpoint. For controls, the paired item-level comparison is the relevant object. For adapters, the evidence is pass count over 90 attempts and distinct problems solved at least once. This keeps the paper from inflating evidence through repeated turns or repeated samples.

Weak and negative regimes. The design intentionally keeps regimes that do not flatter the method. AIME C3 is null. The hard paired dissociation reverses sign. DPO and several plain LoRA variants are weak on the hard harness. R3 shows that structural scaffold alone moves the solver. These boundaries help distinguish typed inference dynamics from generic optimism about longer prompts.

4. Results

Full MATH-500 Level 5. On the full Level-5 subset ($n = 134$), cumulative T2 coverage improves from $35/134$ to $53/134$ at C2, a $+13.4\,\mathrm{pp}$ lift with 95% CI $[+7.5, +19.4]$. At C3 it improves from $36/134$ to $59/134$, a $+17.2\,\mathrm{pp}$ lift with CI $[+11.2, +23.9]$. These are the strongest MATH results because they avoid the original small-n concern and evaluate the same full item set across the carrier ladder.

The MATH result also clarifies scope. It is not a one-off trace demonstration, because the denominator is the full Level-5 subset. It is not a single-carrier accident, because both C2 and C3 move positively. It is not a claim that every added structure helps, because the same paper reports a hard-subset reversal and an AIME null. The supported statement is more precise: the original typed loop moves within-loop cumulative coverage in the full hard MATH setting, and that movement varies by carrier and problem regime.

Cross-benchmark transfer and boundary. On OlympiadBench, C3 improves from $4/29$ to $11/29$, a $+24.1\,\mathrm{pp}$ lift with CI $[+10.3, +41.4]$. C2 is weaker at $+6.7\,\mathrm{pp}$ with a lower bound at zero. On AIME 2024–25, C2 reaches $+10.0\,\mathrm{pp}$ with a CI touching zero, while C3 is exactly null at the reported T2 endpoint. The substrate transfers beyond MATH-500, but the evidence does not support a benchmark-general improvement claim.

Fixed-horizon comparison with a role-matched state updater. To answer the cost and multi-role concern directly, we ran a fresh fixed-horizon comparison on the same MATH-500 Level-5 item set ($n = 134$), with the same 14B-AWQ solver family and no gold-conditioned stopping. Table 3 separates a single solve (A), three independent attempts (B), a schema-lite semantic-state updater (C*), and full typed inference dynamics (D). C* stores three free-form state slots—invalid assumption, established/tentative fact, and next strategy—but does not use the TID operation algebra, deterministic reducer, or accept/reject log. We report the cumulative endpoint as *oracle discovery coverage*: it counts an item if any turn found the answer and therefore overstates what a deployed system can select without gold. Against repeated attempts, D improves oracle discovery by $+6.0\,\mathrm{pp}$ (95% CI $[+0.0, +11.9]$, McNemar $p = 0.096$; Holm-adjusted $p = 0.578$ across exploratory coverage tests). Against schema-lite state update, D differs by $+0.8\,\mathrm{pp}$ (CI $[-6.7, +7.5]$, $p = 1.000$), so no net coverage difference is detected at this sample size. On the realized terminal endpoint, D is $44/134$ versus C* at $36/134$ ($+6.0\,\mathrm{pp}$, CI $[-1.5, +13.4]$, $p = 0.169$), a higher point estimate but not a significant paired difference.

Discovery versus retention. The fresh comparison makes the strongest new diagnosis concrete without replacing the paper's identity. The evaluated TID implementation visits a correct answer on 61 items, but only 44 remain correct at the terminal turn and the frozen majority-with-terminal-tiebreak selector realizes 45. The D trajectory decomposes into 28 persistent-correct items, 16 rescues, 8 corruptions, and 9 T1-only discoveries that are lost before the endpoint.

Table 1. **Evidence stack used in the main claim.** Rows are source-traced. The table is designed to show scale, scope, and known boundaries at once.

Layer	Evidence	Main value	Boundary
Full MATH loop	MATH-500 Level 5, full $n = 134$, C2/C3	C2 $35 \to 53$; C3 $36 \to 59$	Same model family; T2 endpoint
Cross-benchmark	OlympiadBench and AIME 2024–25	Olympiad C3 $4/29 \to 11/29$	AIME C3 null $1/30 \to 1/30$
Mechanism controls	R1 F0-only; R2 gold guard; R3 cross-item shuffle	R2 $+23.3\,\mathrm{pp}$; R3 $+10.0\,\mathrm{pp}$	residual CI includes zero
Trainability	5,000-example transition corpus; 30 problems $\times$ 3 attempts	7B LoRA $6/90 \to 33/90$	aggregate adapter evidence; weights excluded from direct supplement
Proof-like motion	200 trajectories; 50 theorem statements; four paths each	200/200 local SymPy checks	not Lean/Coq/Isabelle validity

Table 2. **Main T2 typed-loop lifts.** Counts are cumulative correct at T0 and through T2.

Benchmark	Carrier	n	Count change	Lift
MATH-500 L5	C2	134	$35 \to 53$	$+13.4\,\mathrm{pp}$
MATH-500 L5	C3	134	$36 \to 59$	$+17.2\,\mathrm{pp}$
OlympiadBench	C3	29	$4 \to 11$	$+24.1\,\mathrm{pp}$
AIME 2024–25	C3	30	$1 \to 1$	$+0.0\,\mathrm{pp}$

Thus the typed transition object is valuable not only as an intervention, but as an audit record showing where discovery is not yet reliably retained.

Measured cost profile. Table 4 reports measured calls, input tokens, generated tokens, total tokens, and serial wall time. Token counts are the more portable accounting; wall time depends on the vLLM/RTX PRO 6000 Blackwell serving stack and is reported as an implementation-dependent GPU-time proxy. The solver call budget and item set are the same for B/C*/D; helper calls and helper prompt sizes are measured consequences of each arm's design. D uses approximately $1.98\times$ the total tokens of repeated sampling (B) and $1.05\times$ those of C*, while wall-time differences are not interpreted as efficiency because serving throughput is implementation-dependent.

Controls bound simple explanations. R1 renders the initial field only and gives a small, non-significant $+3.3\,\mathrm{pp}$ lift (CI $[-16.7, +23.3]$). R2 adds a direct gold-substring guard and leaves the hard-subset lift at $+23.3\,\mathrm{pp}$ (CI $[+10.0, +40.0]$), so direct verbatim answer leakage is not driving that guarded run. R3 replays cross-item operation sequences and gives $+10.0\,\mathrm{pp}$ with a lower confidence bound at zero, suggesting a structural-scaffold contribution without isolating it decisively. The residual content-specific component is estimated at $+13.3\,\mathrm{pp}$, but its CI $[-3.3, +30.0]$ includes zero. The controlled conclusion is therefore stronger than a prompt/no-prompt story but narrower than a fully isolated content-specific mechanism.

Directionality. On the original easy paired set, field rendering gives a 7B solver a positive $+7.06\,\mathrm{pp}$ Δ-of-Δ advantage (CI $[+2.75, +12.16]$). On a hard MATH-500 Level-5 paired subset, the sign reverses at C3: $-32\,\mathrm{pp}$ (CI $[-52, -11]$). This reversal is not the headline improvement, but it is scientifically important. If the method were only extra text, more effort, or answer leakage, monotone benefit would be the natural expectation. Instead, the intervention changes the solver's search region in a way that can help or hurt across carrier and item strata.

Transition traces are trainable. The adapter sweep tests whether the same transition object used for inference and audit can be distilled, not whether LoRA (Hu et al., 2021) itself is novel. The hard harness contains 30 problems and three attempts per problem. Untrained baselines are weak: 1.5B obtains $0/90$, 3B obtains $2/90$, and 7B obtains $6/90$. Plain or weak variants remain low. Enriched r64 adapters are different: 3B reaches $29/90$ and 7B reaches $33/90$. The transition corpus is learnable and parameter-efficient under this harness, but this does not show that the adapter has internalized the full field algebra.

The weak variants matter. They reduce the chance that the training result is merely "any fine-tuning helps." The DPO-3B run reaches only $8/90$; 3B credit weighting reaches $5/90$; plain 7B field v1 reaches $8/90$. The enriched corpus and higher-rank adapter are the settings where the lift becomes large. The paper therefore treats trainability as evidence for the transition representation, not as a claim that parameter-efficient tuning alone is sufficient.

5. Positioning and AI4Math Relevance

Typed inference dynamics is closest to methods that make reasoning more structured, but its claim is narrower than "first mutable reasoning state." Blackboard systems already treated problem solving as shared structured state (Hayes-Roth, 1985; Nii, 1986). Chain-of-thought and rationale supervision expose text; process reward models score local

Table 3. **Fresh fixed-horizon comparator.** Any-turn coverage is an oracle discovery endpoint; terminal and majority+terminal are realized endpoints. Counts are over the same 134 items.

Arm	Mechanism	Any-turn	Terminal	Selector	Main reading
A	Single greedy solve (D turn-0 reference)	36	36	36	Initial frozen-solver level
B	Three independent solver attempts	53	35	40	Repeated-attempt reference
C*	Schema-lite semantic-state update	60	36	38	No detected net coverage difference vs D at $n = 134$
D	Evaluated TID implementation	61	44	45	Higher terminal point estimate; not significant vs C*

Table 4. **Measured cost profile for the main fresh arms.** Tokens are summed from raw call logs; wall is serial summed call time. A is the shared D turn-0 reference and should not be added to D.

Arm	Solver calls	Helper calls	Total calls	Input tok.	Gen. tok.	Total tok.	Wall (s)
A	134	0	134	18,858	63,542	82,400	4,240.5
B	402	0	402	56,574	190,376	246,950	16,072.6
C*	402	268	670	256,442	207,558	464,000	14,306.5
D	402	268	670	270,898	217,614	488,512	14,556.2

Table 5. **Control interpretation.** The controls are used to bound explanations, not to prove a single pure mechanism.

Control	Result	Interpretation
R1 F0-only	$+3.3$ pp; CI crosses 0	Initial visibility not enough
R2 gold guard	$+23.3$ pp; CI excludes 0	Direct answer-string leakage not driving guarded lift
R3 shuffle	$+10.0$ pp; lower bound 0	Suggests scaffold contribution; interval reaches zero
Residual	$+13.3$ pp; CI crosses 0	Content component unresolved at $n = 30$

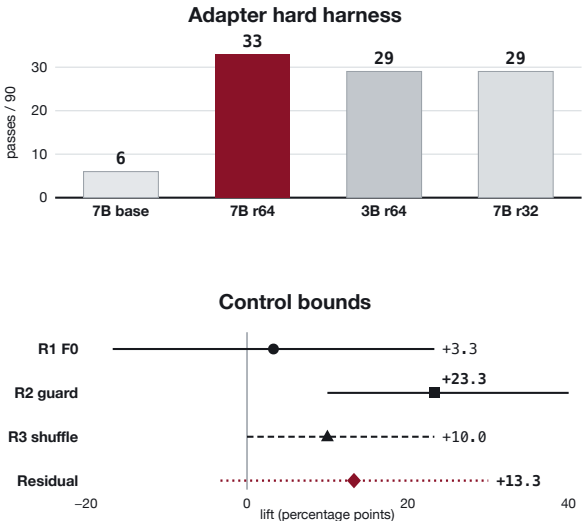

Figure 2. **Trainability and controls.** Top: enriched transition data lifts the 7B hard-harness pass count from $6/90$ to $33/90$, while a 3B r64 adapter reaches $29/90$. Bottom: R1/R2/R3 controls bound visibility, direct leakage, and structural scaffold; intervals are 95% paired bootstrap confidence intervals.

steps; Tree-of-Thoughts and GFlowNet-style methods expand or sample frontiers (Yao et al., 2023a; Bengio et al., 2021); STaR-style methods generate new training traces (Zelikman et al., 2022); and ReAct, Reflexion, PAL, StateFlow, and Kwai-STaR connect reasoning to actions, memory, state machines, or self-improvement (Yao et al., 2023b; Shinn et al., 2023; Gao et al., 2023; Wu et al., 2024; Lu et al., 2024). Recent systems make the overlap even clearer: Canvas-of-Thought uses a mutable structured substrate for multimodal reasoning (Sun et al., 2026), STAR-PólyaMath uses an orchestrated state machine with persistent meta-strategic supervision for competition mathematics (Wu et al., 2026), and *Correct Is Not Enough*, through TraceLift, trains intermediate reasoning as an executor-consumable artifact (Han et al., 2026). TID's narrower contribution is the mathematical state mutation itself: typed operations over paths/constraints/frames, deterministic accept/reject reduction, a replayable operation log, solver-facing rendering, and a shared transition record that can support inference, audit, ablation, and adapter training.

AI4Math fit. The representation gives a modest but useful bridge between informal rationales and proof states. Formal provers expose kernel-checked tactics after a problem has been translated into a proof language (Wu et al., 2022; Jiang et al., 2023; Google DeepMind, 2024); many informal mathematical solutions operate before such states exist. TID offers a pre-formal layer in which a failed route can deposit a constraint, a frame can change, a subgoal can be opened, and malformed state motion can be rejected. The supplementary proof corpus contains 200 typed proof-construction trajectories over 50 theorem statements with local SymPy

Table 6. **Adapter suite on the hard 30-problem harness.** Weak variants are retained to show that the enriched transition corpus and rank setting matter; the claim is not that smaller models generally outperform larger ones.

Run	Base	Rank/style	Passes	Rate	Solved	Role
untrained-1.5B	1.5B	none	$0/90$	0.0%	$0/30$	baseline
untrained-3B	3B	none	$2/90$	2.22%	$2/30$	baseline
untrained-7B	7B	none	$6/90$	6.67%	$4/30$	baseline
LoRA-1.5B-v1	1.5B	field v1	$6/90$	6.67%	$6/30$	weak retained
LoRA-1.5B-v2	1.5B	field v2	$3/90$	3.33%	$2/30$	weak retained
LoRA-3B-credit	3B	credit	$5/90$	5.56%	$4/30$	weak retained
LoRA-3B-v2	3B	field v2	$6/90$	6.67%	$6/30$	weak retained
DPO-3B	3B	preference	$8/90$	8.89%	$6/30$	weak retained
LoRA-7B-v1	7B	field v1	$8/90$	8.89%	$7/30$	weak retained
LoRA-3B-r64-enriched	3B	enriched r64	$29/90$	32.22%	$10/30$	strong
LoRA-7B-r32-enriched	7B	enriched r32	$29/90$	32.22%	$11/30$	strong
LoRA-7B-r64-enriched	7B	enriched r64	$33/90$	36.67%	$12/30$	best

checks, but this is representational evidence, not Lean, Coq, or Isabelle validation.

Credit-facing reading. A flat solution transcript gives sparse credit to a long sequence of tokens. A typed transition record creates natural credit anchors: accepted operation, rejected operation, changed field, rendered observation, downstream answer. This paper only uses supervised adapter training and empirical T0/T2 credit, but the interface is deliberately compatible with later reinforcement learning, counterfactual replay, or process-level reward modeling over mathematical state changes. Those later systems are outside the present claim; the contribution here is to make their target object explicit.

Carrier and stratum effects. Carrier/problem-stratum behavior is not incidental bookkeeping. It is evidence that the typed state is an intervention on a coupled system: the field, the renderer, the solver, and the problem class. Full MATH-500 Level 5 and OlympiadBench show positive regimes; AIME C3 and the hard directionality reversal show that the same intervention can be neutral or harmful. That pattern is scientifically more informative than a single averaged lift. It says that typed transitions are not magic hints; they are structured perturbations whose value varies with whether the current solver is in a region where reframing a constraint or preserving a subgoal changes the next computation in the right direction.

6. Boundaries

Reducer scope. The reducer checks form and local invariants, not mathematical truth. A syntactically valid constraint can still be false, and active constraints are rendered to the solver with stronger authority than the reducer has earned. In the evaluated implementation audit, accepted operations are dominated by constraint addition and reframing; advertised path-fate operations rarely survive reduction, and `advance_path` is accepted in 0/225 proposals. This boundary keeps the substrate in the right place: after free-form rationale and before proof-kernel validation. The paper's contribution is a more inspectable informal transition layer, not a replacement for theorem provers.

Leakage and scaffold. The channel is not formally leak-proof. R2 shows that direct gold-substring leakage is not driving the guarded run, but that is weaker than a theorem about all possible leakage. R3 suggests a structural-scaffold contribution, but its interval reaches zero and the content-specific residual remains unresolved at $n = 30$. The honest reading is that typed rendering can move the solver by organizing state, and may move it further through item-specific content; the present controls bound that decomposition but do not close it.

Benchmark and carrier scope. Empirical scope is bounded but diagnostic: AIME C3 is null at T2; the smallest carrier shows irregular threshold behavior in some runs; the hard-subset directionality reversal shows that typed transitions can move a solver away from a productive path; and all headline carriers are from the same model family. The fresh fixed-horizon comparator directly tests repeated attempts, role structure, and cost, but its D-vs-baseline accuracy differences are not significant after multiplicity control. These boundaries qualify the effect while strengthening the central substrate claim: the transition object is observable enough to reveal where discovery, retention, and selection diverge.

Training evidence scope. Adapter results show that transition traces can be distilled under a controlled hard harness. They do not show that the model has internalized the full field algebra, and they do not compare against every possible untyped or same-token-budget training baseline. The weak variants in Table 6 are still useful: they show that not every training choice produces the headline lift, and they make the enriched transition corpus a real object of study.

Artifact scope. The supplementary artifact package is a reproduction package, not a full internal archive. It includes sanitized row-level outcomes where available, aggregate adapter summaries, reducer code, figures, and scripts. It excludes adapter weights, raw benchmark text, raw gold answers, private logs, hidden files, local absolute paths, and raw source archives. Large-model reruns require external compute and model access.

From discovery to retained state. The discovery-retention gap points to a concrete next system, not an evaluated implementation claim. The next system should condition updates on the persistent field, preserve candidate lineages instead of overwriting earlier endpoints, and calibrate the authority of retained claims against verifier or external evidence. Candidate archives, field-conditioned updating, and verifier- or evidence-calibrated authority are therefore motivated by the present audit, but they are not part of the reported evidence.

Why this still matters. The substrate would be less credible if every condition were positive. The nulls and reversals make the intervention falsifiable. R2 narrows the direct-leakage explanation, R3 suggests but does not isolate scaffold effects, and the residual content-specific component remains unresolved. The adapter sweep shows that transition traces can influence trained behavior. The proof corpus shows that the operation alphabet is not limited to answer extraction. Together, these facts support a disciplined claim: typed inference dynamics is a useful experimental layer for mathematical reasoning, with known limits and clear next tests.

7. Conclusion

Typed inference dynamics turns partial mathematical reasoning into a state-transition object: paths can be retired or reactivated, constraints deposited, frames changed, malformed operations rejected, and downstream credit attached to a trace. Across full MATH-500 Level 5, OlympiadBench, controls, adapter distillation, and proof-like trajectory evidence, the substrate is no longer just a prompt style or a fine-tuning recipe. It is an auditable and trainable layer for studying how mathematical reasoning state changes over time.

The important result is not that every typed transition helps, but that such transitions create a measurable and falsifiable object. The evaluated loop raises within-budget coverage in some regimes, is null or harmful in others, and still leaves an auditable and trainable transition record. The next tests are now concrete: larger controlled mechanism runs, non-Qwen carriers, untyped training baselines, proof-assistant-connected trajectories, and stronger credit assignment over the state-action-transition-observation interface.

Typed inference dynamics therefore makes intermediate mathematical reasoning state not only visible, but experimentally intervenable.

References

Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. *Advances in Neural Information Processing Systems*, 2021.

Luyu Gao, Aman Madaan, Shuyan Zhou, Uri Alon, Pengfei Liu, Yiming Yang, Jamie Callan, and Graham Neubig. PAL: Program-aided language models. In *International Conference on Machine Learning*, 2023.

Google DeepMind. AI achieves silver-medal standard solving international mathematical olympiad problems. Blog post, 2024.

Tianyang Han, Hengyu Shi, Junjie Hu, Xu Yang, Zhiling Wang, and Junhao Su. Correct is not enough: Training reasoning planners with executor-grounded rewards. *arXiv preprint arXiv:2605.03862*, 2026.

Barbara Hayes-Roth. A blackboard architecture for control. In *Artificial Intelligence: An International Perspective*. Elsevier, 1985.

Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Leng Thai, Junhao Shen, Jinyi Hu, Xu Han, Yujie Huang, Yuxiang Zhang, Jie Liu, Lei Qi, Zhiyuan Liu, and Maosong Sun. Olympiadbench: A challenging benchmark for promoting AGI with olympiad-level bilingual multimodal scientific problems. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. *arXiv preprint arXiv:2103.03874*, 2021.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

Albert Q. Jiang, Sean Welleck, Jin Peng Zhou, Wenda Li, Jiacheng Liu, Mateja Jamnik, Timothée Lacroix, Yuhuai Wu, and Guillaume Lample. Draft, sketch, and prove: Guiding formal theorem provers with informal proofs. In *International Conference on Learning Representations*, 2023.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with PagedAttention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*, 2023.

Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. *arXiv preprint arXiv:2305.20050*, 2023.

Xingyu Lu, Yuhang Hu, Changyi Liu, Tianke Zhang, Zhenyu Yang, Zhixiang Ding, Shengsheng Qian, Meng Du, Ruiwen Kang, Kaiyu Tang, Fan Yang, Tingting Gao, Di Zhang, Hai-Tao Zheng, and Bin Wen. Kwai-STaR: Transform LLMs into state-transition reasoners. *arXiv preprint arXiv:2411.04799*, 2024.

Mathematical Association of America. American Invitational Mathematics Examination problems. Problem archive, 2025.

H. Penny Nii. Blackboard systems. *AI Magazine*, 7(2): 38–53, 1986.

Qwen Team. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2025.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems*, 2023.

Lingzhuang Sun, Yuxia Zhu, Ruitong Liu, Hao Liang, Zheng Sun, Caijun Jia, Honghao He, Yuchen Wu, Siyuan Li, Jingxuan Wei, Xiangxiang Zhang, Bihui Yu, and Wentao Zhang. Canvas-of-thought: Grounding reasoning via mutable structured states. *arXiv preprint arXiv:2602.10494*, 2026.

Peiyi Wang, Lei Li, Zhihong Shao, R. X. Xu, Damai Dai, Yifei Li, Deli Chen, Yu Wu, and Zhifang Sui. Math-shepherd: Verify and reinforce LLMs step-by-step without human annotations. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024.

Jiaao Wu, Xian Zhang, Hanzhang Liu, Sophia Zhang, Fan Yang, and Yinpeng Dong. STAR-PólyaMath: Multi-agent reasoning under persistent meta-strategic supervision. *arXiv preprint arXiv:2605.19338*, 2026.

Yiran Wu, Tianwei Yue, Shaokun Zhang, Chi Wang, and Qingyun Wu. StateFlow: Enhancing LLM task-solving through state-driven workflows. *arXiv preprint arXiv:2403.11322*, 2024.

Yuhuai Wu, Albert Q. Jiang, Wenda Li, Markus N. Rabe, Charles Staats, Mateja Jamnik, and Christian Szegedy. Autoformalization with large language models. *Advances in Neural Information Processing Systems*, 2022.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan.

Tree of thoughts: Deliberate problem solving with large language models. In *Advances in Neural Information Processing Systems*, 2023a.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023b.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah D. Goodman. STaR: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems*, 2022.

A. Reducer Schema and Operation Alphabet

The reducer state is a tuple of beliefs, constraints, paths, frame, subgoals, and determined quantities. The operation alphabet used in the cited runs includes `add_constraint`, `retire_constraint`, `weaken`, `strengthen`, `add_path`, `kill_path`, `suspend_path`, `resurrect_path`, `merge_paths`, `transform_construct`, `reframe_field`, `open_subgoal`, `close_subgoal`, and `record_warning`. Legacy code names map to the main-paper prose terms: `kill_path` corresponds to retiring a path, and `resurrect_path` corresponds to reactivating a path. Shape checks include missing-argument checks, source checks for constraints, duplicate-path checks, path-status invariants, non-empty transformed constraints, and frame-id change checks.

B. Upload Artifact Levels

The paper separates what readers can recompute directly from what requires upstream infrastructure. The supplementary artifact package includes sanitized T2-lift rows, figure-generation scripts, reducer code, adapter-count summaries, proof-footprint summaries, run identifiers, configuration metadata, and provenance notes. It does not include GPU rerun infrastructure, raw benchmark problems, raw gold answers, adapter weights, checkpoints, private logs, local paths, or raw source archives.

C. Statistical Procedure

Headline T0-to-T2 intervals are paired item-level bootstrap intervals with $B = 10000$ and seed 20260517 in the source artifacts. The supplementary artifacts include sanitized per-item rows sufficient to recompute point lifts and bootstrap intervals for the headline typed-loop runs. Fresh fixed-horizon comparisons use paired item-level confidence intervals and exact two-sided McNemar tests; displayed exploratory coverage comparisons use Holm adjustment over the stated exploratory coverage tests.

D. Fresh Comparator Contract and Endpoint Definitions

Table 7. **Fresh comparator contract.** This table fixes the contract used for Table 3.

Item	Locked definition
Models and revisions	Solver and helper roles use `Qwen/Qwen2.5-14B-Instruct-AWQ`; vLLM logs record `revision=None` and `tokenizer_revision=None`. Run identity is locked by raw-row hashes, run hashes, and prompt signatures rather than a pinned external commit.
Horizon and stop rule	Same 134-item MATH-500 Level-5 subset; fixed horizon over turns 0, 1, and 2; all arms complete the horizon; no gold-conditioned stopping.
Decoding and seeds	Solver cap is 512 tokens; helper cap is 768 tokens; vLLM engine seed is 0; B uses three independent repeated solver attempts with temperature 0.7 for resampled turns; fresh bootstrap intervals use seed 20260517.
Arm inputs and persistence	A is the D turn-0 single solve. B persists no state across attempts. C^* receives the problem and recent attempt, then persists schema-lite slots for invalid assumption, established/tentative fact, and next strategy. D receives the problem and recent attempt, then persists reducer-accepted typed field updates and rejected-operation logs.
Policy status	A, B, C^*, and D are on-policy for their own mechanisms. Additional diagnostics C, E, F, and G0 are not main-table arms: C is a free-form persistent buffer with truncation/answer-anchoring confounds, E is schema/status-erased replay and off-policy after the first correction, F is an on-policy compiled projection, and G0 is a one-step authority-wording diagnostic.
Endpoints	Any-turn is oracle discovery coverage: an item counts if any turn is correct. Terminal is turn-2 correctness. Selector applies frozen answer normalization, takes majority over normalized turn answers, and uses the terminal answer as the tie-break; unresolved ties therefore resolve to the terminal normalized answer.
Statistics	Fresh intervals are paired item-level percentile bootstrap intervals with $B = 10000$ and seed 20260517; tests are exact two-sided McNemar tests. Holm adjustment is over seven exploratory coverage comparisons: C-vs-B, D-vs-B, D-vs-C, D-vs-C^*, C^*-vs-B, E-vs-B, and D-vs-E.

The main text reports A/B/C^*/D because these are the cleanest fixed-horizon arms for the comparison concern: single solve, repeated attempts, role-matched schema-lite updating, and typed reducer-mediated updating. C/E/F/G0 are retained as diagnostics for mechanism interpretation, not as replacement headline results.

E. Operation and Authority Audit

The v4 operation audit records 1,119 proposed operations, 515 reducer-accepted operations, and 249 state-changing transitions. Accepted operations are dominated by `add_constraint` and `reframe_field`; `advance_path` is proposed 225 times and accepted zero times because it is not executable by the reducer used in the run. A separate authority audit finds that helper-asserted constraints are rendered under "Established" language without truth checking. This supports the method-boundary claim in the main text: the evaluated system is a typed reducer-mediated state-control interface, not a truth-verified or transactional controller.

F. Controls

R1 tests whether the initial field rendering alone explains the lift. R2 applies a direct gold-substring guard to operation strings. R3 shuffles operation sequences across items to estimate the structural-scaffold component. The residual between the guarded baseline and cross-item shuffle is reported as unresolved because its confidence interval includes zero at $n = 30$.

G. Training Details

The enriched corpus contains 5,000 typed-field examples, including generated field trajectories and successful typed-loop examples. Adapter evaluations use a 30-problem hard harness with three attempts per problem. The supplement reports pass counts and problem-solved counts without raw problem statements, gold answers, or model weights.

H. Proof-Corpus Summary

The supplementary proof corpus contains 200 typed-op proof-construction trajectories over 50 theorem statements with four paths per theorem. The source summary reports domains: number theory 52, combinatorics 44, algebra 36, inequalities 24, geometry 20, calculus 12, and analysis 12. Difficulty labels are intro 19, easy 68, intermediate 97, and hard 16. Each trajectory carries a `sympy_verification` record with executed, result, gold, and match fields; the source build raises on any missing or rejected operation trace. The symbolic checks audit local algebraic/arithmetic obligations and trajectory consistency. They are not proof-assistant validation.

I. Related PRM Diagnostic

A separate Phase-A typed-field diagnostic tested whether scalar process scores capture the same field-level structure as typed transitions. The task classified 300 locally verified typed-field traces against 298 operation-shuffled negatives from 107 unique problems with problem-disjoint five-fold cross-validation. Aggregate Qwen2.5-Math-PRM-7B features achieved mean AUC 0.518 (sd 0.014); 12 structural field features achieved mean AUC 0.945 (sd 0.025), with mean gap $+0.427$ and every fold favoring structural features. We include this only as a non-headline diagnostic for the PRM boundary: it is a separate corpus and classifier task, not a final-answer benchmark, not a proof-validity result, and not part of the MATH/Olympiad/AIME evidence stack.

J. Reproducibility and Artifact Limits

The supplementary artifact package contains sanitized rows, aggregate metrics, reducer code, figures, and reproduction scripts. It excludes raw source archives, private logs, local paths, adapter weights, raw benchmark text, and raw gold answers. Large-model inference and adapter reruns require external model access and compute, so the artifact package separates row-level reproduction from infrastructure-dependent reruns.