BISE: BRAIN-INSPIRED MEMORY CONSOLIDATION FOR SELF-EVOLVING AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Accumulating usable knowledge through autonomous exploration of an unfamiliar environment is a core capability of self-evolving agents. The challenge is to preserve the context that makes experience reusable while keeping emerging rules open to correction. We present BISE, a knowledge-accumulating agent inspired by the complementary-learning-systems view of biological memory. BISE separates ordered accounts of individual attempts from a compact, source-linked summary, allowing knowledge to be revised without erasing the experience behind it. On six Jericho games with GLM-5.3, BISE achieves a mean normalized score of 55.7%, 22.1 points above its matched ReAct control and 18.0 points above reproduced EvoTest. Gains over ReAct hold on two further backbones. In ScienceWorld, the 211-pair panel reaches a mean cumulative best score of 95.5 across up to five attempts. Ablations favor ordered episode context, while trajectory audits trace revised advice to subsequent actions. Together, these findings support retaining the context behind reusable procedures while keeping the knowledge drawn from them open to correction.

1 INTRODUCTION

Repeated exploration requires turning past attempts into guidance for the next. A successful action may depend on earlier preparation, while an initially plausible explanation may later prove wrong. LLM agents reuse experience through reflections, retrieval, and revised configurations (Shinn et al., 2023; He et al., 2026). Which structure should memory preserve to make experience reusable? The answer must account for both action dependencies and fallible interpretations.

Memory research offers a way to organize these records. Hippocampal activity can reactivate patterns from experience during sleep and rest; ordered place-cell sequences have been observed in both forward and reverse directions (Wilson & McNaughton, 1994; Lee & Wilson, 2002; Diba & Buzsáki, 2007). Indexing theory proposes that the hippocampus provides an index to distributed cortical activity associated with an episode (Teyler & Rudy, 2007). Multiple-trace and transformation accounts distinguish retained episodic detail from more general knowledge that develops across experiences (Nadel & Moscovitch, 1997; Tse et al., 2007; Moscovitch et al., 2016). Reward context can influence which experiences are retained (Gruber et al., 2016). We use these ideas as functional analogies. An episode-indexed gist retains context from one attempt, while a schema page summarizes information across attempts with links to the source records.

Existing LLM agents retain and use experience in several ways. *Prompt-augmentation* methods add reflections, retrieved passages, or running summaries to the context (Shinn et al., 2023; Lewis et al., 2020; Packer et al., 2023). *Model-training* methods update the backbone on self-generated traces, at the cost of additional training. *Knowledge-base* methods store skill libraries, rule manuals, curated memories, or revised agent configurations (Wang et al., 2024; Chen et al., 2024; Zhao et al., 2024; He et al., 2026; Dai et al., 2026; Suzgun et al., 2026; Wei et al., 2025; Xu et al., 2026). These approaches differ in their *memory representation*. In the implementations compared here, single-step triples and extracted facts do not preserve an entire route as one ordered unit. We test whether retaining that structure supports reuse.

We propose BISE around two complementary requirements: preserve the structure needed to repeat a procedure, and allow its interpretation to change without erasing its history (Figure 1). An *episode-indexed gist* retains a route, its conditions, and its outcome as one retrievable account. A separate

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

schema page consolidates recurring procedures and hazards, with links back to the attempts that support them. The compared EvoTest and RecMem implementations instead use single-step records or extracted facts (He et al., 2026; Dai et al., 2026). Our design makes the sequence surrounding an action available when the actor decides whether to reuse it, including preparation that earned no immediate reward.

We evaluate performance and the contribution of each memory layer. Matched comparisons on three backbones assess the complete memory procedure; ablations test alternatives to ordered gists. The results favor retaining episode context, while the gist-only arm performs close to the full system. The schema page's demonstrated role is therefore to organize revisable knowledge with inspectable sources. Trajectory audits trace how a rejected explanation and zero-scoring preparation change later advice.

We make three contributions.

(1) An ordered unit for experience reuse. We introduce episode-indexed gists that retain action sequences and their conditions together, allowing a later attempt to retrieve a procedure with the context needed to apply it.

(2) Revision without loss of provenance. We separate an append-only gist archive from a source-linked schema page, so revising a rule preserves access to the earlier accounts behind it.

(3) Evidence for the value and limits of structured memory. BISE exceeds matched ReAct and reproduced EvoTest by 22.1 and 18.0 points on GLM-5.3. Further backbones, ScienceWorld, ablations, and trajectory audits establish the scope of the gains and distinguish evidence for ordered gists from the page's role in organizing knowledge.

2 Related Work

Episodic memory, replay, and schema consolidation. Accounts of human memory distinguish working memory, episodic encoding, and integration across experiences (Squire, 2004; Baddeley, 2012; Kumaran et al., 2016). Event segmentation describes how continuous experience is divided into bounded events (Zacks et al., 2007); in our system, each completed attempt defines a record boundary. Studies of hippocampal replay document ordered reactivation of experienced routes (Wilson & McNaughton, 1994; Lee & Wilson, 2002), while indexing and multiple-trace accounts emphasize the retention of episode-specific context (Nadel & Moscovitch, 1997; Teyler & Rudy, 2007). These findings motivate ordered, episode-indexed entries. Complementary learning systems and schema research motivate a separate summary that integrates recurring information (McClelland et al., 1995; Tse et al., 2007; Winocur & Moscovitch, 2011). These functional analogies guide the separation of episode records from cross-attempt summaries. Cognitive architectures for language agents adopt related functional divisions (Sumers et al., 2024), and HippoRAG draws on hippocampal indexing for retrieval (Gutiérrez et al., 2024).

Long-term memory for LLM agents. Agents retain interaction history through retrieval, summaries, and reflections (Lewis et al., 2020; Packer et al., 2023; Shinn et al., 2023); memory banks and agentic memory systems organize accumulated experience for long-horizon use (Park et al., 2023; Zhong et al., 2024; Xu et al., 2025; Chhikara et al., 2025). The closest precedents distill cross-attempt experience into reusable text: RecMem consolidates when related interactions recur (Dai et al., 2026); AutoManual builds and revises rules before assembling a manual (Chen et al., 2024); ExpeL extracts insights from past trajectories (Zhao et al., 2024). Our focus is the retrievable unit: an ordered account of one attempt alongside a revisable summary with source links. The archive preserves episode context when the summary changes. We evaluate RecMem under our Jericho protocol.

Test-time adaptation. Memory supports test-time adaptation in different ways (Suzgun et al., 2026; Wei et al., 2025; Xu et al., 2026). MemGen uses reinforcement learning to train a trigger and weaver that generate latent memory during reasoning (Zhang et al., 2026). JitRL retrieves trajectories to estimate action advantages and modifies output logits without gradient updates (Li et al., 2026). In BISE, retrieved text informs the actor, and adaptation updates stored accounts and summaries. EvoTest revises prompts, memory, hyperparameters, and tools between episodes (He et al., 2026); BISE keeps model parameters and the update procedure fixed, adapting only memory content.

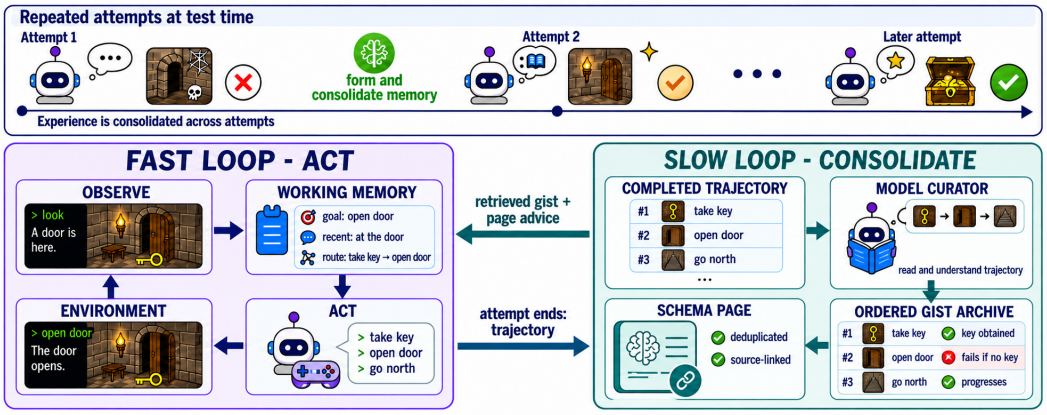

Figure 1: BISE connects within-attempt action to across-attempt consolidation. The curator reads a completed trajectory directly, appends an ordered gist, and updates a source-linked schema page. Retrieved gist and page advice inform the next attempt.

3 BISE

Problem setting. An agent interacts with a partially observed text environment for at most N attempts. Attempt e produces trajectory τ_e and ends when the task terminates or the interaction budget is exhausted. The environment then resets while external memory persists. The memory procedure selects information from each attempt, consolidates it across attempts, and retrieves it to inform later decisions. Algorithm 1 gives the complete cycle.

3.1 Working context and trajectory recording

Within an attempt, the actor's working context serves as an operational analogue of working memory. It holds the task, observations, recent actions, and retrieved entries during an interleaved reasoning-and-action cycle (Yao et al., 2023). Each environment reset starts a fresh working context. The log records location labels, scores, and commands directly. From the completed trajectory, the curator infers ordering requirements, applicability conditions, and causal explanations. Recording the complete attempt allows the curator to connect an early preparation step to a later outcome, even when the preparation itself earns no reward.

3.2 Episode-indexed gists

The gist is designed to preserve a procedure at the scale at which it can be reused. At each attempt boundary, the LLM curator reads the complete trajectory and appends one *episode-indexed gist*, retaining the ordered route, outcome, and relevant conditions. Reading the full trace exposes preparation and intervening failures that a filter for scoring events could omit. The original trajectory remains available for checking inferred explanations; the archive preserves earlier gists and any addenda as experience accumulates.

A gist also retains unsuccessful parts of a route. An attempt that ends in failure may reveal a useful procedure or expose a condition under which an action is unsafe. Preserving that context allows later attempts to reuse partial progress and consult the associated failure record.

Representation properties. We examine three properties of the gist in repeated tasks:

- *Event order.* The gist records actions in execution order, allowing the actor to retrieve and reuse a complete route.
- *Prerequisites and inferred causes.* The curator records inferred prerequisites and explanations alongside scoring actions, including relevant inventory and location conditions. These contextual details may be lost when a trajectory is reduced to isolated triples (§4.3).

- *Failure records.* The curator is instructed to record each death with its command, room, and inferred cause so that the actor can retrieve the hazard on a later attempt.

This representation balances abstraction with episode context: the actor receives a condensed account, but can still associate a conclusion with the sequence that produced it (Brainerd & Reyna, 2002; Teyler & Rudy, 2007). Section 4.3 tests this choice against verbatim logs, isolated triples, and gists with their order removed.

Algorithm 1 One BISE attempt and its boundary memory update.

Require: attempt e, archive E_{e-1}, page K_{e-1}, step budget T, word target B
1: initialize working context $W_{e,0} \leftarrow \varnothing$ and trajectory $\tau_e \leftarrow \varnothing$
2: **for** $t = 1, \ldots, T$ **do**
3: set $W_{e,t} \leftarrow W_{e,t-1}$; append observation and any requested retrieval from $K_{e-1} \cup E_{e-1}$
4: pass $W_{e,t}$ to the LLM actor, which returns a proposed action $a_{e,t}$
5: execute $a_{e,t}$; append the transition to $W_{e,t}$ and τ_e; break if terminal
6: distill the trajectory into one episode-indexed gist: $g_e = C_{\text{gist}}(\tau_e)$
7: append it to the archive: $E_e = E_{e-1} \cup \{g_e\}$
8: rewrite schema page with word target B: $K_e = C_{\text{sem}}(K_{e-1}, E_e; B)$
9: **return** updated (E_e, K_e)

3.3 Auditable distillation: the schema page

The schema page organizes recurring information from the gists into a summary that the actor can retrieve and a reader can inspect. At each attempt boundary, the curator updates one page of scoring procedures, hazards, and open questions, targeting roughly 800 words per game. The curator is instructed to link page entries to supporting attempts.

Provenance and verification. When a later attempt reproduces a recorded outcome, the curator can add a link to it. These links let a reader locate the cited observation and compare it with the current claim. The page can be revised at the next boundary while the append-only gist archive preserves earlier records. Section 4.4 follows a page revision from its source trajectory to later actions.

The separation addresses a tension in accumulating knowledge. Retaining only a running summary makes earlier interpretations difficult to inspect after revision; retaining only episode records leaves recurring procedures distributed across a growing history. The page offers a compact view of current advice while the archive preserves the accounts behind it. A later observation can therefore narrow or replace a rule without deleting the earlier record. This is an organizational property, not a guarantee that a rule is true. Both stores contain generated text, and checking an explanation still requires the original actions and observations. The ablations assess whether this additional organization also improves score.

3.4 Actor-requested retrieval

Retrieval reconnects a stored procedure to the actor's current decision. The actor queries memory using its room, object, or goal; similarity ranking selects complete entries rather than isolated matching sentences. Keeping the surrounding sequence and conditions lets the actor assess whether the earlier route applies and how to continue it. Source links allow the supporting attempt to be inspected when advice is uncertain. The actor may follow, adapt, or ignore the returned text: retrieval supplies contextual guidance rather than an executable policy. Appendix C.2 gives the ranking, thresholds, and query budget.

4 Experiments

4.1 Experimental setup

Benchmarks and budgets. Jericho tests sequential exploration with sparse rewards and irreversible failures (Hausknecht et al., 2020). We use the six-game panel of EvoTest (He et al., 2026), with ten

attempts per game and a 110-step interaction budget per attempt. Cross-attempt memory is shared across the ten attempts of a game and reset between games. ScienceWorld tests scientific procedures in an interactive text environment (Wang et al., 2022). We evaluate 211 task–variation pairs, each with up to five 100-step attempts; a pair stops once it reaches a score of 100. Model parameters are fixed in all conditions. Appendix A gives the full protocols.

Conditions. The matched control, *ReAct* (Yao et al., 2023), shares the backbone and interaction budget with BISE but runs the classic reasoning-and-action loop without cross-attempt memory. We also run RecMem (Dai et al., 2026) through its Jericho adaptation, and five conditions on the EvoTest scaffold: Naive, Sliding window, Summary, RAG, and EvoTest (He et al., 2026). EvoTest revises prompts, state-extractor code, and cross-attempt memory, selecting configurations with an upper-confidence-bound rule. The original method uses o3 as its Evolver; our reproduction uses the shared backbone for all components. In BISE, the same backbone serves as actor and curator. The main comparison uses the GLM-5.3 backbone and the same interaction budget; the matched comparison is repeated on DeepSeek-v4-Pro and GLM-4.7. Published EvoTest and TALES scores, obtained with other backbones and attempt budgets, appear in the reference block of Table 1.

Metric and replication. For Jericho, let $s_{g,e}$ be the highest score reached during attempt e on game g and m_g the reference score for that game (Table 1). We report the mean normalized score $\overline{S} = \frac{100}{|\mathcal{G}|N} \sum_{g \in \mathcal{G}} \sum_{e=1}^{N} s_{g,e}/m_g$ with $N = 10$, and the final-attempt score, which uses $e = N$ only. Recorded scores can exceed the reference, so normalized scores are not clipped. Differences between conditions are reported in percentage points. Because the attempts of a run share accumulated memory, the replication unit is one independent sequential run of ten attempts per game. The GLM-5.3 full system, gist-only arm, and EvoTest reproduction each have three runs; program pre-extraction and RecMem have two, and the remaining conditions have one.

4.2 Performance results

Memory improves repeated-task performance. On GLM-5.3, BISE reaches a mean normalized score of 55.7% and a final-attempt score of 68.0% (three runs; Table 1). The mean gains are 22.1 points over matched ReAct and 18.0 over reproduced EvoTest (37.7 ± 2.6, three runs). Thus the benefit extends beyond adding memory to a fixed actor loop to comparison with an adaptive agent system.

These comparisons answer different questions. The matched ReAct contrast measures the effect of adding the complete memory procedure; EvoTest compares two systems under a shared backbone and interaction budget. Neither identifies which memory component explains the gain. We therefore examine robustness across settings before using controlled variants to assess representation choices in §4.3.

Consistency across backbones and games. BISE exceeds ReAct in mean and final-attempt score on all three backbones and in 15 of 18 game-level comparisons; the exceptions are Zork 1 with DeepSeek-v4-Pro and Balances and Temple with GLM-4.7. Excluding any single game keeps the GLM-5.3 margin over EvoTest positive: 13.7 to 22.1 points (Appendix Table A3); no single game determines the sign of the comparison. The ranking also holds across aggregation schemes: BISE ranks first under unweighted mean, per-game median, and reference-weighted total score (Appendix Table A2).

The aggregate advantage therefore spans games and backbones, while individual exceptions show that stored experience does not help every task equally.

The learning curves show gains emerging as experience accumulates, with different discovery times across games (Figure 2). Detective improves markedly by attempt 4, and all three Library runs reach the reference score by attempt 8. Other games plateau below their references. Memory thus supports reuse of discovered progress, but does not remove the difficulty of finding a successful route in the first place.

Gains emerge early and persist. BISE exceeds reproduced EvoTest's final-attempt score from attempt 3 onward: its attempt-3 mean is 47.0%, compared with EvoTest's attempt-10 mean of 45.9%,

Table 1: Jericho performance on six games, ten attempts per game in our evaluation. Game columns show mean raw scores; parenthesized values are the normalization references. Mean normalizes each attempt's score by its game reference before averaging; Final uses the last attempt only. The GLM-5.3 BISE and EvoTest‡ rows average three runs each and RecMem† two (N=10 attempts per game per run); Run s.d. is 2.1 for BISE, 2.6 for EvoTest, and 0.3 for RecMem. Other rows in our evaluation are single runs. Bold marks the best value within each evaluated backbone block. †Our reproduction of the original method under our protocol (Appendix A). ‡Reproduced on the EvoTest scaffold (He et al., 2026). The top reference block uses different backbones and attempt budgets; TALES scores come from its online leaderboard (Cui et al., 2025; Microsoft, 2026). Gemini-2.5-F, Claude-4-S, and Claude-3.7-R abbreviate Gemini-2.5-Flash, Claude-4-Sonnet, and Claude-3.7-Sonnet with reasoning. EvoTest game scores are rescaled from published normalized AUCs; its Gemini-2.5-F Final is estimated from the episode-50 learning curves.

Method	Backbone	Det. (360)	Lib. (30)	Zork 1 (350)	Zork 3 (7)	Bal. (51)	Tem. (35)	**Mean** (%)	**Final** (%)
External references:									
EvoTest (N=50)	Gemini-2.5-F	338.4	23.1	49.0	2.5	16.3	10.9	47.0	64.0
EvoTest (N=50)	Claude-4-S	342.0	24.0	56.0	2.7	17.9	11.9	50.0	–
TALES (N=1)	Claude-3.7-R	259.2	11.4	49.0	4.3	24.5	12.6	45.0	–
Our evaluation:									
ReAct †	GLM-5.3	60.0	22.6	46.6	3.4	14.0	7.1	33.6	33.4
Naive‡	GLM-5.3	20.0	0.0	3.5	0.0	5.0	0.0	2.7	2.6
Sliding window‡	GLM-5.3	93.0	10.0	38.9	2.4	8.5	6.2	23.2	18.8
Summary‡	GLM-5.3	76.0	11.4	48.0	2.6	11.0	6.2	24.9	20.9
RAG‡	GLM-5.3	82.0	9.0	40.7	2.4	5.5	5.3	20.8	19.1
RecMem†	GLM-5.3	111.5	19.6	48.4	2.2	11.2	10.6	32.4	35.7
EvoTest‡	GLM-5.3	**261.3**	14.3	50.2	2.3	17.5	8.4	37.7	45.9
BISE	GLM-5.3	254.7	**26.3**	**86.3**	**4.1**	**21.9**	**17.2**	**55.7**	**68.0**
ReAct	DeepSeek-v4-Pro	150.0	23.5	**55.8**	2.8	14.0	10.5	38.9	43.2
BISE	DeepSeek-v4-Pro	**208.0**	**25.3**	52.2	**2.9**	**22.5**	**14.2**	**47.2**	**52.4**
ReAct	GLM-4.7	69.0	20.5	43.8	2.0	**13.5**	**15.1**	33.0	26.1
BISE	GLM-4.7	**212.0**	**28.0**	**51.6**	**3.3**	9.5	12.6	**44.8**	**51.2**

both over three runs (Table 1; Appendix Table A8). The advantage is therefore visible before the interaction budget is exhausted, rather than arising only from a strong final attempt. This comparison concerns environment interactions; it does not establish lower token use or inference cost. Published fifty-attempt EvoTest results remain contextual references because their backbones differ. Together, these curves show that the performance gain emerges early and persists across subsequent attempts.

Variation across runs. Across the three GLM-5.3 runs (Appendix Figure 6), the mean over attempts 6–10 is 64.7%, and the final-attempt s.d. is 2.4 points. The time needed to discover a high-scoring route varies; §4.4 examines whether those routes are reproduced after reset.

ScienceWorld: progress beyond initial competence. The 211-pair panel reaches a mean cumulative best score of **95.5**, with 187 pairs reaching 100 within five attempts (Figure 3). Since 148 pairs already succeed before memory exists, the informative learning subset is the 63 initially imperfect pairs. Their mean rises from **44.4 on attempt 1 to 70.0 on attempt 2**, a 25.7-point increase using unrounded means; 39 improve, 22 are unchanged, and two decline. By attempt 5, 39 reach 100 and 24 remain below. This shows substantial recovery after an imperfect start, although a repeated-attempt no-memory control is needed to attribute that recovery to memory.

Procedural tasks offer scope for reuse. All seven eligible task families improve in mean score, with substantial variation (Appendix Table A1). Matter, the largest subset with 20 pairs, rises from 23.2 to 69.1; Biology growth gains 8.7 points over 16 pairs, and Electricity gains 8.1 over 13. Several tasks require preparation followed by a later outcome, such as changing a substance's state or tending

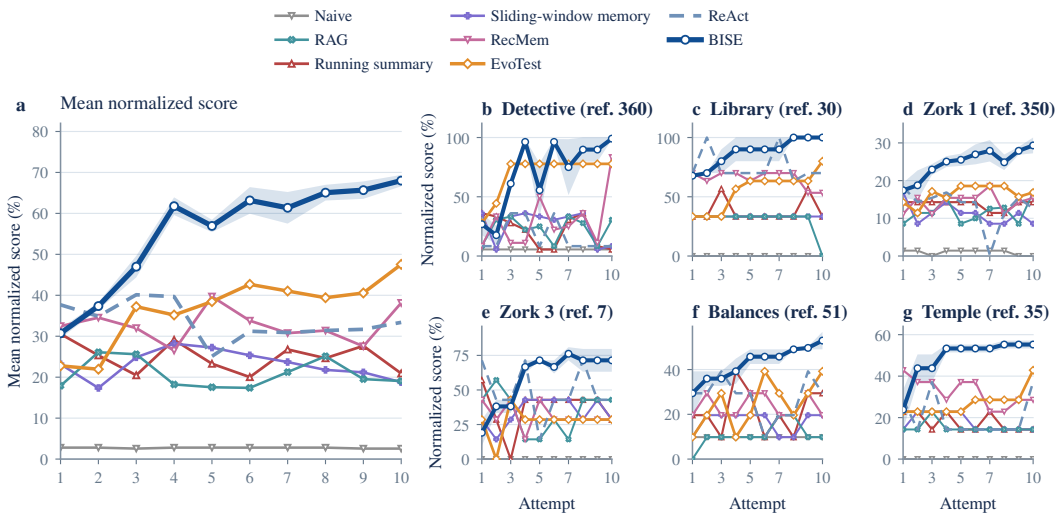

Figure 2: Jericho performance over successive attempts. a, Cross-game normalized score. **b–g**, Normalized score for each game. BISE curves show the mean and s.e.m. over three runs of the final configuration; the EvoTest curve shows its second run, while its Table 1 row averages three. Other curves show single runs under the same protocol. RecMem appears in all six games.

a plant. Family means cannot isolate sequence reuse, and small subsets limit their interpretation. Together, the results show progress after initial failure across several scientific procedures, extending the evidence beyond navigation tasks.

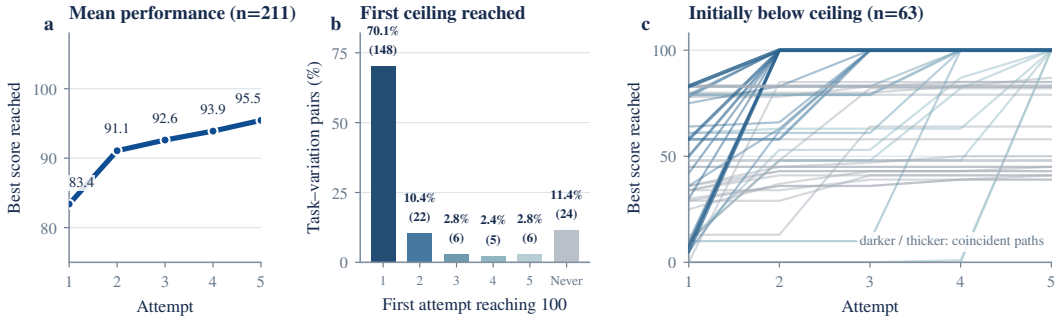

Figure 3: ScienceWorld performance across attempts. a, Mean cumulative best score across all 211 pairs. **b**, Attempt at which each pair first reaches 100. **c**, Cumulative best-score trajectories for the 63 pairs initially below 100. After a pair reaches 100, the value is carried forward for plotting only.

4.3 Ablations: Representation and Episode Context

Which structure should survive when a trajectory becomes memory? Table 2 compares verbatim records, local triples, gists without order, and ordered gists without a schema page. These variants distinguish the value of retaining experience from that of its representation, and test whether consolidation adds performance beyond the archive.

Table 2: Memory representation and extraction ablations (GLM-5.3). n is the number of independent runs; $\pm$ denotes run s.d., and Δ is relative to BISE. [a]The ten no-memory attempts form one run and are not ten run-level replications.

Arm	Representation	n	Mean (%)	Δ
Representation variants				
No memory (ReAct)	–	1[a]	33.6	-22.1
Verbatim event log	raw record stream, no abstraction	1	49.6	-6.1
Atomized triples	unordered (room, command, Δ)	1	49.7	-6.0
De-temporalized gist	gist content, order stripped	1	48.2	-7.5
Gist-only	ordered narrative, no schema page	3	55.0 ± 2.7	-0.7
BISE	episode-indexed gist + schema page	3	$\mathbf{55.7 \pm 2.1}$	–
Extraction stage				
Program pre-extraction	deterministic parse feeds the curator	2	53.2 ± 1.2	-2.5

Ordered gists retain most of the gain. Removing order lowers the mean by 7.5 points; gist-only stays within 0.7 of the full system. These contrasts favor episode structure (Hikosaka et al., 1999; Lee & Wilson, 2002), although the single-run order variant needs replication. The page shows no clear extra score benefit; below we examine its organization and revision.

Context beyond reward. Direct reading exceeds event pre-extraction by 2.5 points (three versus two runs). Together, these comparisons favor retaining action order and context around reward.

4.4 What the final trajectories retain and reuse

Across 180 trajectories, action sequences stabilize and knowledge pages remain compact (Figure 4a,b). At attempt 10, page word-edit distance is 31.9%, despite only 3.2% change in length (panel c): stable size does not mean static text. In Library, all 15 attempts after the first win remain successful with fewer actions (panel d).

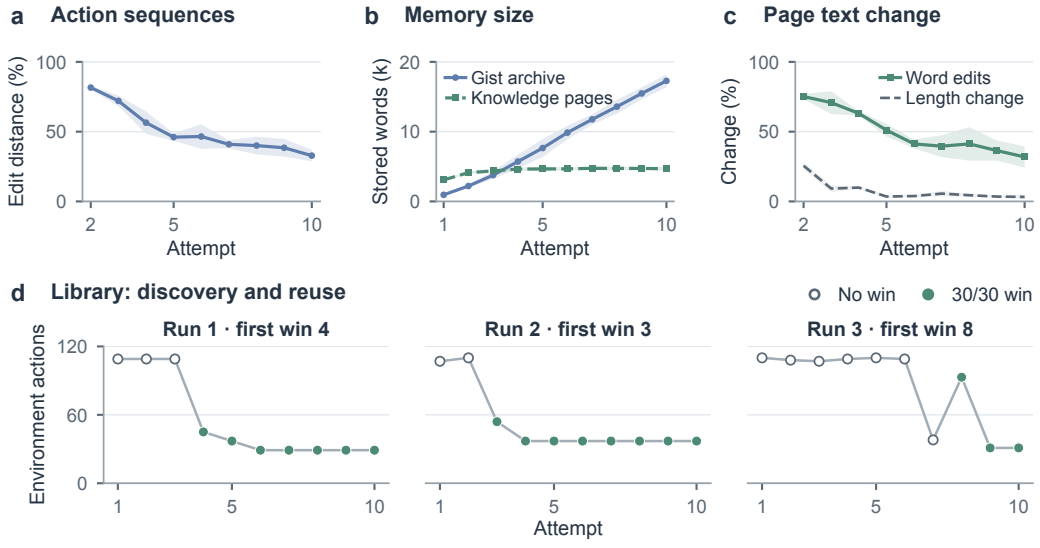

Figure 4: Behavioral stabilization, memory revision, and route reuse. a, Normalized command-sequence edit distance between attempts. **b,** Stored words. **c,** Page word edits versus absolute length change, both normalized. Curves show mean $\pm$ s.d. across three runs of six games (games averaged in a,c; summed in b). **d,** All ten Library attempts per run; filled points reach 30/30 with an explicit win, open points do not win. Short failures are retained. Text edits do not measure knowledge accuracy. Definitions and sources: Appendix B.

Text edits include rewording and source links; Figure 5 examines two revisions and the actions that follow.

a Library: correct a mistaken rule

1 Episode evidence

Earlier claim · A2 (incorrect)
The biography must be copied before leaving.

New evidence · A4
The actor exits with the **original** after obtaining an encyclopedia printout for the librarian.

2 Revised memory · after A4

Replace the copying rule with this route:

1. `ask technician about gates` (+2)
   Technician leaves.
2. `examine printers` (+5)
   Obtain encyclopedia printout.
3. `give librarian printout` (+2)
   Gates allow the original to leave.

Source link: A4.

3 Subsequent execution

A5 receives two page matches and repeats the route.

Attempt	A4	A5	A6–10
Actions	**45**	**37**	**29**

30/30 on every attempt from A4 onward.

b Zork 3: retain a prerequisite

1 Episode evidence

Same command: `touch table` (+1).

A5 · with a torch:
Safe arrival in the Timber Room.

A6 · without light:
Dark arrival; `east` ends in death.

2 Revised memory · after A6

Retain conditions around the reward:

Preparation
`take torch` (0); carry it lit.

State check
Confirm the table's indicator is "I".

Scoring action
`touch table` (+1); keep light on arrival.

Source links: A4, A6.

3 Subsequent execution

A7 takes the torch, touches the table, and arrives in the lit Timber Room.

6/7 on all four attempts A7–10.
No recorded deaths in those attempts.

Figure 5: What changed in memory, and what happened next. Read each column downward: episode evidence, saved page update, then subsequent execution. Both cases are from run 1; A denotes attempt and parentheses give score changes. Text is condensed from the records. Stored conditions are not exhaustively validated, and these selected cases do not establish causal effects. Full records are in Appendix B.1.

Revision preserves useful context. The cases illustrate two benefits of separating experience from advice: Library can replace an explanation while retaining its history, and Zork 3 can retain preparation that immediate reward would miss. Saved updates precede the later retrievals and actions, making these links inspectable. Source-linked revision thus lets advice change while its earlier basis remains available for inspection.

Limitations. Evaluation covers six Jericho games and one ScienceWorld panel. The matched control and several representation variants have one run and need replication. Full-system comparisons cannot isolate event order or extraction effects; the three-run ablation shows no clear extra score benefit from schema pages. Without a repeated-attempt no-memory control, ScienceWorld gains cannot be attributed to memory alone. Selected cases, retrieval hits, and source links do not establish general accuracy or applicability. Task drift and longer histories remain untested; biological connections are functional analogies. Equal interaction budgets need not imply equal token use or inference costs. Published references differ in backbone and budget.

5 Conclusion

Our results support preserving ordered experience as a reusable memory unit. BISE improves Jericho score over matched ReAct by 22.1 points, with gains on two further backbones; the ablations favor episode context but show no clear extra score benefit from the schema page. The trajectory audits establish its complementary role: organizing advice that can be revised while earlier accounts remain accessible. The resulting design principle is to retain the context behind a procedure and keep its generalization open to correction.

Ethics statement. The experiments use simulated text environments; recorded hazards guide avoidance in these settings, and transferring such guidance to settings where a wrong decision could cause real harm would require validation for that setting.

AI use statement. Large language models assisted with drafting, language editing, consistency checks, and analysis code. The authors are responsible for the text, experimental records, analyses, and claims.

Reproducibility statement. The appendix describes the evaluation protocols, memory-use instructions, and retrieval settings. The supplementary material includes per-attempt score tables, available memory snapshots, and analysis and plotting scripts. The final-run trajectory audit includes per-attempt metrics, source hashes, and exact case records for the three BISE runs; the data index distinguishes these records from remaining baseline and ScienceWorld prompt gaps.

REFERENCES

Alan Baddeley. Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63:1–29, 2012. doi: 10.1146/annurev-psych-120710-100422.

C. J. Brainerd and V. F. Reyna. Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 11(5):164–169, 2002. doi: 10.1111/1467-8721.00192.

Minghao Chen, Yihang Li, Yanting Yang, Shiyu Yu, Binbin Lin, and Xiaofei He. AutoManual: Constructing instruction manuals by LLM agents via interactive environmental learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 589–631, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/0142921fad7ef9192bd87229cdafa9d4-Abstract-Conference.html. arXiv:2405.16247.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. Mem0: Building production-ready ai agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*, 2025.

Christopher Zhang Cui, Xingdi Yuan, Ziang Xiao, Prithviraj Ammanabrolu, and Marc-Alexandre Côté. TALES: Text adventure learning environment suite. *arXiv preprint arXiv:2504.14128*, 2025. doi: 10.48550/arXiv.2504.14128. URL https://arxiv.org/abs/2504.14128.

Zijie Dai, Shiyuan Deng, Sheng Guan, Yizhou Tian, Xin Yao, Xiao Yan, and James Cheng. RecMem: Recurrence-based memory consolidation for efficient and effective long-running LLM agents. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 32353–32376. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.findings-acl.1619. URL https://aclanthology.org/2026.findings-acl.1619/.

Kamran Diba and Gy"orgy Buzsáki. Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, 10(10):1241–1242, 2007. doi: 10.1038/nn1961.

Matthias J. Gruber, Maureen Ritchey, Shao-Fang Wang, Manoj K. Doss, and Charan Ranganath. Post-learning hippocampal dynamics promote preferential retention of rewarding events. *Neuron*, 89(5):1110–1120, 2016. doi: 10.1016/j.neuron.2016.01.017.

Bernal Jiménez Gutiérrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. HippoRAG: Neurobiologically inspired long-term memory for large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. arXiv:2405.14831.

Matthew Hausknecht, Prithviraj Ammanabrolu, Marc-Alexandre Côté, and Xingdi Yuan. Interactive fiction games: A colossal adventure. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pp. 7903–7910, 2020. doi: 10.1609/aaai.v34i05.6297.

Yufei He, Juncheng Liu, Yue Liu, Yibo Li, Tri Cao, Zhiyuan Hu, Xinxing Xu, and Bryan Hooi. EvoTest: Evolutionary test-time learning for self-improving agentic systems. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://arxiv.org/abs/2510.13220. arXiv:2510.13220.

Okihide Hikosaka, Hiroyuki Nakahara, Miya K. Rand, Katsuyuki Sakai, Xiaofeng Lu, Kae Nakamura, Shigehiro Miyachi, and Kenji Doya. Parallel neural networks for learning sequential procedures. *Trends in Neurosciences*, 22(10):464–471, 1999. doi: 10.1016/S0166-2236(99)01439-3.

Dharshan Kumaran, Demis Hassabis, and James L. McClelland. What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends in Cognitive Sciences*, 20 (7):512–534, 2016. doi: 10.1016/j.tics.2016.05.004.

Albert K. Lee and Matthew A. Wilson. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, 36(6):1183–1194, 2002. doi: 10.1016/S0896-6273(02)01096-6.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Yibo Li, Zijie Lin, Ailin Deng, Xuan Zhang, Yufei He, Shuo Ji, Tri Cao, and Bryan Hooi. Just-in-time reinforcement learning: Continual learning in LLM agents without gradient updates. *arXiv preprint arXiv:2601.18510*, 2026. doi: 10.48550/arXiv.2601.18510. URL https://arxiv.org/abs/2601.18510.

James L. McClelland, Bruce L. McNaughton, and Randall C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995. doi: 10.1037/0033-295X.102.3.419.

Microsoft. TALES leaderboard. https://microsoft.github.io/tale-suite/, 2026. Online scores; accessed 25 September 2026.

Morris Moscovitch, Roberto Cabeza, Gordon Winocur, and Lynn Nadel. Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annual Review of Psychology*, 67(1): 105–134, 2016. doi: 10.1146/annurev-psych-113011-143733.

Lynn Nadel and Morris Moscovitch. Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, 7(2):217–227, 1997. doi: 10.1016/S0959-4388(97)80010-4.

Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G. Patil, Ion Stoica, and Joseph E. Gonzalez. MemGPT: Towards LLMs as operating systems. *arXiv preprint arXiv:2310.08560*, 2023.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 1–22, 2023. doi: 10.1145/3586183.3606763.

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 (NeurIPS)*, volume 36, 2023.

Larry R. Squire. Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3):171–177, 2004. doi: 10.1016/j.nlm.2004.06.005.

Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. Cognitive architectures for language agents. *Transactions on Machine Learning Research*, 2024.

Mirac Suzgun, Mert Yuksekgonul, Federico Bianchi, Dan Jurafsky, and James Zou. Dynamic cheatsheet: Test-time learning with adaptive memory. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 7080–7106. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.eacl-long.333.

Timothy J. Teyler and Jerry W. Rudy. The hippocampal indexing theory and episodic memory: Updating the index. *Hippocampus*, 17(12):1158–1169, 2007. doi: 10.1002/hipo.20350.

Dorothy Tse, Rosamund F. Langston, Masaki Kakeyama, Ingrid Bethus, Patrick A. Spooner, Emma R. Wood, Menno P. Witter, and Richard G. M. Morris. Schemas and memory consolidation. *Science*, 316(5821):76–82, 2007. doi: 10.1126/science.1135935.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models. *Transactions on Machine Learning Research*, 2024. arXiv:2305.16291.

Ruoyao Wang, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. ScienceWorld: Is your agent smarter than a 5th grader? In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 11279–11298. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.emnlp-main.775.

Tianxin Wei, Noveen Sachdeva, Benjamin Coleman, Zhankui He, Yuanchen Bei, Xuying Ning, Mengting Ai, Yunzhe Li, Jingrui He, Ed H. Chi, Chi Wang, Shuo Chen, Fernando Pereira, Wang-Cheng Kang, and Derek Zhiyuan Cheng. Evo-Memory: Benchmarking LLM agent test-time learning with self-evolving memory. *arXiv preprint arXiv:2511.20857*, 2025. doi: 10.48550/arXiv.2511.20857. URL https://arxiv.org/abs/2511.20857.

Matthew A. Wilson and Bruce L. McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679, 1994. doi: 10.1126/science.8036517.

Gordon Winocur and Morris Moscovitch. Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, 17(5):766–780, 2011. doi: 10.1017/S1355617711000683.

Weijia Xu, Alessandro Sordoni, Chandan Singh, Zelalem Gero, Michel Galley, Xingdi Yuan, and Jianfeng Gao. Test-time learning with an evolving library. *arXiv preprint arXiv:2605.14477*, 2026. doi: 10.48550/arXiv.2605.14477. URL https://arxiv.org/abs/2605.14477.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. A-Mem: Agentic memory for LLM agents. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2502.12110.

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 (ICLR)*, 2023.

Jeffrey M. Zacks, Nicole K. Speer, Khena M. Swallow, Todd S. Braver, and Jeremy R. Reynolds. Event perception: A mind–brain perspective. *Psychological Bulletin*, 133(2):273–293, 2007. doi: 10.1037/0033-2909.133.2.273.

Guibin Zhang, Muxin Fu, and Shuicheng Yan. MemGen: Weaving generative latent memory for self-evolving agents. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://arxiv.org/abs/2509.24704. arXiv:2509.24704.

Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. ExpeL: LLM agents are experiential learners. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38 (17):19632–19642, 2024. doi: 10.1609/aaai.v38i17.29936.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. MemoryBank: Enhancing large language models with long-term memory. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(17):19724–19731, 2024. doi: 10.1609/aaai.v38i17.29946.

A Evaluation Protocols and Supporting Analyses

Matched control. ReAct (ReAct) shares the backbone, interaction budget, rolling context, and repetition detector with BISE. The full condition adds the gist archive, the schema page, and retrieval. The EvoTest-scaffold conditions share the game panel, backbone, and interaction budget but use their own scaffold.

Jericho baselines. All GLM-5.3 conditions use ten 110-step attempts per game. RecMem provides per-attempt scores for all six games; its condition adapts the memory procedure to Jericho.

The EvoTest scaffold uses temperature 0.4. Naive has no memory; Sliding window keeps the last 30 states; Summary maintains a two- to three-sentence progress summary updated at each step; RAG retrieves the three most similar within-attempt interactions with text-embedding-3-small; EvoTest additionally revises its reusable configuration between attempts. Only the four static baselines are restricted to within-attempt memory.

ScienceWorld and statistical units. Each task–variation pair receives up to five 100-step attempts and up to 30 memory-update steps after each attempt. A score of 100 ends the sequence. The attempt 1-to-2 comparison uses all 63 pairs initially below 100, each as one paired observation. The Wilcoxon signed-rank test is two-sided and excludes zero differences. Family-wise tests apply Holm correction across the seven eligible families. Cumulative-best plots carry 100 forward after completion. Jericho uses the independent sequential run as its replication unit.

Table A1: ScienceWorld change from attempt 1 to 2 for the 63 pairs initially below 100. Family p values are two-sided Wilcoxon signed-rank tests with Holm correction across the seven eligible families; the pooled test is unadjusted. Changes use unrounded means. Tests concern within-system changes across attempts. †Every Measurement pair reaches 100 on the first attempt.

Task family	n	Attempt 1	Attempt 2	Change	p
Matter (state change)	20	23.2	69.1	+45.9	0.010
Measurement	0†	–	–	–	–
Electricity	13	74.1	82.2	+8.1	1.000
Classification	2	79.0	91.5	+12.5	1.000
Biology (growth)	16	35.8	44.4	+8.7	0.034
Chemistry	2	36.0	100.0	+64.0	1.000
Biology (lifespan)	5	69.8	100.0	+30.2	0.313
Biology (life stages)	5	43.4	73.6	+30.2	1.000
All eligible pairs	63	44.4	70.0	+25.7	3.73×10^{-8}

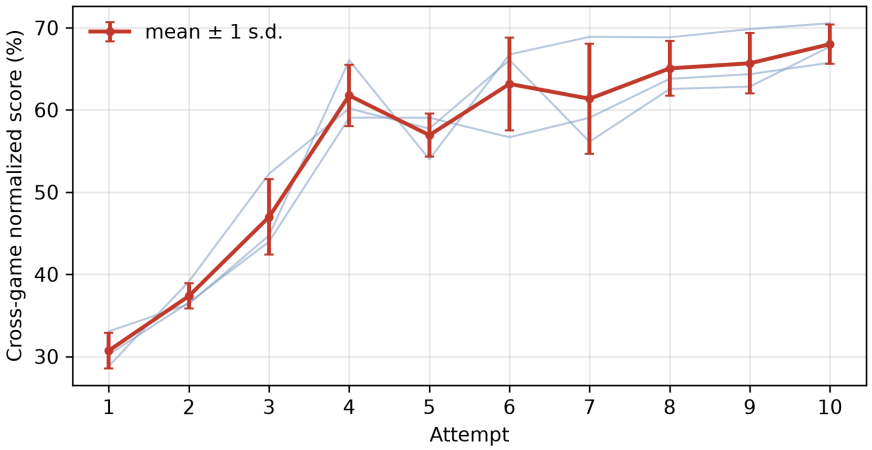

Figure 6: Variation across the three GLM-5.3 runs of Table 1. Thin lines are individual runs; the red line and error bars are the mean and one standard deviation of the cross-game normalized score at each attempt. The runs differ in learning speed; the final-attempt s.d. is 2.4 points.

Table A2: Robustness of the comparisons to the aggregation scheme. Mean: unweighted mean of per-game normalized scores (as in Table 1); Median: median of the six per-game normalized means; Weighted: total raw score over total reference score. BISE ranks first under all three aggregations on all three backbones.

Method	Backbone	Mean	Median	Weighted
BISE	GLM-5.3	**55.7**	**54.0**	**49.3**
EvoTest	GLM-5.3	37.7	33.8	42.5
ReAct	GLM-5.3	33.6	23.9	18.5
RecMem	GLM-5.3	32.4	30.7	24.6
BISE	DeepSeek-v4-Pro	**47.2**	**42.8**	**39.0**
ReAct	DeepSeek-v4-Pro	38.9	35.0	30.8
BISE	GLM-4.7	**44.8**	**41.6**	**38.1**
ReAct	GLM-4.7	33.0	27.5	19.7

Table A3: Leave-one-game-out comparison with EvoTest (our evaluation). Both methods average three runs. The row with no game excluded uses the run-level means from Table 1; leave-one-game-out rows use its displayed per-game means, and differences are computed before rounding.

Excluded game	BISE	EvoTest	Difference
None	55.7	37.7	18.0
Detective	52.7	30.6	22.1
Library	49.3	35.6	13.7
Zork 1	61.9	42.3	19.6
Zork 3	55.0	38.6	16.5
Balances	58.3	38.3	20.0
Temple	57.0	40.4	16.7

Table A4: Normalized-mean intermediate values behind Table 1. Per-game normalized mean (each attempt's score divided by the game reference, averaged over attempts and runs), then the cross-game mean of these six values (the Mean column) and the normalized last attempt (the Final column).

Method	Det.	Lib.	Z1	Z3	Bal.	Tem.	Mean / Final
BISE (n=3)	70.7	87.8	24.7	59.0	42.9	49.0	55.7 / 68.0
EvoTest (n=3)	72.6	47.8	14.3	33.3	34.3	24.1	37.7 / 45.9
ReAct (n=1)	16.7	75.3	13.3	48.6	27.5	20.3	33.6 / 33.4

B FINAL-TRAJECTORY AUDIT

Scope and reproducibility. The audit reads the supplied final-data archive and joins trajectories by experiment, game, and attempt: abl41, abl45, and abl47 correspond to runs 1–3 in Table A8. All 180 play trajectories and their score maxima agree with the run manifests and plotting arrays. The abl47 run was resumed; some manifest paths point to the restart timestamp even for skipped, previously completed attempts. Matching by game and attempt recovers exactly one saved play trace for each cell. No trajectories are pooled across configurations. The script `scripts/analyze_final_trajectories.py` exports all attempt metrics, exact case records with JSONL line numbers, and source-member SHA-256 hashes to `data/final_trajectories/`.

Counting rules. An action is one `tale_jericho_action` record; retrieval calls, curator tool calls, and the game engine's move counter are excluded. The audit contains 14,465 such actions. A retrieval hit means `total_matches` is positive, not that the returned text is relevant or followed. Query counts include attempted calls that return no matches, including calls beyond the retrieval budget. Word counts use whitespace splitting of the complete saved Markdown files, including

headings, links, addenda, and auxiliary entries. They compare final stored representations, not the raw trajectory or actual model input tokens. The six principal schema pages per run contain 635–827 words each; run 1 additionally saves a 510-word Library route page. This accounts for the difference between principal and total knowledge-page counts in Table A5.

Table A5: Saved memory size for the final configuration. Counts include every Markdown file in the corresponding archive or knowledge directory. Ratio divides archive words by all knowledge-page words.

Run	Archive words	Principal pages	All knowledge pages	Ratio
1 (abl41)	17,327	4,494	5,004	3.46
2 (abl45)	16,324	4,503	4,503	3.63
3 (abl47)	18,192	4,549	4,549	4.00

Table A6: Library route reuse across the three final runs. Actions count logged environment calls, excluding retrieval. The last column counts full-score attempts strictly after the first 30/30. All counted completions also contain the game's explicit win message. These are within-run observations, not independent replications.

Run	First 30/30 (attempt)	Actions to completion		Later 30/30 (attempts)
		First	Attempt 10	
1 (abl41)	4	45	29	6/6
2 (abl45)	3	54	37	7/7
3 (abl47)	8	93	31	2/2

Action-sequence dynamics. Figure 4a uses every play trajectory, without filtering for success. For each game and run, we compare the complete command sequence at each attempt with the preceding one using unit-cost Levenshtein distance (insertions, deletions, and substitutions), divided by the longer sequence length. Commands are lowercased and whitespace-normalized; synonyms such as `n` and `north` remain distinct. There is no distance for the first attempt. We average the six games within each run before computing means and sample standard deviations across three runs; the 162 adjacent pairs are not independent replicates. Mean distance decreases from 0.816 ± 0.014 at attempt 2 to 0.328 ± 0.042 at attempt 10. This measures literal sequence stability, not semantic equivalence or optimality. The aggregate action-count series, retained in the source data, changes from 71.0 ± 4.9 to 79.3 ± 5.3, while normalized score rises from 30.7 to 68.0; shorter failed attempts prevent interpreting action count alone as efficiency.

Memory growth reconstruction. We reconstruct the memory state after each attempt from logged literal writes and edits, including actor-side addenda before curation. Counts sum all archive and knowledge files across the six games within each run. At attempts 1 and 10, mean archive size is 947 and 17,281 words; mean knowledge size is 3,085 and 4,685. All 50 reconstructed final files match the saved files' word counts. Four curator phases in run 3 (Library 7, Zork1 5, Zork3 8, Balances 6) contain only API-limit failures in their manager logs and execute no writes; their existing state is retained, not interpolated. A failed write to an invalid path is excluded using the recorded shell error. The analysis script `scripts/analyze_trajectory_dynamics.py` parses these records as data without executing their commands and exports command sequences, per-attempt word counts, write provenance, and aggregate statistics. These curves describe retained file size, not token costs or validated knowledge.

Knowledge-page text revision. We reconstruct all knowledge files after each attempt using the audited literal writes, including actor and curator edits, and match all 180 boundary word counts and final saved token sequences. For each file, we compute unit-cost Levenshtein distance between consecutive whitespace-token sequences, preserving case and punctuation. We sum edits across files and divide by the sum of the longer token counts in each file pair; added files are compared with empty text. Length change is the absolute change in total knowledge words divided by the larger total. Both measures are averaged over six games within each run before reporting three-run means and sample standard deviations. At attempt 10, token-edit distance is $31.9 \pm 7.6\%$, while length change is $3.2 \pm 1.5\%$. Over attempts 8–10, mean token-edit distance is $36.5 \pm 8.7\%$, using one window mean per

run. Edits include wording, Markdown tokens, and source links; they do not count new facts, semantic corrections, or verified knowledge. The script `scripts/analyze_memory_revision.py` exports file-, attempt-, and run-level measurements without executing logged commands.

Historical audit. The earlier pre-extraction audit is retained separately for provenance. Its command-and-location lexical rule covers 59 of 72 unique score events and 9 of 11 hazards. These are not coverage estimates for the final configuration, whose new audit measures retrieval, stored size, and the action sequences below.

B.1 Contemporaneous evidence for the two cases

Both cases use run 1. Table A7 indexes actions within each play trajectory. The accompanying JSON records retain the commands, before/after observations, score changes, and source paths, so the examples can be checked without relying on a curator's description. The table paraphrases observations; quoted commands retain their recorded spelling.

Table A7: Action-level evidence from final run 1. A denotes attempt, and action positions exclude non-environment tool calls. Memory-update evidence is described below.

Game, attempt	Action positions	Recorded transition
Library, A4	$24 \to 25 \to 30$	Ask technician about gates ($+2$), examine printers ($+5$), give librarian printout ($+2$).
Library, A5	$17 \to 18 \to 23$	Repeat the chain; the last command is `give reference printout`. All three score increments recur.
Zork 3, A4	47	`touch table`: $+1$, arrival at the Sacrificial Altar, then explicit death text.
Zork 3, A5	$48 \to 50$	Take torch (0), indicator changes to "I"; `touch table` ($+1$) reaches the Timber Room.
Zork 3, A6	$34 \to 35$	`touch table` ($+1$) reaches darkness; `east` (0) produces the grue death.
Zork 3, A7	$21 \to 22 \to 24$	Turn off lamp (0), put down lamp (0), enter lake ($+1$); the lamp stays ashore.
Zork 3, A7	$31 \to 33 \to 43$	Take torch (0), touch table ($+1$), later turn on the recovered lamp (0): the lamp works.

Library update and reuse. Line 1 of the attempt-4 curation trace reads the existing page, whose opening says the missing points require the copier. Line 6 appends the new episode account; line 7 rewrites the page to label the copier a red herring and record the technician–printer–librarian chain with attempt-4 links. These writes precede the attempt-5 retrieval, which returns two knowledge-page matches in round 1. We do not infer the earlier memory state from the final page alone. The 15/15 later full-score attempts in Table A6 come from all three runs. Their observations explicitly contain the win message despite the wrapper setting `lost=true` in these Library completions; we therefore do not classify them as deaths from that flag.

Zork 3 update and limits. Lines 8–10 of the attempt-6 curation trace append the failure account and rewrite the page with the lamp, torch, and indicator conditions, before the attempt-7 query. All four attempts 7–10 reach 6/7 and end without a death flag or death text. However, the page also generalizes beyond the observations, including a claim that every non-"I" indicator phase is fatal. The selected traces establish contrasting observed outcomes and retained conditions, not exhaustive validation of that rule. Memory contents remain fallible, as the Library copying hypothesis also demonstrates. These cases were selected for an observable memory update followed by corresponding actions; they do not estimate the frequency or causal effect of such updates across tasks.

C Implementation and Prompts

C.1 Instructions for Memory Use and Consolidation

Actor. The actor is instructed to pursue the task until termination or the interaction limit. On later attempts, it consults memory for useful actions, checks retrieved procedures against the current observation, and explores alternatives when a suggestion fails or memory is incomplete. Recorded hazards are retrieved as advice, with the location where they were observed.

Curator. After each completed attempt, the curator appends one ordered gist containing the outcome, route, hazards, and conclusions, then rewrites the schema page from the previous page and the new evidence. The instructions set an 800-word target, require source attempts for every entry, and group scoring procedures, hazards, and unresolved cases. Actions with both scoring and fatal observations are recorded as unresolved cases rather than unconditional fatal rules. The gist archive grows incrementally; existing entries are preserved as written.

C.2 Control and Retrieval Details

Within-attempt controls. A repetition detector warns after the same command–location pair occurs five times within a ten-step window. It resets with the environment and is shared by the matched ReAct and BISE conditions, including the ablations.

Retrieval configuration. Candidates are embedded with BGE-M3 and ranked by cosine similarity. The implementation down-weights entries from other games by a factor of 0.3 when they are present. They are absent in this evaluation because memory is reset between games. The tool returns *complete* entries, preserving the prerequisites and continuation of an action sequence. It returns up to three entries per query, with score thresholds of 0.30 for the schema page, 0.20 for the gist archive, and 0.25 when a query draws from both. The configured budget is three queries per attempt. Attempted calls beyond the budget are still logged: four final-run attempts contain a fourth call, each returning zero matches.

D PER-ATTEMPT RESULTS

Table A8: Per-attempt GLM-5.3 scores on the six-game panel. The BISE rows are the three per-run trajectories of the final configuration; Table 1 reports their means.

Condition	Game	Ref.	1	2	3	4	5	6	7	8	9	10	Best	Mean
BISE (run 1)	Det.	360	130	30	210	350	90	350	350	360	360	360	360	259.0
	Lib.	30	19	21	21	30	30	30	30	30	30	30	30	27.1
	Zork1	350	59	73	73	95	95	110	105	74	95	110	110	88.9
	Zork3	7	2	3	3	5	5	5	6	6	6	6	6	4.7
	Bal.	51	20	20	20	25	25	25	25	25	25	25	25	23.5
	Tem.	35	5	13	13	18	18	18	18	20	20	20	20	16.3
BISE (run 2)	Det.	360	130	40	210	330	150	330	100	250	250	350	350	214.0
	Lib.	30	21	21	30	30	30	30	30	30	30	30	30	28.2
	Zork1	350	50	85	90	90	95	95	110	99	105	110	110	92.9
	Zork3	7	0	3	3	4	5	5	5	5	5	5	5	4.0
	Bal.	51	5	15	15	15	25	25	25	25	25	25	25	20.0
	Tem.	35	15	20	20	20	20	20	20	20	20	20	20	19.5
BISE (run 3)	Det.	360	30	120	240	360	360	360	360	360	360	360	360	291.0
	Lib.	30	21	21	21	21	21	21	21	30	30	30	30	23.7
	Zork1	350	74	39	78	78	78	78	78	88	93	88	93	77.2
	Zork3	7	2	2	2	5	5	4	5	4	4	4	5	3.7
	Bal.	51	20	20	20	20	20	20	20	25	26	31	31	22.2
	Tem.	35	5	13	13	18	18	18	18	18	18	18	18	15.7
ReAct	Det.	360	30	30	130	130	30	130	30	30	30	30	130	60.0
	Lib.	30	21	30	21	21	21	21	30	19	21	21	30	22.6
	Zork1	350	64	50	54	59	50	50	1	40	54	44	64	46.6
	Zork3	7	5	3	3	5	1	3	3	5	3	3	5	3.4
	Bal.	51	15	15	20	15	15	5	10	10	20	15	20	14.0
	Tem.	35	10	5	13	5	5	5	5	5	5	13	13	7.1

Table A9: Per-attempt GLM-5.3 scores for EvoTest-scaffold conditions. EvoTest runs 1, 2, and 3 are shown; its three-run aggregate appears in Table 1. Naive, Sliding window, Summary, and RAG each have one run.

Condition	Game	Ref.	1	2	3	4	5	6	7	8	9	10	Best	Mean
Naive	Det.	360	20	20	20	20	20	20	20	20	20	20	20	20.0
	Lib.	30	0	0	0	0	0	0	0	0	0	0	0	0.0
	Zork1	350	5	5	0	5	5	5	5	5	0	0	5	3.5
	Zork3	7	0	0	0	0	0	0	0	0	0	0	0	0.0
	Bal.	51	5	5	5	5	5	5	5	5	5	5	5	5.0
	Tem.	35	0	0	0	0	0	0	0	0	0	0	0	0.0
Sliding window	Det.	360	130	20	120	130	120	110	120	130	20	30	130	93.0
	Lib.	30	10	10	10	10	10	10	10	10	10	10	10	10.0
	Zork1	350	59	30	40	50	40	40	30	30	40	30	59	38.9
	Zork3	7	2	1	2	3	3	3	3	2	3	2	3	2.4
	Bal.	51	5	10	10	10	10	10	5	5	10	10	10	8.5
	Tem.	35	5	8	8	8	8	5	5	5	5	5	8	6.2
Summary	Det.	360	130	120	100	80	20	20	110	130	30	20	130	76.0
	Lib.	30	10	10	17	10	10	10	10	10	17	10	17	11.4
	Zork1	350	50	50	50	50	50	50	40	40	50	50	50	48.0
	Zork3	7	4	2	0	3	3	3	3	3	3	2	4	2.6
	Bal.	51	10	10	5	20	15	5	10	5	15	15	20	11.0
	Tem.	35	8	8	5	8	5	5	8	5	5	5	8	6.2
RAG	Det.	360	30	110	120	80	90	30	120	100	30	110	120	82.0
	Lib.	30	10	10	10	10	10	10	10	10	10	0	10	9.0
	Zork1	350	30	40	40	54	30	35	44	45	30	59	59	40.7
	Zork3	7	3	4	3	1	1	2	1	3	3	3	4	2.4
	Bal.	51	0	5	5	5	5	5	10	10	5	5	10	5.5
	Tem.	35	5	5	8	5	5	5	5	5	5	5	8	5.3
EvoTest (run 1)	Det.	360	100	100	290	290	290	290	280	280	290	290	290	250.0
	Lib.	30	0	10	10	10	10	10	12	17	19	19	19	11.7
	Zork1	350	40	50	40	50	40	50	50	50	40	50	50	46.0
	Zork3	7	3	2	2	3	2	2	3	3	3	3	3	2.6
	Bal.	51	15	10	15	10	20	20	20	20	20	20	20	17.0
	Tem.	35	8	8	8	8	8	8	8	8	8	8	8	8.0
EvoTest (run 2)	Det.	360	100	160	280	280	280	280	280	280	280	280	280	250.0
	Lib.	30	10	10	10	17	19	19	19	19	19	24	24	16.6
	Zork1	350	50	40	60	54	65	65	65	65	55	59	65	57.8
	Zork3	7	2	0	3	2	2	2	2	2	2	2	3	1.9
	Bal.	51	5	10	15	5	10	20	15	10	15	20	20	12.5
	Tem.	35	8	8	8	8	8	10	10	10	10	15	15	9.5
EvoTest (run 3)	Det.	360	120	300	300	300	300	300	310	300	300	310	310	284.0
	Lib.	30	10	10	10	10	12	19	19	19	19	19	19	14.7
	Zork1	350	50	69	25	40	54	44	55	55	50	25	69	46.7
	Zork3	7	3	3	1	3	3	3	3	2	1	3	3	2.5
	Bal.	51	10	20	25	25	25	25	25	25	25	25	25	23.0
	Tem.	35	5	5	8	8	8	8	8	8	10	10	10	7.8

Table A10: Per-attempt matched comparisons on DeepSeek-v4-Pro and GLM-4.7. One run per condition.

Backbone	Condition	Game	Ref.	1	2	3	4	5	6	7	8	9	10	Best	Mean
DeepSeek-v4-Pro	BISE	Det.	360	120	130	200	130	190	310	310	330	30	330	330	208.0
		Lib.	30	21	21	21	21	21	28	30	30	30	30	30	25.3
		Zork1	350	50	65	50	35	25	29	49	73	73	73	73	52.2
		Zork3	7	3	3	3	3	3	4	4	3	3	0	4	2.9
		Bal.	51	15	15	20	20	20	20	25	30	30	30	30	22.5
		Tem.	35	10	15	15	12	15	15	15	15	15	15	15	14.2
	ReAct	Det.	360	230	100	120	130	340	100	100	130	120	130	340	150.0
		Lib.	30	21	21	21	30	21	29	21	21	21	29	30	23.5
		Zork1	350	74	50	65	70	50	50	54	50	40	55	74	55.8
		Zork3	7	5	1	2	1	3	3	3	4	3	3	5	2.8
		Bal.	51	15	15	15	15	10	0	15	15	20	20	20	14.0
		Tem.	35	10	5	10	15	10	15	5	15	10	10	15	10.5
GLM-4.7	BISE	Det.	360	120	130	320	90	200	310	320	300	30	300	320	212.0
		Lib.	30	21	21	28	30	30	30	30	30	30	30	30	28.0
		Zork1	350	54	45	45	59	45	50	45	59	50	64	64	51.6
		Zork3	7	3	3	3	3	5	3	4	3	3	3	5	3.3
		Bal.	51	10	10	10	10	10	5	10	10	10	10	10	9.5
		Tem.	35	5	8	15	15	15	13	13	15	12	15	15	12.6
	ReAct	Det.	360	30	120	20	20	130	30	90	110	120	20	130	69.0
		Lib.	30	20	21	21	20	20	21	29	21	19	13	29	20.5
		Zork1	350	35	29	50	45	30	55	59	25	55	55	59	43.8
		Zork3	7	3	3	1	0	2	1	4	1	3	2	4	2.0
		Bal.	51	15	10	15	15	15	15	10	15	10	15	15	13.5
		Tem.	35	20	10	15	22	12	10	15	20	15	12	22	15.1