Memory Retention Is Not Enough to Master Memory Tasks in Reinforcement Learning

Extended Abstract

Oleg Shchendrigin
MIRIAI & Innopolis University
Moscow, Russia
shchendrigin.o@miriai.org

Egor Cherepanov
AXXX & MIRIAI
Moscow, Russia
cherepanov.e@miriai.org

Alexey K. Kovalev
AXXX & MIRIAI
Moscow, Russia

Aleksandr I. Panov
AXXX & MIRIAI
Moscow, Russia

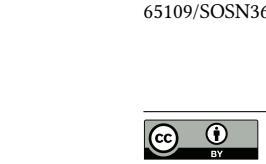
ABSTRACT

Effective decision-making in the real world depends on memory that is both stable and adaptive: environments change over time, and agents must retain relevant information over long horizons while also updating or overwriting outdated content when circumstances shift. Existing Reinforcement Learning (RL) benchmarks and memory-augmented agents focus primarily on retention, leaving the critical ability of memory rewriting largely unexplored. To address this gap, we introduce a benchmark that explicitly tests continual memory updating under partial observability, and use it to compare recurrent, transformer-based, and structured memory architectures. Our experiments reveal that classic recurrent models, despite their simplicity, demonstrate greater flexibility and robustness in memory rewriting tasks than modern structured memories, which succeed only under narrow conditions, and transformer-based agents. These findings expose a fundamental limitation of current approaches and emphasize the necessity of memory mechanisms that balance stable retention with adaptive updating. Our work highlights this overlooked challenge, introduces benchmarks to evaluate it, and offers insights for designing future RL agents with explicit and trainable forgetting mechanisms. Full paper and code: https://quartz-admirer.github.io/Memory-Rewriting/.

KEYWORDS

Reinforcement Learning; POMDP; Memory; Benchmark

ACM Reference Format:
Oleg Shchendrigin, Egor Cherepanov, Alexey K. Kovalev, and Aleksandr I. Panov. 2026. Memory Retention Is Not Enough to Master Memory Tasks in Reinforcement Learning: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/SOSN3643



Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/SOSN3643

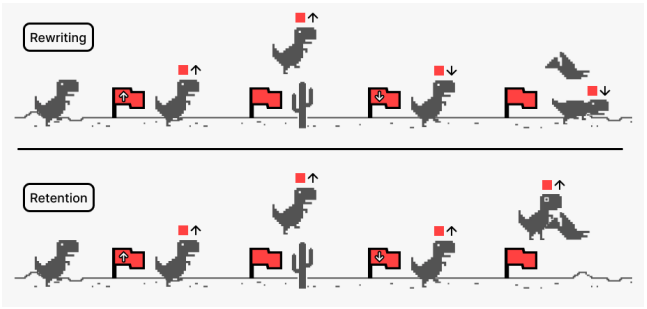


Figure 1: Illustration of memory rewriting vs. retention.

1 INTRODUCTION

Many everyday tasks demand memory that can both preserve and revise information [1, 6, 17]. A pedestrian following a sequence of directional signs must replace an earlier instruction (*"turn left at the park"*) when a new sign ahead indicates a detour, while a warehouse robot tracking object locations must update its internal map each time an item is moved. In both cases, previously stored information becomes misleading unless the agent rewrites its memory with current observations (see Figure 1 for a representative illustration).

RL agents must exhibit the same capacity to revise stored information because observations are often incomplete and the latent state cannot be inferred from a single frame [7, 8, 10, 23]. Under partial observability, effective decision-making requires agents to construct and maintain a memory that summarizes the relevant aspects of past experience [3, 4, 9, 14, 20, 22].

While existing work [2, 5, 11, 15, 16, 19, 21] focuses on retention, we isolate *memory rewriting* in RL through four diagnostic tasks that enforce continuous memory updates under partial observability. **`Endless T-Maze`** consists of sequential corridors where each new cue invalidates the previous one, requiring the agent to actively overwrite outdated instructions. **`Color-Cubes`** (in `Trivial`, `Medium`, and `Extreme` variants) is a grid-world task where colored cubes stochastically teleport, forcing agents to continually update their internal map and avoid acting on stale information.

We evaluate three representative families of memory-augmented RL agents providing a detailed characterization of their respective strengths and limitations. We focus on how each architecture adapts when memory rewriting becomes essential for success, exposing distinct strategies and failure modes across these paradigms.

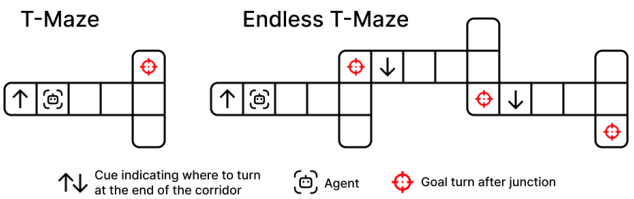


Figure 2: `T-Maze` vs. `Endless T-Maze`.

Our contributions are threefold:

(1) **Benchmarks for rewriting.** We introduce `Endless T-Maze` and `Color-Cubes`, two families of environments, which provide four options for isolate the ability to perform continual, selective memory updates beyond simple cue retention.
(2) **Systematic evaluation.** We compare recurrent [18], transformer [14], and structured-memory [9, 12] agents, identifying when rewriting mechanisms succeed and how performance degrades with increasing memory update frequency.
(3) **Design principles.** We analyze architectural factors linked to rewriting competence, highlighting the effectiveness of *explicit, adaptive forgetting* (e.g., learnable forget gates).

2 BENCHMARKING MEMORY REWRITING

Endless T-Maze. This environment extends the classic T-Maze [13] into an infinite sequence of interconnected corridors, forming a continual version of the cue-based navigation task. At the start of each corridor, the agent receives a binary cue indicating whether to turn left or right at the upcoming junction. Once a turn is made, the cue changes, invalidating all previous information – thereby requiring continual memory overwriting rather than static retention.

Color-Cubes. `Color-Cubes` is a stress test for adaptive memory, designed to evaluate whether an agent can store, update, and selectively forget internal representations under partial observability. The environment consists of a $G \times G$ grid with N uniquely colored cubes. At the beginning of each phase, the agent observes the colors and positions of all cubes and receives a target color. This information is then hidden, forcing the agent to rely on memory.

The agent must navigate to the cube with the target color and execute an interaction based on the remembered color, then this cube is teleported and another color becomes the target. Episodes consist of multiple phases with changing targets. During movement, non-target cubes may stochastically teleport, making previously stored information outdated. The agent must detect inconsistencies and rewrite its internal map to maintain an accurate belief about the environment. It evaluates selective adaptive forgetting: information that was previously irrelevant can suddenly become critical, requiring the agent to discard obsolete memories while retaining only task-relevant content.

We define three difficulty levels that progressively increase the demands on memory rewriting: **`Trivial`** – Single cube and single target (N=K=1); **`Medium`** – Multiple cubes with complete state updates (positions and colors); **`Extreme`** – Multiple cubes with *incomplete* teleportation updates (positions only, colors hidden).

3 RESULTS

The results establish a consistent performance hierarchy among the tested architectures. Across memory rewriting tasks the observed ranking was: PPO-LSTM [18], FFM [12], GTrXL [14], SHM [9], and MLP. This ranking correlates with the architectural approach each model takes to managing outdated information – specifically, with its mechanism for forgetting.

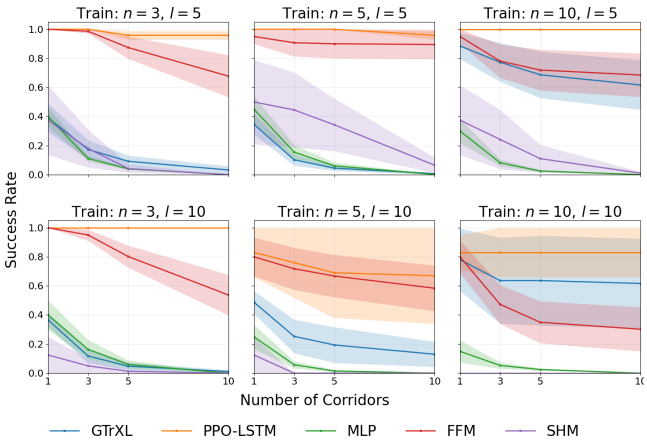


Figure 3: Baseline comparison on `Endless T-Maze` under validation interpolation and extrapolation conditions, where corridor lengths and fixed sampling are the same, and the parameter of the number of corridors varies. The result for Success Rate is mean±sem.

An analysis of the architectures reveals a trend:

- **PPO-LSTM**, the top-performing model, utilizes a dedicated forget gate with learnable parameters.
- **FFM** architecture implements a more defined, rule-based forgetting process where older information is systematically replaced.
- **GTrXL** show generalization in some scenarios. The stabilization of architecture learning occurs with the help of gating, which once again emphasizes their necessity.
- **SHM** has a similar concept of matrix memory structuring to FFM, but lacks an explicit forgetting mechanism.

A comparison in an ablation study with PPO-RNN confirmed that PPO-LSTM gates contributed to its success, while a comparison of its gates with PPO-GRU gates highlighted the need for an adaptive forgetting mechanism.

4 CONCLUSION

Our study reveals that memory retention alone is insufficient for solving RL tasks that require continual adaptation under partial observability. Through the proposed benchmarks, we isolate and evaluate the ability of RL agents to rewrite memory – to selectively discard obsolete information and integrate new evidence as environments evolve. We observe a consistent hierarchy of adaptive competence: agents with explicit, learnable forgetting mechanisms exhibit strong performance and generalization when rewriting is essential rather than optional. These findings suggest that the core challenge in memory-intensive decision-making is not the preservation of information but its controlled transformation over time.

ACKNOWLEDGMENTS

The study was supported by the Ministry of Economic Development of the Russian Federation (agreement No. 139-15-2025-013, dated June 20, 2025, IGK 000000C313925P4B0002).

REFERENCES

[1] Zachary H Bretton, Hyojeong Kim, Marie T Banich, and Jarrod A Lewis-Peacock. 2024. Suppressing the maintenance of information in working memory alters long-term memory traces. *Journal of Cognitive Neuroscience* 36, 10 (2024), 2117–2136.

[2] Egor Cherepanov, Nikita Kachaev, Alexey K Kovalev, and Aleksandr I Panov. 2025. Memory, benchmark & robots: A benchmark for solving complex tasks with reinforcement learning. *arXiv preprint arXiv:2502.10550* (2025).

[3] Egor Cherepanov, Alexey Kovalev, and Aleksandr Panov. 2025. ELMUR: External Layer Memory with Update/Rewrite for Long-Horizon RL. In *CoRL 2025 Workshop RembeBeRL*. https://openreview.net/forum?id=H2dvLYqlaa

[4] Egor Cherepanov, Alexey Staroverov, Dmitry Yudin, Alexey K Kovalev, and Aleksandr I Panov. 2023. Recurrent action transformer with memory. *arXiv preprint arXiv:2306.09459* (2023).

[5] Maxime Chevalier-Boisvert, Bolun Dai, Mark Towers, Rodrigo de Lazcano, Lucas Willems, Salem Lahlou, Suman Pal, Pablo Samuel Castro, and Jordan Terry. 2023. Minigrid & Miniworld: Modular & Customizable Reinforcement Learning Environments for Goal-Oriented Tasks. arXiv:2306.13831 [cs.LG] https://arxiv.org/abs/2306.13831

[6] Seth R Koslov, Arjun Mukerji, Katlyn R Hedgpeth, and Jarrod A Lewis-Peacock. 2019. Cognitive flexibility improves memory for delayed intentions. *ENeuro* 6, 6 (2019).

[7] Hanna Kurniawati. 2022. Partially Observable Markov Decision Processes and Robotics. *Annu. Rev. Control. Robotics Auton. Syst.* 5 (2022), 253–277. https://api.semanticscholar.org/CorpusID:245789500

[8] Mikko Lauri, David Hsu, and Joni Pajarinen. 2022. Partially observable markov decision processes in robotics: A survey. *IEEE Transactions on Robotics* 39, 1 (2022), 21–40.

[9] Hung Le, Kien Do, Dung Nguyen, Sunil Gupta, and Svetha Venkatesh. 2024. Stable Hadamard Memory: Revitalizing Memory-Augmented Agents for Reinforcement Learning. arXiv:2410.10132 [cs.LG] https://arxiv.org/abs/2410.10132

[10] Lingheng Meng, Rob Gorbet, and Dana Kulić. 2021. Memory-based deep reinforcement learning for pomdps. In *2021 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 5619–5626.

[11] Steven Morad, Ryan Kortvelesy, Matteo Bettini, Stephan Liwicki, and Amanda Prorok. 2023. POPGym: Benchmarking Partially Observable Reinforcement Learning. arXiv:2303.01859 [cs.LG] https://arxiv.org/abs/2303.01859

[12] Steven Morad, Ryan Kortvelesy, Stephan Liwicki, and Amanda Prorok. 2023. Reinforcement Learning with Fast and Forgetful Memory. arXiv:2310.04128 [cs.LG] https://arxiv.org/abs/2310.04128

[13] Tianwei Ni, Michel Ma, Benjamin Eysenbach, and Pierre-Luc Bacon. 2023. When do transformers shine in rl? decoupling memory from credit assignment. *Advances in Neural Information Processing Systems* 36 (2023), 50429–50452.

[14] Emilio Parisotto, Francis Song, Jack Rae, Razvan Pascanu, Caglar Gulcehre, Siddhant Jayakumar, Max Jaderberg, Raphael Lopez Kaufman, Aidan Clark, Seb Noury, et al. 2020. Stabilizing transformers for reinforcement learning. In *International conference on machine learning*. PMLR, 7487–7498.

[15] Jurgis Pasukonis, Timothy Lillicrap, and Danijar Hafner. 2022. Evaluating Long-Term Memory in 3D Mazes. arXiv:2210.13383 [cs.AI] https://arxiv.org/abs/2210.13383

[16] Marco Pleines, Matthias Pallasch, Frank Zimmer, and Mike Preuss. 2025. Memory Gym: Towards Endless Tasks to Benchmark Memory Capabilities of Agents. *Journal of Machine Learning Research* 26, 6 (2025), 1–40.

[17] John J Sakon and Roozbeh Kiani. 2022. Differences in memory for what, where, and when components of recently formed episodes. *Journal of neurophysiology* 128, 2 (2022), 310–325.

[18] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. arXiv:1707.06347 [cs.LG] https://arxiv.org/abs/1707.06347

[19] Artyom Sorokin, Nazar Buzun, Leonid Pugachev, and Mikhail Burtsev. 2022. Explain My Surprise: Learning Efficient Long-Term Memory by Predicting Uncertain Outcomes. (07 2022). https://doi.org/10.48550/arXiv.2207.13649

[20] Ruo Yu Tao, Kaicheng Guo, Cameron S. Allen, and George Dimitri Konidaris. 2025. Benchmarking Partial Observability in Reinforcement Learning with a Suite of Memory-Improvable Domains. *ArXiv* abs/2508.00046 (2025). https://api.semanticscholar.org/CorpusID:280297434

[21] Zekang Wang, Zhe He, Borong Zhang, Edan Toledo, and Steven Morad. 2025. POPGym Arcade: Parallel Pixelated POMDPs. arXiv:2503.01450 [cs.LG] https://arxiv.org/abs/2503.01450

[22] Daniil Zelezetsky, Egor Cherepanov, Alexey K Kovalev, and Aleksandr I Panov. 2025. Re: Frame–Retrieving Experience From Associative Memory. *arXiv preprint arXiv:2508.19344* (2025).

[23] Haoxu Zhang, Parham M Kebria, Shady Mohamed, Samson Yu, and Saeid Nahavandi. 2023. A Review on Robot Manipulation Methods in Human-Robot Interactions. *arXiv preprint arXiv:2309.04687* (2023).