Unsupervised Hebbian Learning Drives Biologically Interpretable Pattern Separation in a Hippocampal–Striatal Network

Jiachuan Wang[1], **Vachan Jagadeesh Shetru**[1], **M Ganesh Kumar**[2], **Camilo Libedinsky**[1], **Shih-Cheng Yen**[1], **Andrew Yong-Yi Tan**[1], **Jai S Polepalli**[1]

[1]National University of Singapore, Singapore
[2]Max Planck Institute for Biological Cybernetics, Germany
jiachuan.wang@u.nus.edu, vachan.sj@gmail.com, mganeshkumar@seas.harvard.edu, camilo@nus.edu.sg, shihcheng@nus.edu.sg, phstya@nus.edu.sg, jpolepalli@nus.edu.sg

Abstract

Pattern separation, essential for encoding distinct memories of overlapping contexts, relies on dentate gyrus coding, which is shaped by entorhinal input and strong lateral inhibition. The pattern-separated state space provided by the hippocampus is thought to facilitate striatal-dependent reinforcement learning, enabling associations between sensory features and outcomes. Although synaptic plasticity, value prediction error modulation, and adult neurogenesis have been implicated in this process, their precise contributions remain unclear. To investigate the computational mechanisms underlying pattern separation, we developed neural network models incorporating an entorhinal cortex–dentate gyrus–striatal circuit. Simulations suggest that lateral inhibition is necessary for forming a decorrelated coding subspace, whereas hippocampal plasticity and dopamine modulation are not required for value learning. These findings dissociate neural pattern separation in hidden-layer representations from behavioral discrimination at the model output, highlighting how biologically grounded architectures and learning rules can enhance interpretability.

Introduction

Pattern separation, a memory process that enables animals to distinguish between similar environments, is essential for adapting to new conditions. Imaging and behavioral studies highlight the dentate gyrus (DG) as a key locus of this process, where dissimilarities in DG activity and behavioral patterns increase with training (Cholvin and Bartos 2022; Sahay et al. 2011; Yassa and Stark 2011). Since the Marr–Albus theory, computational mechanisms underlying hippocampal pattern separation have been framed in terms of the dimensionality of the coding subspace, shaped by divergent feedforward projections from the entorhinal cortex (EC) and lateral inhibition within the DG (Cayco-Gajic and Silver 2019). However, activity-dependent processes such as adult neurogenesis and synaptic plasticity have also been implicated, though their precise roles remain unclear. Notably, conflicting evidence suggests that neurogenesis can both enhance and impair hippocampus-dependent learning and memory, including pattern separation (Evans et al. 2022; Sahay et al. 2011). The hippocampus also projects to the ventral striatum, where environmental value is encoded, and influences midbrain dopamine neurons that compute value prediction errors (VPEs) (Schultz, Dayan, and Montague 1997; van der Meer and Redish 2011). These neurons project back to the DG, where dopamine-modulated plasticity has been proposed to contribute to pattern separation (Hamilton et al. 2010). To investigate how synaptic plasticity, lateral inhibition, dopamine modulation, and neurogenesis shape DG pattern separation and value estimation, we developed computational models of the EC–DG–striatal circuit implementing different learning rules.

Methods

We modeled the classical contextual fear discrimination task used in biological studies (McHugh et al. 2007; Sahay et al. 2011). Agents underwent 14 days of pattern separation training with daily 3-minute exposures to chambers A and B, which differed slightly in sensory cues. A 2-second foot shock was delivered only in chamber A. EC activity in each chamber was represented as a noisy version of two vectors drawn from a normal distribution, with cosine similarity controlled by a similarity index (SI). DG neurons were modeled as rate-based leaky units receiving EC input. Feedforward weights were updated by one of several learning rules: Hebbian ($\Delta W = \alpha_{EC\to DG}(yx^T - W)$), three-factor ($\Delta W = \alpha_{EC\to DG}\delta yx^T$), backpropagation ($\Delta W = \alpha_{EC\to DG}[(W^2)^T\delta \odot f^{'}(u) - \lambda f^{'}(u)]x^T$), or direct feedback alignment (DFA; W^2 replaced by a random matrix); in the fixed-weight condition (Kumar et al. 2022), no learning occurred (Fig. 1A). In these conditions, DG neurons received no lateral inhibitory input.

In the Hebbian/anti-Hebbian network (Qin et al. 2023), DG neurons also received lateral inhibitory input from other DG neurons, with inhibitory weights updated by a Hebbian rule ($\Delta M = \alpha_{lat}(yy^T - M)$) (Fig. 1A). The ratio $r_\alpha = \frac{\alpha_{lat}}{\alpha_{EC\to DG}}$ defined the sparsity constraint (Wang et al. 2025). Additive neurogenesis was modeled by incrementally activating new DG neurons each day.

Results

We first evaluated the network's baseline performance in the behavioral task. As EC input similarity decreased under noise, the Hebbian/anti-Hebbian network progressively

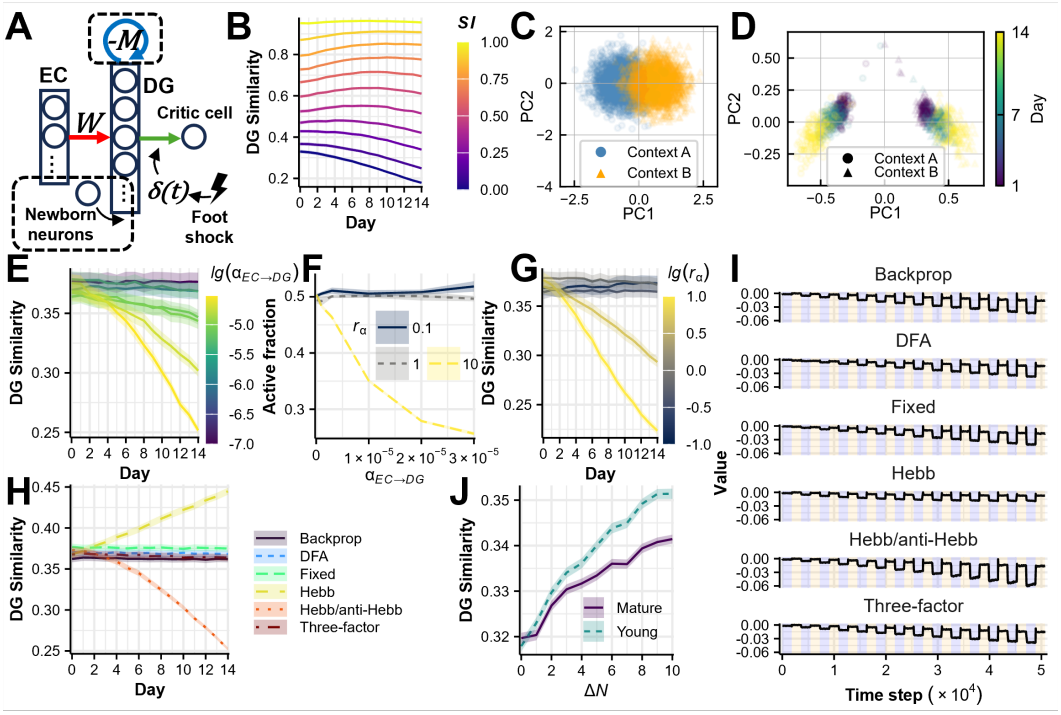

Figure 1: Biologically grounded EC–DG–striatal network model and simulation results.

reduced DG activity similarity, indicating emergent pattern separation (Fig. 1B). With overlapping EC inputs representing contexts A and B (Fig. 1C), the DG produced low-dimensional representations dominated by the first principal component, along which contextual representations became linearly separable over training (Fig. 1D).

To examine the computational mechanisms supporting pattern separation, we systematically varied key network parameters. Increasing $\alpha_{EC\to DG}$, representing stronger EC→DG synaptic plasticity, enhanced pattern separation (Fig. 1E). Higher r_α values promoted sparser coding (Fig. 1F) and improved pattern separation (Fig. 1G), consistent with prior models embedding lateral inhibition in feedforward updates (Sanger 1989).

Alternative learning rules were then compared. Networks without EC→DG plasticity or using VPE modulation (backpropagation, DFA, or three-factor rule) failed to increase DG dissimilarity (Fig. 1H), possibly due to the limited punishment events. Despite qualitatively similar value representations across models, only the Hebbian/anti-Hebbian network achieved stronger contextual discrimination and reproduced experimental findings where both DG activity and behavioral dissimilarities increased with training (Fig. 1I).

Finally, we tested DG neurogenesis. In biological systems, young adult-born DG neurons exhibit stronger EC→DG plasticity (higher $\alpha_{EC\to DG}$) and weaker lateral inhibition (lower r_α) (Marín-Burgin et al. 2012; Schmidt-Hieber, Jonas, and Bischofberger 2004). Increasing daily neuron addition (ΔN) led to impaired pattern separation, regardless of their maturation stage (Fig. 1J).

Discussion

We developed a biologically grounded model incorporating EC→DG synaptic plasticity, lateral inhibition, and DG neurogenesis to investigate hippocampal pattern separation. The model reproduced key aspects of pattern separation in DG coding and value representations. Disrupting EC→DG plasticity, lateral inhibition, or increasing neurogenesis each impaired pattern separation, indicating their individually sufficient yet converging contributions.

Only the Hebbian/anti-Hebbian network, mimicking DG lateral inhibition, reproduced experimentally observed pattern separation and enhanced value discrimination. Unsupervised Hebbian learning self-organized network components to produce decorrelated representations, whereas backpropagation failed to increase hidden-layer dissimilarity, implying that canonical reinforcement learning models may miss parts of the data-compatible solution space (Kumar et al. 2025; Levenstein et al. 2024; Laskin, Srinivas, and Abbeel 2020).

Although hippocampal synaptic plasticity is considered a mechanistic substrate of learning and memory, it alone is not sufficient for expressing learned behavior. For memories to be expressed as behavior, hippocampal activity must engage downstream circuits such as the prefrontal cortex, striatum, and amygdala (Frankland and Bontempi 2005). The hippocampus can exhibit strong plasticity or replay without overt behavioral change (Carr, Jadhav, and Frank 2011; Morris, Davis, and Butcher 1990). In sum, local hippocampal plasticity can support distinct neural representations, whereas behavioral expression depends on how these patterns are integrated and reactivated across broader brain net-

works.

Acknowledgments

This research was supported by the Healthy Longevity Translational Research Programme Pilot Studies Grant, NUS Artificial Intelligence Institute Seed Funding, and Singapore Ministry of Education Academic Research Fund Tier 2 (MOE-T2EP30122-0016).

References

Carr, M. F.; Jadhav, S. P.; and Frank, L. M. 2011. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience*, 14(2): 147–153.

Cayco-Gajic, N. A.; and Silver, R. A. 2019. Re-evaluating Circuit Mechanisms Underlying Pattern Separation. *Neuron*, 101(4): 584–602.

Cholvin, T.; and Bartos, M. 2022. Hemisphere-specific spatial representation by hippocampal granule cells. *Nature Communications*, 13(1): 6227.

Evans, A.; Terstege, D. J.; Scott, G. A.; Tsutsui, M.; and Epp, J. R. 2022. Neurogenesis mediated plasticity is associated with reduced neuronal activity in CA1 during context fear memory retrieval. *Scientific Reports*, 12(1): 7016.

Frankland, P. W.; and Bontempi, B. 2005. The organization of recent and remote memories. *Nature Reviews Neuroscience*, 6(2): 119–130.

Hamilton, T. J.; Wheatley, B. M.; Sinclair, D. B.; Bachmann, M.; Larkum, M. E.; and Colmers, W. F. 2010. Dopamine modulates synaptic plasticity in dendrites of rat and human dentate granule cells. *Proceedings of the National Academy of Sciences*, 107(42): 18185–18190.

Kumar, M. G.; Bordelon, B.; Zavatone-Veth, J. A.; and Pehlevan, C. 2025. A Model of Place Field Reorganization During Reward Maximization. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 31892–31929. PMLR.

Kumar, M. G.; Tan, C.; Libedinsky, C.; Yen, S.-C.; and Tan, A. Y. Y. 2022. A nonlinear hidden layer enables actor–critic agents to learn multiple paired association navigation. *Cerebral Cortex*, 32(18): 3917–3936.

Laskin, M.; Srinivas, A.; and Abbeel, P. 2020. CURL: Contrastive Unsupervised Representations for Reinforcement Learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 5639–5650. PMLR.

Levenstein, D.; Efremov, A.; Eyono, R. H.; Peyrache, A.; and Richards, B. 2024. Sequential predictive learning is a unifying theory for hippocampal representation and replay. *bioRxiv*.

Marín-Burgin, A.; Mongiat, L. A.; Pardi, M. B.; and Schinder, A. F. 2012. Unique Processing During a Period of High Excitation/Inhibition Balance in Adult-Born Neurons. *Science*, 335(6073): 1238–1242.

McHugh, T. J.; Jones, M. W.; Quinn, J. J.; Balthasar, N.; Coppari, R.; Elmquist, J. K.; Lowell, B. B.; Fanselow, M. S.; Wilson, M. A.; and Tonegawa, S. 2007. Dentate Gyrus NMDA Receptors Mediate Rapid Pattern Separation in the Hippocampal Network. *Science*, 317(5834): 94–99.

Morris, R. G. M.; Davis, S.; and Butcher, S. P. 1990. Hippocampal synaptic plasticity and NMDA receptors: a role in information storage? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 329(1253): 187–204.

Qin, S.; Farashahi, S.; Lipshutz, D.; Sengupta, A. M.; Chklovskii, D. B.; and Pehlevan, C. 2023. Coordinated drift of receptive fields in Hebbian/anti-Hebbian network models during noisy representation learning. *Nature Neuroscience*, 26(2): 339–349.

Sahay, A.; Scobie, K. N.; Hill, A. S.; O'Carroll, C. M.; Kheirbek, M. A.; Burghardt, N. S.; Fenton, A. A.; Dranovsky, A.; and Hen, R. 2011. Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature*, 472(7344): 466–470.

Sanger, T. D. 1989. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2(6): 459–473.

Schmidt-Hieber, C.; Jonas, P.; and Bischofberger, J. 2004. Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. *Nature*, 429(6988): 184–187.

Schultz, W.; Dayan, P.; and Montague, P. R. 1997. A Neural Substrate of Prediction and Reward. *Science*, 275(5306): 1593–1599.

van der Meer, M. A. A.; and Redish, A. D. 2011. Theta Phase Precession in Rat Ventral Striatum Links Place and Reward Information. *The Journal of Neuroscience*, 31(8): 2843–2854.

Wang, J.; Shetru, V. J.; Kumar, M. G.; Libedinsky, C.; Yen, S.-C.; Tan, A. Y.-Y.; and Polepalli, J. S. 2025. A biologically plausible computational model of hippocampal neurogenesis and pattern separation in memory. In *Cognitive Computational Neuroscience*.

Yassa, M. A.; and Stark, C. E. 2011. Pattern separation in the hippocampus. *Trends in Neurosciences*, 34(10): 515–525.