Preserving TD-Driven Place-Field Reorganization at Scale: A Scaling-Limit Perspective
M Ganesh Kumar, Cengiz Pehlevan (Harvard University) COSYNE 2026

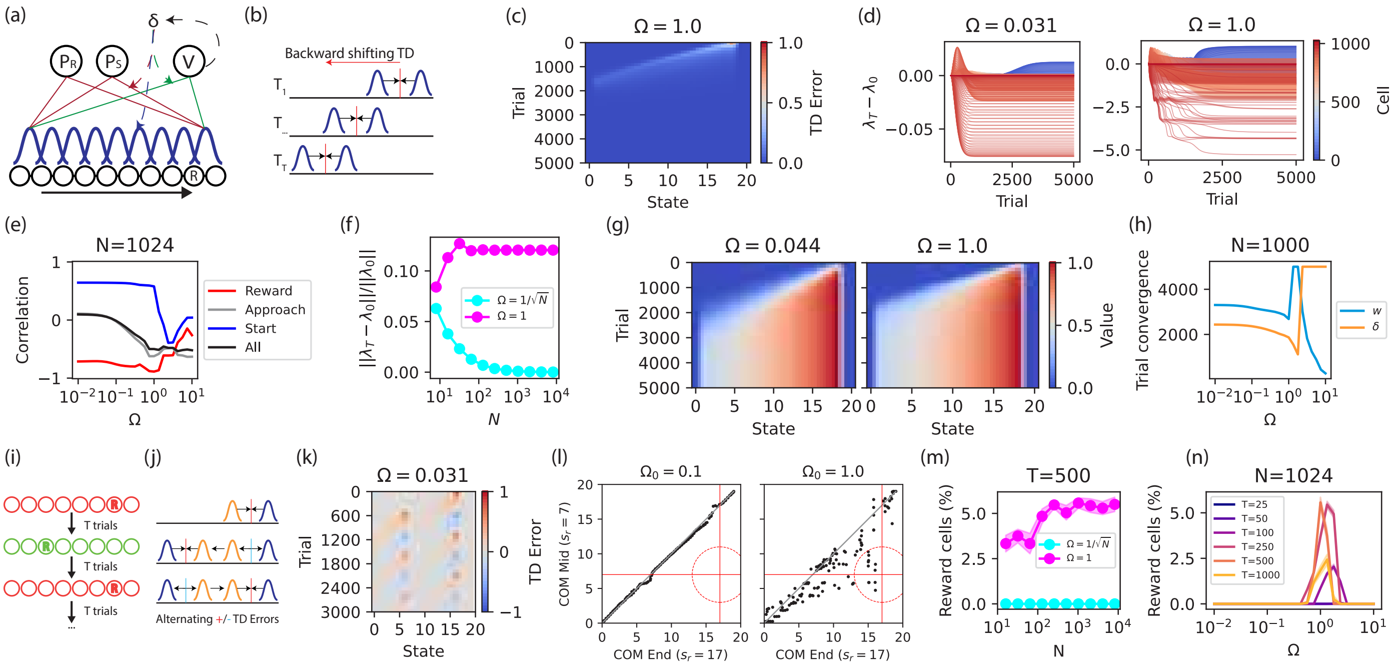
Abstract: A desired property when building network models to understand neural phenomena is that observed behaviors persist as network size grows to experimentally relevant scales. Is this requirement met by models of learning-dependent neural reorganization? We address this question by integrating scaling-limit theory from deep learning with a model of hippocampal place-field reorganization based on temporal-difference (TD) learning—a canonical mechanism of reinforcement learning in the brain. We analyze how representation updates scale with network size (N) under two parameterizations: Neural Tangent Parameterization (NTP) and Maximal Update Parameterization (μP). NTP drives representation updates toward zero as N grows, causing learned phenomena to vanish at scale, while μP maintains updates at $\mathcal{O}(1)$, thereby preserving phenomena as networks grow. We use a TD learning model of hippocampal place fields as a testbed, where place fields encode spatial representations and are modeled as radial basis function units. These networks are trained using the reinforcement learning framework where the TD error updates both the synaptic weights and the center of mass (COM) of place fields, enabling spatial reorganization. We find that under the "lazy" NTP setting, hallmark phenomena e.g. (1) reward coding place fields shifting backward and (2) emergence of reward-coding cell subpopulations during alternating rewards—disappear as N increases. Place field shifts depend on backward-propagating TD error, with positive and negative TD errors respectively attracting or repelling place fields. Maintaining representation learning with μP not only preserves these effects across scales, but also leads to faster value convergence, indicating improved credit assignment and model flexibility. Adopting a scaling-limit perspective, we show that persistence of TD-driven neural reorganization in large-scale models depends critically on principled parameterization choices. Our findings provide a framework for constructing neural network models that capture key, experimentally observed hippocampal phenomena at scale, and help distinguish robust effects from artifacts of restricted model size.

Central Question: A fundamental goal in neuroscience is to understand how neuronal representations reorganize to support adaptive behavior, and how computational models can accurately capture this process at biologically realistic scales. Many neural network models provide insight into representation learning, but commonly used parameterizations (rules for scaling learning rates and outputs with network size) may fail to preserve key neural phenomena as models approach large network limits [1]. This raises the question: How does the choice of parameterization affect a model's ability to replicate both the representational and behavioral signatures of hippocampal learning observed in experiments? Specifically, can an appropriate parameterization robustly account for hallmark effects such as (1) the backward shifting of place fields from reward to start and (2) the emergence of place cell subpopulations that dynamically track alternating reward positions, even as network size is increased to biologically relevant scales?

Approach: Agents comprised of N Gaussian place cells learn either value estimation or a navigation policy using the actor-critic framework (Fig. **a**) [1]. Each place cell i produces activation $\phi_i(s;\lambda_i) = \exp[-(s-\lambda_i)^2/(2\sigma^2)]$ centered at λ_i, forming an activity vector $\phi(s)$. The outputs for value and action are:

$$v(s) = \frac{1}{\Omega N}\sum_{i=1}^{N} w_i\phi_i(s,\lambda_i), \quad a_j(s) = \frac{1}{\Omega N}\sum_{i=1}^{N} W_{ij}\phi_i(s,\lambda_i), \quad \eta = \eta_0\Omega^2 N, \tag{1}$$

where η is the learning rate scaled using a base learning rate $\eta_0 = 0.1$, the number of place fields N and the richness hyperparameterΩ which smoothly controls the transition between the NTP ($\Omega = 1/\sqrt{N}$) and μP ($\Omega = 1$) parameterizations [2]. NTP causes representation learning to diminish in $\mathcal{O}(1/\sqrt{N})$ as the number of neurons N is increased while μP maintains representation learning of $\mathcal{O}(1)$ independent of network scale [3,4]. Place field center of masses (COMs) λ_i are optimized using gradients of the TD error $\delta_t = r(s_t, a_t) + \gamma v(s_{t+1}) - v(s_t)$ with $\gamma = 0.95$. This translates to each fields COM shifting according to $\Delta\lambda_t = \frac{1}{\Omega N}\delta_t w_t \phi(s)(\frac{s-\lambda_t}{\sigma})$. Thus, fields will shift towards states with a positive δ_t (Fig. **b**), and shift away from states with a negative δ_t (Fig. **j**). Agents were evaluated on two tasks: (1) learning value estimation in a 20-state linear track (states 1=start, 18=reward) to probe backward shifting using an optimal policy π^*, and (2) learning to navigate in a 20-state circular track with reward alternating between two states 7 and 17 every T trials to probe the emergence of reward-coding subpopulations. By systematically varying the network parameterization using Ω and N, we examine the conditions at which spatial reorganization persists or vanishes at large scale. Deterministic value-estimation agents were run with one seed, and stochastic actor–critic agents with five seeds.



Results: Across both the NTP and μP parameterizations, TD error propagates backward from the reward to the start (Fig. c), producing a systematic backward shift in the centers of mass (COMs) of place fields. This effect is pronounced under μP, which shows significantly larger and more sustained backward shifts (Fig. d, with negative values indicating shift direction), closely matching experimental data [5]. "Reward", "Approach", and "Start" cells denote place fields near the reward, between start and reward, and near the start, respectively. The correlation between COM position and trial number remains largely negative for reward (red) and start (blue) cells across all Ω values (Fig. e). Increasing Ω makes the correlation for approach cells (gray) strongly negative, producing pronounced backward shifts in all cell types (black), and demonstrating that this effect is specific to the "rich" (μP) regime. The magnitude of representation updates for COMs decreases with N in the NTP regime but is maintained in the μP regime (Fig. f). Notably, backward shifting also accelerates value convergence: the number of trials required for both value and TD error to stabilize is reduced under μP. When $\Omega > 2$, however, representations and value learning collapse (Fig. e, h).

When the reward state alternates, TD error oscillates between positive and negative (Fig. k), and place fields are respectively attracted or repelled (Fig. j). In the NTP regime, these shifts are minimal; in contrast, the μP regime induces substantial shifts, causing place cells coding for state 7 to remap to state 17, and vice versa (Fig. i; COMs within red circles code for both reward states). Approximately 5% of place cells shift between reward locations for $N \in \{256, 8192\}$ (Fig. m), closely matching the 4.2% observed experimentally [6]. This dedicated population of reward-coding cells emerges only for intermediate Ω $(0.6 < \Omega < 1.8)$ across various trial intervals T (Fig. h). Shaded areas indicate standard error over five seeds.

Conclusion: Our results show that TD error–driven place field reorganization with the μP parameterization provides a scalable and unified mechanism for both reward-predictive and reward-coding dynamics in large networks. We demonstrate that (1) robust spatial reorganization across hundreds to thousands of place cells, (2) preserved representation learning enables faster behavioral convergence through improved credit assignment, and (3) reward-coding cell subpopulations emerge flexibly in response to TD error dynamics, rather than specific cell identities (i.e. place versus reward cells). These findings underscore the necessity of principled parameterization for neural network models to robustly capture key, experimentally observed phenomena as network size increases. Limitations include studying other neural phenomena and network models. Still, this approach can be generalized to other neural models, such as recurrent networks, and informs future work on the connections between representational learning, neural adaptation, and neural dysfunction in high-dimensional, biologically relevant contexts.

References: [1] Kumar et al. 2025 [2] Graldi et al. 2025 [3] Yang & Hu 2020 [4] Bordelon & Pehlevan 2022 [5] Yaghoubi et al. 2026 [6] Gauthier et al. 2018