A Model of Place Field Reorganization During Learning
M Ganesh Kumar, Cengiz Pehlevan (Harvard University) COSYNE 2025

When rodents learn to navigate in a new environment, high place field density emerge at reward locations, fields elongate against the trajectory, and individual fields change spatial selectivity or drift while demonstrating stable behavior. Why place fields demonstrate these characteristic phenomena during learning remains elusive. We develop a model that optimizes for a normative goal, which is to maximize cumulative rewards. Place fields are modelled using Gaussian basis functions to represent spatial information of an environment, and directly synapse to an actor-critic for policy learning in a 1D and 2D environment. Each field's amplitude, center and width, as well as actor weights, are updated online to maximize cumulative reward while the critic minimizes the TD error. We demonstrate that place fields near the target increase their firing rates and move closer to the target during learning. We analytically argue that each place field's dynamics is modulated by the value of a location. Next, we show that both our normative model and fields trained to learn the successor representation increase in size and the center of mass shifts backwards towards the start location. Interestingly, each model's spatial representation evolves differently during early learning stages before becoming aligned. Furthermore, we show that within a certain noise regime, the population vector correlation decreases while the representation similarity remains fairly stable, and fields that are important for stable navigation performance drift less. Finally, we show that incorporating these place field phenomena improves the speed of policy convergence when learning to navigate to a single target, and when the target is shifted to a new location, suggesting a functional role for place field reorganization and noise in continual learning. To conclude, we develop a normative model that recapitulates three aspects of place field learning dynamics and unifies mechanisms to offer testable predictions for future experiments.

Central Question: A place field is canonically described as a localized region in an environment where the firing rate of a hippocampal neuron is maximal and robust across trials. Ablation studies show that the hippocampal representation is useful for learning to navigate to new targets. Importantly, each field's spatial selectivity evolves with experience in a new environment before stabilizing in the later stages of learning. Specifically, a high density of place fields emerge at reward locations (Lee et al. 2020, Cell), place fields elongate backward against the trajectory (Mehta et al. 1997, PNAS), and individual place field's spatial selectivity continues to change or "drift" even when animals demonstrate stable behavior (Qin et al. 2023, Nat. Neuro.). Although disparate mechanisms have been proposed to model these phenomena, a framework that can unify their phenomena and clarify their computational role remains elusive.

Approach: The task is to navigate from a start location (Fig. 1A, green dash) to a target (red area) to receive rewards in either a 1D track or 2D arena with an obstacle. We develop a spatial representation learning model (Fig. A) whose normative goal is to maximize the cumulative discounted reward, described as the Temporal Difference (TD) learning objective $\mathcal{J} = \mathbb{E}[\delta_t] = \mathbb{E}[r_{t+1} + \gamma v_{t+1} - v_t]$. The model has two-layers, with the first containing a population of Gaussian radial basis functions that transform continuous spatial information x_t into a relevant representational substrate $\phi(x_t)$,

$$\phi_i(x_t) = \alpha_i^2 \exp\left(-\frac{||x_t - \lambda_i||_2^2}{2\sigma_i^2}\right) \quad , \quad \nabla\mathcal{J}(W^\pi, w^v, \theta) = \mathbb{E}\left[\sum_t^T \nabla \log \pi(g_t|x_t; W^\pi, \theta)\delta_t + \nabla v(x_t; w^v, \theta)\delta_t\right].$$

which synapse to an actor network in the second layer that learns a navigational policy π to maximize cumulative discounted returns, and a critic network that estimates the value function v_t to compute the TD objective. Besides the actor W^π and critic w^v weights, each place field's firing rate α_i, center of mass λ_i and width σ_i, collectively as θ are optimized using the TD objective. To induce drift, we introduce Gaussian noise when updating field parameters $\theta_{t+1} \leftarrow \theta_t + \xi_t$.

Results: In early learning stages (Fig. 1B, $T = 1000$), our Reward Maximization (RM) agent spends a higher proportion of time ($p(x)$, black) at the start location, while a high field density ($d(x)$, red) emerges at the reward location as individual fields ($\phi(x)$, colored) near the target move closer and increase in firing rate. Hence, the correlation between the field density and proportion of time spent in a location becomes anti-correlated (Fig. 1C, orange). As learning progresses, the agent spends a higher proportion of time at the target while the field density increases slightly at the start location. Other fields along the track elongate and shift backwards (Fig. 1B, colored; Fig. 1D, orange), causing the correlation between the field density and time spent at a location to become positive (Fig. 1C, orange). Our baseline Successor Representation (SR) agent demonstrates similar field elongation and backward shift (Fig. D, blue). However, the field density is

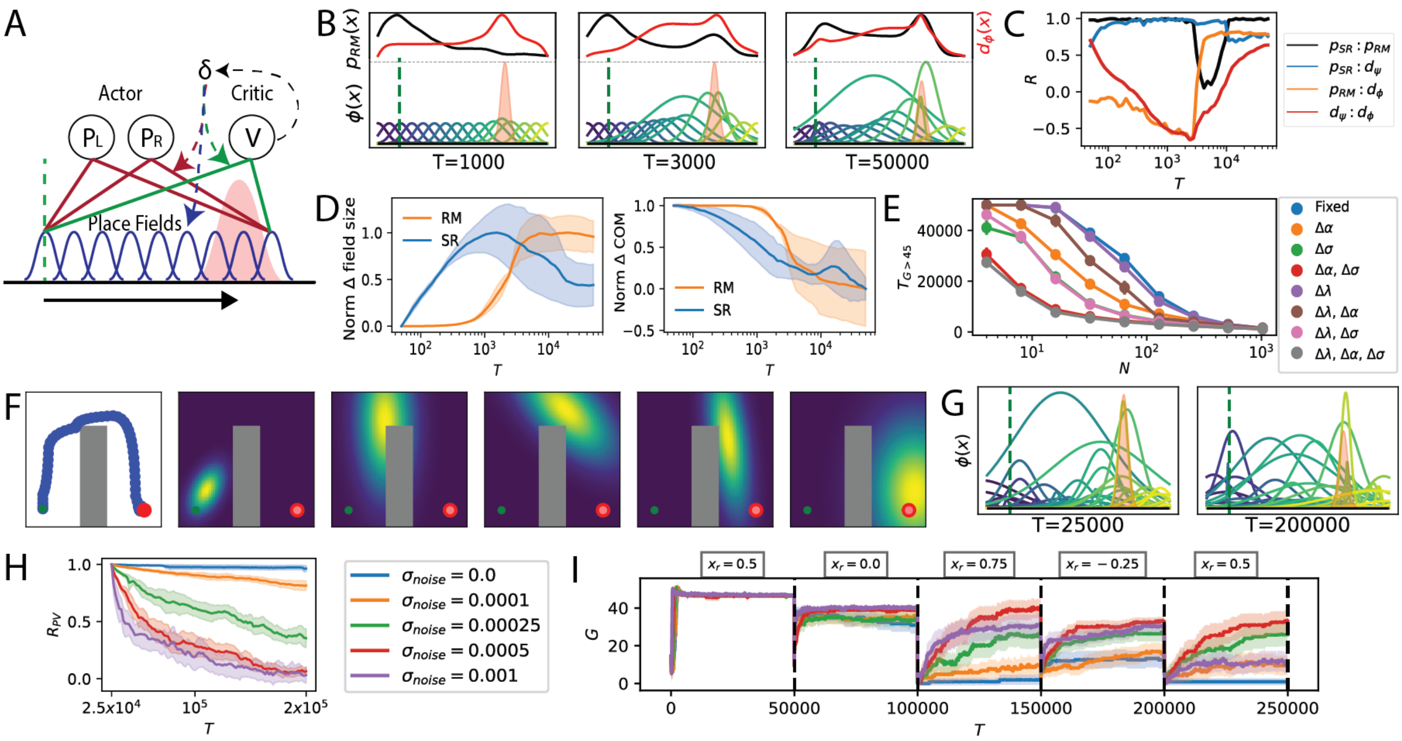

Figure 1: **(A)** Navigation agent with tunable place fields to maximize rewards. **(B)** Change in proportion of time spent in a location or location occupancy ($p(x)$, black), field density ($d(x)$, red) and individual field's spatial selectivity ($\phi(x)$, colored) during learning. **(C)** Correlation between location occupancy and field densities between Successor Representation (SR) and Reward Maximization (RM) agents. **(D)** Both SR and RM agents' place fields increase in size and the center of mass shift backwards towards the start location on averaged. **(E)** Ablation study shows optimizing field width σ, then amplitude α and finally center λ reduces the number of trials needed for policy convergence, when the agent achieves cumulative discounted return (G) of 45 per trial averaged over 300 trials. **(F)** Example agent trajectory (left) and place fields in a 2D arena with an obstacle (grey) after 50,000 learning trials. **(G)** After policy convergence, place field selectivity changes across trials with noisy field parameter updates, yet a high field density remains at the target (red). **(H)** Population vector (PV) correlation (R_{PV}) decreases faster with higher Gaussian noise. **(I)** When the target shifts to a new location after every 50,000 trials, adding Gaussian noise increased the agents' flexibility while agents without noise struggled to learn new targets (blue).

always aligned with the proportion of time spent in a location throughout learning (Fig. 1C, blue). Hence, the place field density learned by the RM ($d_{RM}(x)$) and SR ($d_{SR}(x)$) agents initially become anti-correlated but subsequently become correlated (Fig. 1C, red). In a 2D arena with an obstacle, individual place fields elongate along the obstacle and against the agent's trajectory (Fig. 1F) during reward maximization. Ablation studies (Fig. 1E) demonstrate that optimizing place field parameters reduced the number of trials needed for policy convergence, prioritizing width, then amplitude, and finally centers. However, increasing the field numbers extinguished the computational advantage of field optimization. Introducing different magnitudes of Gaussian noise to place field parameters caused fields to drift across trials (Fig. 1G), and the population vector correlation to decrease (Fig. 1H), although the agent maintained stable navigation performance (first 50,000 trials in Fig. 1I). When the target changed to a new location every 50,000 trials (Fig. 1I), agents with no noise (blue) struggled to learn the new target while agents with noise showed greater flexibility in new target learning. These results are consistent in a 2D arena with an obstacle.

Conclusion: Our normative model has the ability to learn a navigational policy while recapitulating several place field representation learning phenomena seen in experiments to make testable predictions. These include an amplification in firing rates besides fields moving towards the reward location. Whether the field density converges before policy convergence (RM dynamics) or if the density is always aligned with the animal's location occupancy (SR dynamics) needs to be experimentally verified. Lastly, noisy place field updates seem to suggest a functional role for noise in continual learning. Casually manipulating drift rates in place fields to influence continual learning will be an interesting experiment to conduct. We will expand our framework to other place cell descriptions to determine if the same phenomena can be recapitulated.