

# Abstract

Over the last decade, deep meta-reinforcement learning (deep meta-RL) has emerged to overcome the massive data requirements of traditional RL systems by enabling agents to essentially "learn to learn". The core concept involves using a standard, slow-updating deep RL algorithm to train the synaptic weights of a recurrent neural network (RNN) across a broad distribution of interrelated tasks. Through this process, the network's recurrent dynamics implement an emergent, entirely separate, and fast RL procedure configured to exploit the underlying structure of the training domain. To better characterize this process, we focused on structured two-armed bandit tasks with correlated reward probabilities, which provide a controlled environment to assess the emergence of this secondary RL algorithm (Wang et al. 2017; Wang et al. 2018). Foundational work in deep meta-RL has typically employed complex Long Short-Term Memory (LSTM) architectures to achieve this, successfully replicating learning behaviors observed in primates. However, their intricate mechanisms present significant barriers to interpretability. To achieve a straightforward, mechanistic grasp of the relationship between connectivity, activity and behavior, we focus on simpler networks where dynamics can be interpreted and compared to biological circuits. Our evaluations demonstrate that these simpler architectures effectively "learn to learn", successfully exploiting the underlying covariance structure of the tasks. By analyzing the behavior of our RNNs, and comparing it to that of LSTMs and Gated Recurrent Units (GRUs), we provide clearer insights into how minimal, low-dimensional connectivity structures can support complex meta-learning behaviors. Building on these results, we plan to systematically study the internal neural dynamics of our networks to uncover the mechanisms driving this meta-learning capability.

## References:

- Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D., & Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, 21(6), 860–868. <https://doi.org/10.1038/s41593-018-0147-8>
- Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., Blundell, C., Kumaran, D., & Botvinick, M. (2017). Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*. <https://doi.org/10.48550/arXiv.1611.05763>