Why Do Animals Need Shaping?
A Theory of Compositional Learning and Curriculum Learning

Jin Hwa Lee (jin.lee.22@ucl.ac.uk)

Stefano Sarao Mannelli (s.saraomannelli@ucl.ac.uk)

Andrew Saxe (a.saxe@ucl.ac.uk)

Sainsbury Wellcome Centre & Gatsby Computational Neuroscience Unit,
25 Howland St, London, United Kingdom

Abstract

Diverse studies in systems neuroscience begin with extended periods of training known as 'shaping' procedures. These involve progressively training on components of more complex tasks, and can make the difference between learning a task quickly, slowly or not at all. Despite the importance of shaping to the acquisition of complex tasks, there is as yet no theory that can help guide the design of shaping procedures, or more fundamentally, provide insight into its key role in learning. In this light, we propose and analyse a model of deep policy gradient learning on compositional reinforcement learning (RL) tasks. Using the tools of statistical physics, we solve exactly the learning dynamics and characterise different learning strategies including *primitives pre-training*, in which task primitives are studied individually before learning compositional tasks. We find a complex interplay between task complexity and the efficacy of shaping strategies. Overall, our theory provides an analytical understanding of the benefits of shaping in a class of compositional tasks and a quantitative account of how training protocols can disclose useful task primitives, ultimately yielding faster and more robust learning.

Keywords: compositionality; curriculum; learning theory

Introduction

Shaping is critical for effective learning in animals and humans (Skinner, 2019; Pavlov & Anrep, 1927; Elio & Anderson, 1984; Clerkin, Hart, Rehg, Yu, & Smith, 2017; Pashler & Mozer, 2013; Eckstein & Collins, 2021; Dekker, Otto, & Summerfield, 2022). Rather than teaching a complex task directly, shaping aims to gradually teach the components—*primitive tasks*—of a complex task and it is often exploited in the behavioural training of animals (Mushiake, Saito, Sakamoto, Sato, & Tanji, 2001; Laboratory et al., 2021; Grossman, Bari, & Cohen, 2022; Makino, 2023). Nevertheless, we do not have a theory that can quantitatively explain the role of shaping and how it changes the learning dynamics of intelligent systems which could give us deeper insights into these procedures.

Shaping is a form of curriculum learning that allows animals to learn and integrate primitive tasks to complete the higher level tasks (Schulz, Tenenbaum, Duvenaud, Speekenbrink, & Gershman, 2017; Hupkes, Dankers, Mul, & Bruni, 2020) leveraging a compositional structure of the tasks. This property, which is a crucial feature of shaping, is often called *systematic compositionality* that enables us to flexibly reuse previously acquired *primitives* by combining them (Chomsky, 2014; Smolensky, 1990; Lake, Linzen, & Baroni, 2019; Dehaene, Al Roumi, Lakretz, Planton, & Sablé-Meyer, 2022).

Here, we develop a simple theory of compositional task learning to obtain conceptual insight into the factors affecting learning performance. We borrow tools from statistical mechanics and the recent results in RL theory (Patel, Lee, Mannelli, Goldt, & Saxe, 2023; Bordelon, Masset, Kuo, & Pehlevan, 2023) to shed light on the learning dynamics of compositional tasks. By characterizing the curricula *primitives pre-training* and *vanilla training*, we reveal that curricula result in substantial differences in training time and robustness to noise during training.

Task and Model Setup

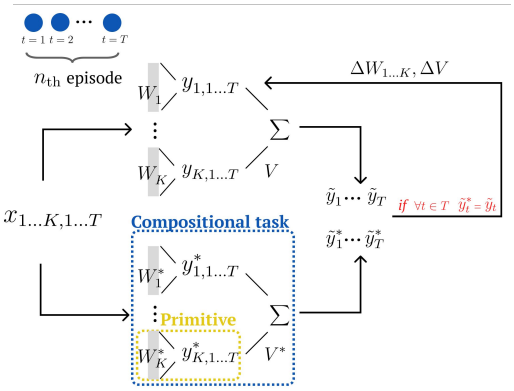

Figure 1: A compositional task with sequence length T and K primitives. A 'student' network learns to make the same decisions as a 'teacher'. Each primitive task is modelled as a pair of teacher-student network and K primitive tasks are linearly combined via the teacher context vector V^* that the student has to learn (V).

We consider a sequential decision-making task in which a student makes T binary choices in an episode. At time step t, given an observation x_t (for instance, representing visual input), the student makes a decision and all T decisions for all steps $t = 1 \cdots T$ need to be correct to get a reward. The correct decision for a compositional task is determined according to the compositional rule obtained by a linear combination of K *primitives*. Each k-th primitive task is a ran-

domly generated teacher network with parameters W_k^* and it is fixed throughout learning. At time step t, given a random task input $x_{k,t} \sim \mathcal{N}(0, \mathbb{I}_N)$, the teacher defines a correct decision $y_{k,t}^* = \text{sign}\left(W_k^* \cdot x_{k,t}/\sqrt{N}\right)$ for every k-th primitive task. The compositional task is defined by a linear combination of the primitives which we call the teacher context, $V^* \in \mathbb{R}_+^K$, $\tilde{y}_t^* = \text{sign}\left(\sum_{i=1}^K V_i^* \frac{W_i^* \cdot x_{i,t}}{\sqrt{N}}\right)$. This can be interpreted as linearly combining a set of task rules in an appropriate context to generate the correct decision. The student network has the same architecture as the teacher and learns its weights ($W_{1\ldots K}$ and V)– to generate the same decision generated by the teacher and maximize the reward. The student can update its weights only after it makes all T decisions, thus T can be interpreted as a task difficulty.

Results

In the following, we first show how to achieve an analytical solution of the learning dynamics and, finally, we demonstrate the benefits of primitives pre-training curriculum, i.e. shaping.

Learning Dynamics Analysis

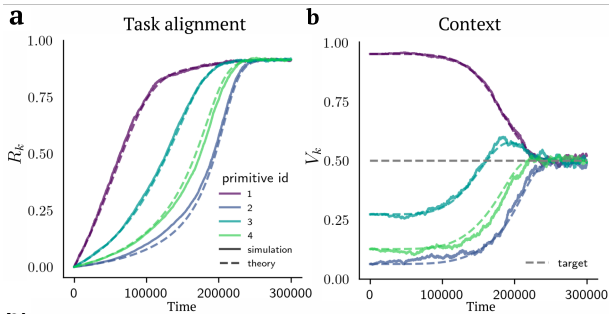

Figure 2: **a)** Learning dynamics of each primitive in compositional RL ($K = 4, T = 6$). Each primitive is learned in a different timescale. **b)** Learning dynamics of context in compositional RL. The student context for each primitive V_k gets aligned to the teacher value (target) $V^* = [0.5, 0.5, 0.5, 0.5]$ as learning proceeds.

The weight update rule of the student follows an approximate online policy gradient update. In the high-dimensional limit ($N \to \infty$), the stochastic learning dynamics concentrate to deterministic dynamics. We characterise the compositional learning dynamics by tracking the evolution of the two order parameters: the teacher-student alignment of each k-th primitive, $R_k = \frac{W_k \cdot W_k^*}{N}$, where R_k is a proxy for performance on the k-th primitive; and the context V_k. Using methods from statistical physics (Saad & Solla, 1995), we derive ordinary differential equations (ODEs) for order parameters which allow us to capture the learning dynamics as shown in Figure 2. We analyse two learning protocols: primitives pre-training and vanilla training. In primitives pre-training, each primitive is trained individually first, i.e. updating W_k if $y_{k,t} = y_{k,t}^* \forall t \in \lfloor T \rfloor$. Once the primitives are learned, the compositional task is learned - $W_{1\ldots K}$ and context vector V are updated when a sequence of T compositional decisions is made correctly; $\tilde{y}_t = \tilde{y}_t^* \forall t \in \lfloor T \rfloor$. In vanilla training, the student directly learns the compositional task without pre-training. From the ODEs, we derive the typical timescale for curriculum learning and vanilla learning when $K = 2$:

$$\tau_{\text{curriculum}} \sim (K 2^{T-2} + \tilde{P}_0^{1-T}), \tag{1}$$

$$\tau_{\text{vanilla}}^{(K=2)} \sim 2^{T-2} \frac{1}{(V_1^0 V_1^*)^2 + (V_2^0 V_2^*)^2}; \tag{2}$$

with $\tilde{P} = 1 - \frac{\tilde{\theta}}{\pi}$ and $\tilde{\theta} = \cos^{-1}\left(\frac{\sum_{i=1}^K V_i^* V_i R_i}{\sqrt{\sum_{i=1}^K (V_i^*)^2}\sqrt{\sum_{i=1}^K (V_i)^2}}\right)$, and where V_k^0 refers to the initial context value of k-th primitive.

Benefits of Curriculum Learning

Our proposed model of compositional RL confers two benefits of having primitive curriculum, namely, faster learning and robustness to noise during training.

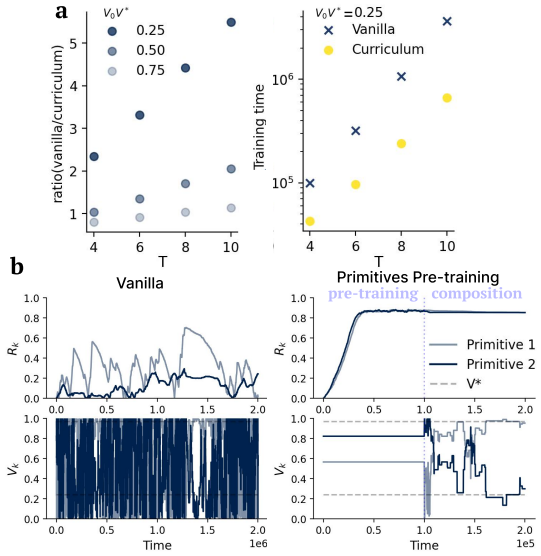

Figure 3: **a)** Speed boost from curriculum learning (K=2). Left: Ratio of the vanilla learning time to curriculum learning. Right: Total training time for vanilla vs. curriculum learning. **b)** Effect of noise during vanilla vs. curriculum learning ($\sigma_w = 0.01$, $\sigma_v = 0.1$).

Faster Learning We find that primitive pre-training can offer substantial learning speed benefits compared to vanilla training. As the task difficulty T increases, the training time in both vanilla learning and primitive pre-training grows exponentially (Figure 3a right), while their growth rate differs. Furthermore, having larger T and smaller cosine similarity between initial context V^0 and target context V^* significantly increases the learning speed boost from the primitives pre-training curriculum (Figure 3a left).

Robust Learning In the real world, learning is often noisy. In the presence of noise during gradient update, we compare the robustness of curriculum learning and vanilla learning. We inject i.i.d. Gaussian noise $\varepsilon_w \sim \mathcal{N}(0, \sigma_w)$ and $\varepsilon_v \sim \mathcal{N}(0, \sigma_v)$ into each element of the gradient of W and V, respectively, and compare the learning efficiency in simulation of the two training protocols. We varied the noise levels σ_w and σ_v and

found out that when σ_w is small but σ_v is relatively large, primitives pre-training provides significantly better learning than vanilla training as shown in Figure 3b.

Conclusion

In this study, we provide a theory of a simple case of task composition and curriculum learning. By formulating a compositional task with primitives and compositional context in the teacher-student setup, we derive a set of ODEs that can describe the learning dynamics of the task. This allows us to analytically study the distinct learning dynamics emerging in two different curricula, namely primitives pre-training and vanilla training. In our setting, we characterise potential benefits of curriculum learning: a speed boost in learning, and robustness to the noise during learning. Our model provides a quantitative understanding of the importance of shaping in learning compositional tasks.

Acknowledgments

We thank Athena Akrami, Chunyu A. Duan, Maria Eckstein, Kishore Kuchibhotla and Nishil Patel for useful discussions. SSM acknowledges Bocconi University for the hospitality during part of this project. This work was supported by a Sir Henry Dale Fellowship from the Wellcome Trust and Royal Society (216386/Z/19/Z) to AS, and the Sainsbury Wellcome Centre Core Grant from Wellcome (219627/Z/19/Z) and the Gatsby Charitable Foundation (GAT3755).

References

Bordelon, B., Masset, P., Kuo, H., & Pehlevan, C. (2023). Loss dynamics of temporal difference reinforcement learning. In *Thirty-seventh conference on neural information processing systems.*

Chomsky, N. (2014). *Aspects of the theory of syntax* (No. 11). MIT press.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055.

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences*.

Dekker, R. B., Otto, F., & Summerfield, C. (2022). Curriculum learning for human compositional generalization. *Proceedings of the National Academy of Sciences*, *119*(41), e2205582119.

Eckstein, M. K., & Collins, A. G. (2021). How the mind creates structure: Hierarchical learning of action sequences. In *Cogsci... annual conference of the cognitive science society. cognitive science society (us). conference* (Vol. 43, p. 618).

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & cognition*, *12*(1), 20–30.

Grossman, C. D., Bari, B. A., & Cohen, J. Y. (2022). Serotonin neurons modulate learning rate through uncertainty. *Current Biology*, *32*(3), 586–599.

Hupkes, D., Dankers, V., Mul, M., & Bruni, E. (2020). Compositionality decomposed: How do neural networks generalise? *Journal of Artificial Intelligence Research*, *67*, 757–795.

Laboratory, T. I. B., Aguillon-Rodriguez, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., ... others (2021). Standardized and reproducible measurement of decision-making in mice. *Elife*, *10*.

Lake, B. M., Linzen, T., & Baroni, M. (2019). Human few-shot learning of compositional instructions. *arXiv preprint arXiv:1901.04587*.

Makino, H. (2023). Arithmetic value representation for hierarchical behavior composition. *Nature Neuroscience*, *26*(1), 140–149.

Mushiake, H., Saito, N., Sakamoto, K., Sato, Y., & Tanji, J. (2001). Visually based path-planning by japanese monkeys. *Cognitive Brain Research*, *11*(1), 165–169.

Pashler, H., & Mozer, M. C. (2013). When does fading enhance perceptual category learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1162.

Patel, N., Lee, S., Mannelli, S. S., Goldt, S., & Saxe, A. (2023). The rl perceptron: Generalisation dynamics of policy learning in high dimensions. *arXiv preprint arXiv:2306.10404*.

Pavlov, I. P., & Anrep, G. V. (1927). *Conditioned reflexes. an investigation of the physiological activity of the cerebral cortex... translated and edited by gv anrep*. London.

Saad, D., & Solla, S. A. (1995). Exact solution for on-line learning in multilayer neural networks. *Physical Review Letters*, *74*(21), 4337.

Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M., & Gershman, S. J. (2017). Compositional inductive biases in function learning. *Cognitive psychology*, *99*, 44–79.

Skinner, B. F. (2019). *The behavior of organisms: An experimental analysis*. BF Skinner Foundation.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, *46*(1-2), 159–216.