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

Jin Hwa Lee [1] **Stefano Sarao Mannelli** [1] **Andrew Saxe** [1]

Abstract

Diverse studies in systems neuroscience begin with extended periods of curriculum training known as 'shaping' procedures. These involve progressively studying component parts 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. Modern deep reinforcement learning systems might implicitly learn compositional primitives within their multilayer policy networks. Inspired by these models, we propose and analyse a model of deep policy gradient learning of simple compositional reinforcement learning tasks. Using the tools of statistical physics, we solve for exact 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.

1. Introduction

Shaping is critical for effective learning in animals and humans (Skinner, 2019; Pavlov & Anrep, 1927). Rather than teaching a complex task directly, shaping aims to gradually teach components of the complex task. Imagine a mouse in a laboratory learning a decision-making task to get water rewards. To learn this task, the mouse needs to learn many sub-tasks often through reinforcement, for example, how to interact with a reward port, or how a light bulb predicts rewards. Depending on what shaping procedures are applied, the animals might learn faster, slower or even completely fail to learn a given task. Indeed, many studies in neuroscience exploit shaping stages for behavioural training of animals (Mushiake et al., 2001; Aguillon-Rodriguez et al., 2021; Grossman et al., 2022; Makino, 2023). In humans, studies similarly show that curriculum—the order in which tasks are trained—is important for generalisation (Elio & Anderson, 1984; Clerkin et al., 2017; Pashler & Mozer, 2013; Eckstein & Collins, 2021; Dekker et al., 2022). 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 and guide design choices in experiments.

One crucial feature of shaping is compositionality. Shaping stages involve breaking down complex tasks into simpler components and learning them individually before animal learns to integrate those primitive tasks in the right context to complete the higher level tasks (Schulz et al., 2017; Hupkes et al., 2020). In other words, shaping is a form of *curriculum learning*, leveraging the compositional structure of complex tasks. Many real world tasks require mastering different skills and combining them in the right way. This property is often referred to as systematic compositionality, which enables us to flexibly reuse previously acquired *primitives* by combining them (Chomsky, 2014; Smolensky, 1990; Lake et al., 2019; Dehaene et al., 2022).

On the other hand, the recent success of deep reinforcement learning (RL) in highly complex tasks motivates using these models to study compositional task and curriculum learning. Many empirical studies indicate that multilayer policy networks might implicitly learn compositional primitives if not explicitly (Baker et al., 2020; Zhang et al., 2020; Nachum et al., 2018; Andreas et al., 2017; Barreto et al., 2020). This motivates using a deep policy network model of composi-

[1]Sainsbury Wellcome Centre for Neural Circuits and Behaviour & Gatsby Computational Neuroscience Unit, University College London, London, UK
. Correspondence to: Jin Hwa Lee <jin.lee.22@ucl.ac.uk>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

tional RL tasks to study learning dynamics and understand the role of shaping.

Our goal in this paper is to develop a simple theory of compositional task learning, to obtain conceptual insight into the factors affecting learning performance. The theoretical formulation of compositional learning in RL is challenging due to the limited mathematical tools available. In this study, we borrow tools from statistical mechanics and the recent framework of analyzing RL dynamics using the teacher-student setup (Patel et al., 2023; Bordelon et al., 2023) to shed light on the learning dynamics of compositional learning in RL. We characterize different curricula including *primitives pre-training* and *vanilla training*. We reveal that different curricula result in differences in training time and robustness to noise during training, indicating potential ramifications of shaping and curriculum learning in complex compositional tasks.

Main contributions

- We model a compositional RL task and its primitives in the teacher-student setup and derive an asymptotically exact set of ordinary differential equations (ODEs) describing the evolution of the order parameters in online learning.
- We analyze learning dynamics under two different curricula: 1) *primitives pre-training*, in which an agent is first trained on primitives before learning the compositional task; and 2) *vanilla training*, in which an agent is trained directly on the compositional task.
- We investigate the benefits of curriculum learning based on our learning dynamics analysis—principally, faster learning and robustness to noise.
- Finally, we connect our study to the literature from psychology and neuroscience.

Related Work

The **teacher-student framework** and the tools of statistical physics have been widely used to study learning dynamics in neural network models (Gardner & Derrida, 1989; Seung et al., 1992; Biehl & Schwarze, 1995; Saad & Solla, 1995; Engel, 2001; Zdeborová & Krzakala, 2016). Diverse learning behaviors and phenomena in different settings have been studied using the paradigm (Goldt et al., 2019; Mannelli et al., 2019; Advani et al., 2020; Bahri et al., 2020; Saglietti et al., 2022; Gerace et al., 2022). In particular, our model formulation of a compositional task is largely inspired by the RL perceptron (Patel et al., 2023), a recent theoretical framework to study the average dynamics of a prototypical policy gradient RL network in high-dimensions. The agent faces a high-dimensional input and learns to select actions through sparse rewards. In our work, we introduce hierarchy to the RL perceptron to bring a notion of primitives and higher-level task composition, which is defined by a combination of primitives.

Curriculum learning has been studied extensively in neuroscience and psychology (Ahissar & Hochstein, 1997; Hornsby & Love, 2014; Skinner, 2019; Dekker et al., 2022; Makino, 2023; Hocker et al., 2024) across different species (humans, rodents, pigeons) and tasks (decision making, perceptual learning, motor learning). In the machine learning community, while its importance has long been acknowledged (Bengio et al., 2009; Wang et al., 2021), a recent survey shows only marginal benefits of curriculum in standard supervised learning benchmarks (Wu et al., 2021). However, the right curricula have shown greater effects in the context of reinforcement learning (Narvekar et al., 2020; Tessler et al., 2017; Karpathy & Van De Panne, 2012; Tong et al., 2023), especially with hierarchical task structure and network architectures.

Leveraging **compositionality** for efficient machine learning systems has been proposed in many modalities and contexts. In vision models, disentangled representations seek to factorize input variance into semantic constituents, modeling the input data as a composition of those factors (Montero et al., 2020). In the domain of language, meta-training on memory-augmented neural networks exhibits certain forms of compositional generalisation (Lake, 2019). Furthermore, it has been shown that modular motifs emerge from training with multiple cognitive tasks and they can be repurposed to rapidly learn new tasks that combine these motifs (Driscoll et al., 2022; Yang et al., 2019). Yet theoretical understanding of learning compositional tasks is limited, though there have been recent studies which formulate compositionality in a data-generating process perspective (Wiedemer et al., 2024; Okawa et al., 2023).

In RL, compositionality has been often studied in the context of temporally extended action planning. Hierarchical RL instantiates the idea of decomposing planning using temporal abstraction to chain learned skills to perform a complex behavior on a longer timescale (Sutton et al., 1999b; Bacon et al., 2017). The compositionality in our study is conceptually more similar to *functional composition*, which has been addressed with explicitly modularised neural networks for each task or policy primitive (Devin et al., 2017; Saxe et al., 2017; Goyal et al., 2021; Yang et al., 2020; Mittal et al., 2020; Mendez et al., 2022). In a similar spirit, recent theoretical work showed that modular networks can discover compositional structure through learning sparse combinations of modules (Schug et al., 2024). Our work focuses on compositional learning in RL with policy gradient methods.

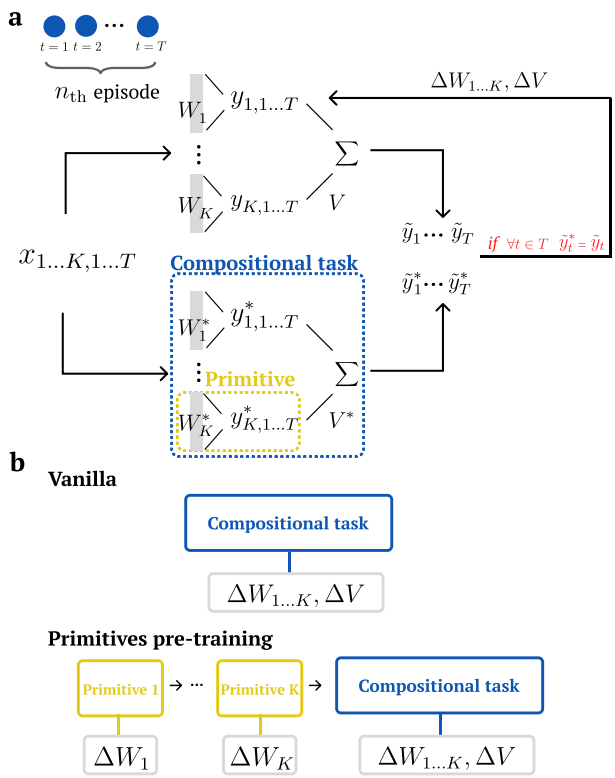

Figure 1. **Schema of primitives and a compositional task in the teacher-student setup. a)** Illustration of a *compositional task* with K *primitives* in the teacher-student setup. Each primitive $k = 1...K$ is modelled by a fixed teacher network W^*_k and learnable student network W_k. At each timestep t of an episode with total length T, the output of each primitive teacher W^*_k from task-dependent input $x_{k,t}$ is blended using fixed teacher context V^*_k to generate a compositional choice $\tilde{y}^*_t$. The goal of learning for the student primitives W_k and context V is to generate the same compositional choice $\tilde{y}_t$ as the teacher choice. Importantly, the student updates the weights only when it makes all T choices in the episode correctly, reflecting a sparse reward setting. **b)** Schematic of two curricula: vanilla training and primitives pre-train ing. In vanilla training, an agent directly learns the compositional task. In primitives pre-training, an agent learns each primitive first to expert level and then learns the compositional task.

2. Compositional Learning and Primitives

To begin with, we conceptualize compositional learning and primitives with the following example. Ito et al. (2022) studied a compositional task that consists of a combination of rules of three domains (sensory, logic, motor) that a subject needs to combine to make a decision. For example, the sensory rule can be "red" or "blue", the logic rule can be "both" or "either", and the motor rule can be "right" or "left". With such rules, an example compositional task built from the rules could be "If $<$ **both** $>$ stimuli are $<$ **red** $>$, press $<$ **right** $>$". The primitives in this context are individual rules, which need to be selected and combined appropriately to make a correct decision for the compositional task. Following this concept, we formulate a compositional generalisation task with *primitives* and *compositional context*, where a compositional task is modeled as a linear combination of the primitives given by the context.

In order to theoretically evaluate a network trained to perform a given task, we need a tractable data generative model that generates the ground-truth input-output mapping that a network needs to learn. One standard way is using the teacher-student framework (Seung et al., 1992; Zdeborová & Krzakala, 2016). Within this framework, a teacher is a network used to label randomly sampled inputs, while a student is a network whose goal is to optimise its weights and match the label generated by the teacher. The teacher network can be thought of as an oracle network or an ideal post-learning weight configuration. Using a teacher-student setup as a generative model, we construct synthetic inputs x and labels y^* from a teacher network (we will denote the teacher network and correct labels with $*$), as schematized in Figure 1a.

In the following, we describe our teacher-student model of *primitives* and *compositional tasks*.

Primitives. We define K primitive tasks, where each primitive $k = 1, ..., K$ is constructed with an independent teacher-student pair. Specifically, each primitive task is represented as an RL perceptron (Patel et al., 2023) which allows us to model policy learning with high-dimensional inputs. In each primitive task, the student needs to make T binary choices: T will be an important parameter of our problem acting as a proxy for difficulty that we call *episode length*. At each time step t of primitive task k, the student makes a binary choice given a task-dependent high-dimensional input $x_{k,t} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}^N)$ of dimension N. At time step t, the choice of student is correct if $y_{k,t} = \text{sign}(W_k \cdot x_{k,t}/\sqrt{N})$ matches $y^*_{k,t} = \text{sign}(W^*_k \cdot x_{k,t}/\sqrt{N})$ provided by the teacher. After the student makes choices for $t = 1, ..., T$, the student receives a delayed reward and updates its weights following a perceptron-like update rule,

$$\Delta W_k^{n+1} = \frac{\eta_w}{\sqrt{N}} \left(\frac{1}{T} \sum_{t=1}^{T} y_{k,t} x_{k,t} \right) \mathbb{I}(\Phi(y_{k,1:T}, y^*_{k,1:T})) \tag{1}$$

where n indicates the episode index, $\mathbb{I}$ is an indicator function and Φ is a boolean criterion (i.e., the student receives a sparse reward and updates its weights at the end of each episode only when it makes all T choices correctly). Figure 1a illustrates the weight update scheme. We use this specific learning rule since Patel et al. (2023) showed that this is analogous to the REINFORCE algorithm (Sutton et al., 1999a), a widely adopted policy gradient learning

algorithm. See Appendix A for details.

Compositional Task. The compositional task is constructed by blending the outputs of K primitives via a context vector $V^* \in \mathbb{R}_+^K$. We denote variables related to the compositional task with " ˜ ". The correct choice at timestep t in the compositional task is given by $\tilde{y}_t^* = \text{sign}\left(\sum_{k=1}^K V_k^* \frac{W_k^* \cdot x_{k,t}}{\sqrt{N}}\right)$, a linear combination of teacher primitives and context weights. Similarly, the student makes a choice from student primitives and context weights, $\tilde{y}_t = \text{sign}\left(\sum_{k=1}^K V_k \frac{W_k \cdot x_{k,t}}{\sqrt{N}}\right)$. This can be interpreted as linearly combining a set of task rules in an appropriate context to generate the choice. The scheme is shown in the blue-dotted rectangle of Figure 1a.

We derive the compositional version of the Eq. 1, which allows us to model a policy gradient learning equivalent to REINFORCE on this compositional task. When the student is trained for the compositional task, both primitives and context are updated by the following update rules:

$$\Delta W_k^{n+1} = \frac{\eta_w}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T} \tilde{y}_t V_k x_{k,t}\right)\mathbb{I}(\Phi(\tilde{y}_{1:T}, \tilde{y}_{1:T}^*)), \tag{2}$$

$$\Delta V_k^{n+1} = \frac{\eta_v}{N}\left(\frac{1}{T}\sum_{t=1}^{T} \tilde{y}_t \frac{W_k \cdot x_{k,t}}{\sqrt{N}}\right)\mathbb{I}(\Phi(\tilde{y}_{1:T}, \tilde{y}_{1:T}^*)), \tag{3}$$

where η_w and η_v are learning rates for the primitive weights and context weights, respectively. Analogously to the primitive task update rule, the student only updates its primitive weights and context weights at the end of each episode, when all T choices are made correctly. This learning rule is a proxy of policy gradient learning for our model. See Appendix A for details.

Note that our model of compositional tasks relies on learning each primitive independently using separate networks and combining them post hoc, which is only one form of compositional learning. This is equivalent to assume that primitives are orthogonal which naturally yields a modular structure similar to that in the literature (Schug et al., 2024; Jarvis et al., 2023; Driscoll et al., 2022).

3. Learning Dynamics in Different Curricula

Often curriculum learning in machine learning (Bengio et al., 2009; Wu et al., 2021; Saglietti et al., 2022) is implemented by feeding training data in order of difficulty. The concept of compositionality adds a degree of complexity in task structure and naturally suggests a curriculum for primitives and compositional tasks. In this section, we introduce two curriculum strategies—*primitives pre-training* and *vanilla learning*—and investigate their different learning dynamics. In *vanilla training*, a naïve agent is trained directly with the compositional task while in *primitives pre-training*, an agent first learns the primitives to expert level and then learns the compositional task (as illustrated in Figure 1b).

Following the approach of (Saad & Solla, 1995; Biehl & Schwarze, 1995; Patel et al., 2023), we derive a set of dynamical equations that capture the learning dynamics in the high dimensional limit ($N \to \infty$). This approach allows us to describe the evolution of the high-dimensional student weights with a handful of order parameters,

$$Q_k = \frac{W_k \cdot W_k}{N},\ R_k = \frac{W_k \cdot W_k^*}{N},\ S_k = \frac{W_k^* \cdot W_k^*}{N}. \tag{4}$$

Intuitively, Q_k is the norm of the student primitive task k. Similarly, S_k is the norm of the teacher primitive task k. R_k is interpreted as the alignment or correlation between the student and teacher of the primitive task k. The generalisation error for a single primitive task k is defined by the *average* disagreement between y_k^* and y_k and it can be described with the order parameters as follows,

$$\epsilon_{k,g} = \frac{1}{\pi}\cos^{-1}(R_k/\sqrt{Q_k S_k}). \tag{5}$$

Similarly, the generalisation error on the compositional task $\tilde{\epsilon}_g$ is defined by the average disagreement between $\tilde{y}^*$ and $\tilde{y}$ and becomes

$$\tilde{\epsilon}_g = \frac{1}{\pi}\cos^{-1}\left(\frac{\sum_{i=1}^K V_i^* V_i R_i}{\sqrt{\sum_{i=1}^K V_i^{*2} S_i}\sqrt{\sum_{i=1}^K V_i^2 Q_i}}\right). \tag{6}$$

Therefore, tracking the order parameters (Eq. 4) and V_k allows us to characterise the generalisation error of the compositional and primitive tasks during the entire learning period as shown in Eq. 5-6. As we approach the high dimensional limit ($N \to \infty$), the learning curves concentrate to a typical trajectory described by the ODEs below. We introduce a variable $z = \frac{n}{N}$, the ratio of the number of episodes to the input dimension, which is analogous to continuous 'time' in the limit $N \to \infty$. for the full derivation refer to Appendix B. The equations describe the time evolution of the order parameters and V_k in continuous time (z),

$$\frac{dQ_k}{dz} = \frac{2\eta_w}{\sqrt{2\pi}}\left(\frac{V_k^2 Q_k}{\sqrt{\sum_{i=1}^K V_i^2 Q_i}} + \frac{V_k V_k^* R_k}{\sqrt{\sum_{i=1}^K V_i^{*2} S_i}}\right)\tilde{P}^{T-1} + \frac{\eta_w^2}{T} V_k^2 \tilde{P}^T \tag{7}$$

$$\frac{dR_k}{dz} = \frac{\eta_w}{\sqrt{2\pi}}\left(\frac{V_k^2 R_k}{\sqrt{\sum_{i=1}^K V_i^2 Q_i}} + \frac{V_k V_k^* S_k}{\sqrt{\sum_{i=1}^K V_i^{*2} S_i}}\right)\tilde{P}^{T-1} \tag{8}$$

$$\frac{dV_k}{dz} = \frac{\eta_v}{\sqrt{2\pi}}\left(\frac{V_k Q_k}{\sqrt{\sum_{i=1}^K V_i^2 Q_i}} + \frac{V_k^* R_k}{\sqrt{\sum_{i=1}^K V_i^{*2} S_i}}\right)\tilde{P}^{T-1} \tag{9}$$

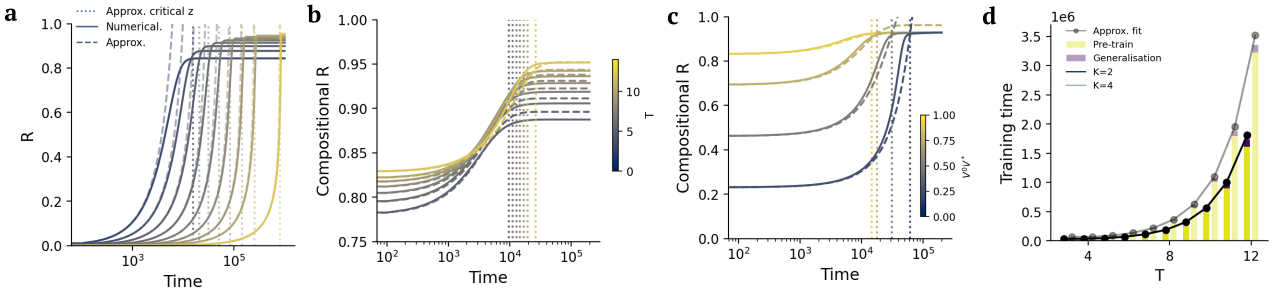

Figure 2. **Learning dynamics for primitives pre-training curriculum. a)** Learning dynamics of a single primitive task of episode length T. The dashed line depicts the approximate dynamics of R from Eq. 10, which captures the numerical integration of the ODE near the beginning of learning ($N = 1000, \eta_w = 1$). The dotted line denotes the approximate timescale τ. **b)** Learning dynamics during the compositional generalization phase after primitives pre-training across different T ($N = 1000, \eta_w = 1, \eta_v = 1, K = 2, V^0V^* = 0.5$). **c)** Learning dynamics of compositional generalization across different cosine similarities between initial V^0 and target V^* ($N = 1000, \eta_w = 1, \eta_v = 1, K = 2, T = 8$). **d)** Simulated total training time (bars) and approximated total training time from our analytical derivation (lines) for the primitives pre-training curriculum across different T (x-axis) and K (indicated by different transparency). The pre-training time of the primitives (yellow) is the major burden compared to the relatively faster compositional generalisation training time (purple) ($\eta_w = 1, \eta_v = 1, V^0V^* = 0.5$).

where $\tilde{P} = 1 - \frac{1}{\pi}\tilde{\epsilon}_g$ is the probability of making a correct compositional choice. In this study, we focus on the spherical case where $\|V^*\| = \|V\| = 1$ and $S_k = Q_k = 1$. This simplifies the expression of $\tilde{\epsilon}_g$ and $\epsilon_{g,k}$ and we only need to keep track of V_k and R_k.

In the next sections we analyse evolution of the order parameters (Eqs. 7-9) for *primitives pre-training* and *vanilla training* in detail and obtain the relevant timescales of learning.

3.1. Primitives Pre-training

Leveraging the compositional structure of the task, the primitives pre-training strategy aims to train a student on primitives first, and later generalise to the compositional task. During pre-training, only primitive student weights $W_{1\ldots K}$ are updated (Eq. 1) without information from the context V. The compositional task is presented to the student model once it reaches expert level on the primitives R_0 (defined as 99% of the maximum accuracy that the student can reach with T), and both $W_{1\ldots K}$ and V are updated according to Eqs. 2-3.

Timescale of primitive learning. Using the ODEs for a single primitive task from the RL perceptron (Patel et al., 2023), we analytically derive a closed-form approximation of $R_k(z)$,

$$R_k(z) = -\frac{\pi}{2} + \pi\left(\frac{\eta_w}{\sqrt{2\pi\pi}}(2-T)z + 2^{(T-2)}\right)^{\frac{1}{2-T}} \quad (10)$$

which captures the learning timescale of a single primitive of episode length T

$$\tau_{\text{primitive}} \sim \frac{1}{\eta_w(T-2)}2^{(T-2)}. \quad (11)$$

See Appendix C.1 for the derivation. In Figure 2a, we show the learning dynamics of a single primitive from the numerical integration (solid line), approximated dynamics (dashed line) and the approximated timescale $\tau_{\text{primitive}}$ (dotted line). We can see that the approximation captures the dependence of the training timescale of a primitive task on episode length T. From this result we see that the learning timescale for a single primitive task, $\tau_{\text{primitive}}$, is exponential in T. Namely, the larger the episode length T, the harder the primitive task is to learn. Intuitively, this is due to sparser rewards and fewer updates in the early stages of learning which make the overall dynamics slower.

After primitives pre-training, the student is trained on the compositional task.

Compositional generalization. In the compositional generalization phase of the curriculum, the student learns to adapt the acquired primitives to the compositional task. In this learning stage, both the student primitive weights $W_{1\ldots K}$ and context V are updated following Eqs. 2-3.

After primitives pre-training, each primitive reaches expert alignment R_0 and we can derive the approximate dynamics of $\tilde{R}$ around $R_k \approx R_0$. This allows us to determine the dominant factor of the training time for compositional generalization,

$$\tau_{\text{composition}} \sim \frac{1}{\eta_v(T-1)}\tilde{P}_0^{2-T} \quad (12)$$

$$\tilde{P}_0 = \left(1 - \frac{1}{\pi}\cos^{-1}\left(R_0 \sum_k V_k^0 V_k^*\right)\right) \tag{13}$$

where V_k^0 is a initial student context weight for primitive k. See Appendix C.2 for the detailed expression and derivation. This approximation accords well with simulation results as shown in Figure 2b-c.

Altogether, given K primitives and episode length T, the total training time of the *primitives pre-training* curriculum is of the order

$$\tau_{\text{curriculum}} \sim (K2^{T-2} + \tilde{P}_0^{2-T}). \tag{14}$$

Since $\tilde{P}_0 \geq 1/2$ and $K \geq 2$ for the compositional task, as the primitive task complexity (episode length T) increases, the burden of training time mostly comes from primitive training rather than compositional generalization as clearly illustrated in Figure 2d. As a consequence, after the student masters the primitives, it is relatively fast to adapt to new compositional tasks. Note that an important role is played by the alignment of the initial context of the student V^0 with the target context V^* as the better alignment makes larger $\tilde{P}_0$.

3.2. Vanilla Training

In vanilla training, the student faces the compositional task immediately, updating the weights following Eqs. 2-3. In Figure 3a, we can see that the student successfully learns the primitive tasks and context. Notably, each primitive is learned with a different timescale in vanilla learning while the context learning for each primitive occurs with a similar timescale.

In the following section, we aim to understand the learning dynamics of primitives and context in the vanilla learning strategy. We restrict our consideration to the $K = 2$ case for simplicity.

Initial context alignment drives primitive learning speed. In vanilla learning, the learning dynamics is much more convoluted due to the coupling of the order parameters R and V. Yet, we can still capture important factors which determine the early learning timescale for each primitive in vanilla learning. In this section, we first show how the alignment of the initial context V^0 and the target context V^* modulates the learning time of the primitives, explaining the different learning timescale of each primitive observed in Figure 3a.

In the early phase of vanilla learning, we can assume that $R_{k' \neq k} \approx 0, \tilde{R} \approx R_k V_k V_k^* \ll 1$. Under these assumptions, we approximate the early dynamics of R_k as follows:

$$\frac{dR_k}{dz} \approx \frac{\eta_w}{\sqrt{2\pi}} V_k^0 V_k^* \left(\frac{1}{2} + \frac{V_k^0 V_k^* R_k}{\pi}\right)^{T-1}. \tag{15}$$

Solving the ODE above, we get a closed-form approximation of early R_k dynamics and the characteristic time in vanilla learning as following,

$$R_k(z) = \frac{1}{V_k^0 V_k^*}\Big(-\frac{\pi}{2} + \pi\Big(\frac{\eta_w}{\sqrt{2\pi}\pi}(2-T)\Big(V_k^0 V_k^*\Big)^2 z + 2^{T-2}\Big)^{\frac{1}{2-T}}\Big) \tag{16}$$

$$\tau_{\text{vanilla},k} \sim \frac{1}{\eta_w (T-2)} 2^{(T-2)} \Big(\frac{1}{V_k^0 V_k^*}\Big)^2. \tag{17}$$

From Eq. 17 we can observe that the learning timescale of primitive k is attenuated by a $1/(V_k^0 V_k^*)^2$ term compared to the primitive training shown in Eq. 10. In other words, each primitive has a different timescale for learning determined by $1/(V_k^0 V_k^*)^2$.

Intuitively, V_k^0 represents the initial attention of the model to the primitive k. V_k^* represents the actual relevance of the primitive k in the compositional task. Hence higher $V_k^0 V_k^*$ means that the primitive stands out during learning and is in fact important, leading to faster learning. Given that the context vector has norm 1, the student is forced to trade-off in which direction to put more relevance or attention. For sake of comprehension, we denote the first trained primitive as $k = 1$ and the later trained one as $k = 2$ in the following. In Figure 3b, we can clearly see that the approximation holds for $V_k V_k^*$ sufficiently apart. Nevertheless it fails when the difference between $\Delta(V^0 V^*)^2 = (V_1^0 V_1^*)^2 - (V_2^0 V_2^*)^2$ becomes small. In the next section, we further investigate the cause of this gap, which brings us to a *multiplicative effect* in primitive learning during vanilla learning.

Emergence of multiplicative primitive learning. So far in considering R_k, we assume $R_{k'} \approx 0, k \neq k'$. However, as learning progresses, the overall $\tilde{R}$ becomes larger due to the contribution of $R_{k'}$, and this accelerates the learning of primitive k. This effect explains why the approximation for $R_{k=1}$ predicts slower learning than the actual dynamics, and why the gap is bigger when $\Delta V^0 V^*$ is smaller (Figure 3b top). On the other hand, from the perspective of R_2, large $\Delta(V^0 V^*)^2$ indicates that R_1 already has been significantly increased, and causes a strong divergence from the numerical integration (Figure 3b bottom).

These observations indicate that learning one primitive accelerates learning of the rest of the primitives. Interestingly, despite the task design where all primitives are modularised and independent, the learning of one primitive increases $\tilde{R}$ and leads to rapid learning of the other primitives. We probe this multiplicative primitive learning effect in vanilla compositional task learning with a controlled experiment. We initialise the first primitive R_1 at different levels and simulate the learning dynamics of the second primitive task under the vanilla learning strategy. As shown in Figure. 3c,

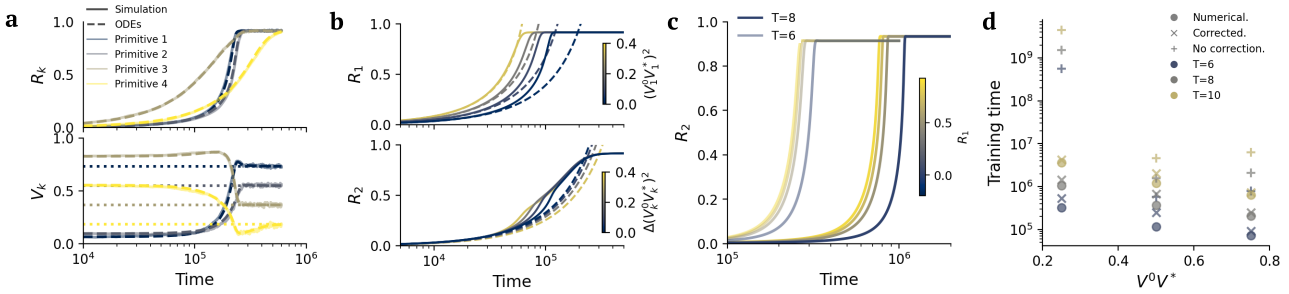

Figure 3. **Learning dynamics in vanilla learning. a)** Learning dynamics of each primitive task R_k (top) and context V_k (bottom) during vanilla training ($T = 6$, $K = 4$, $V^0V^* = 0.5$, $\eta_w = 1$, $\eta_v = 1$, $N = 1000$). Dynamics from simulation of the update rules Eq. 2-3) (solid) are well described by the ODEs of Eq. 7-9 (dashed). The dotted line is the target context V^*. **b)** $V_k^0V_k^*$ controls the delay of primitive learning. Top: Learning dynamics of the first learned primitive task (R_1) for varying $(V_1^0V_1^*)^2$. Smaller $V_1^0V_1^*$ leads to more delayed learning of the primitive. Bottom: Learning dynamics of the later learned primitive task R_2 for different $\Delta(V_k^0V_k^*)^2$ values. As the difference of $\Delta(V_k^0V_k^*)^2$ is bigger, the gap between approximation and the numerical integration is bigger. **c)** Demonstration of *multiplicative learning effect* of the primitives. The second primitive learns faster as the first primitive has better knowledge. Each learning curve depicts the second primitive learning dynamics for a given initial overlap of the first primitive R_1 (color scale). **d)** The numerical training time (circle), approximated training time corrected for the multiplicative effect (cross), and non-corrected (plus) with respect to cosine similarity of the initial V_0 and the target V^* and the episode length T. We can appreciate the importance of multiplicative effect in vanilla learning.

a high initial value of R_1 leads to fast learning of the second primitive task, R_2.

In the next section, we derive the timescale for vanilla learning based on our understanding of attenuation from V_0V^* and the multiplicative effect.

Timescale of vanilla learning. To derive the approximate total learning timescale of vanilla learning, we consider the primitive which is learned last (R_2), since the last task to be learnt will decide the total compositional task training timescale. We take the multiplicative effect caused by the first primitive into account and add a correction term from the contribution of R_1 in Eq. 16. See Appendix C.3 for the full derivation. In essence, the total learning time of vanilla learning can be described as

$$\tau_{\text{vanilla}}^{(K=2)} \sim \frac{1}{\eta_w(T-2)} 2^{(T-2)} \frac{1}{(V_1^0V_1^*)^2 + (V_2^0V_2^*)^2}. \quad (18)$$

The correction from the multiplicative effect is critical in determining the total learning time scale and accounting for it provides good agreement with the total training time in vanilla learning obtained from numerical integration as shown in Figure. 3d.

4. Curriculum Learning Effects

4.1. Improved Learning Speed

Finally, after analysing the learning time for the two different curricula strategies, we compare the learning time across T and the alignment between the initial and the target context V^0V^* for $K = 2$. We find that primitive pretraining can offer substantial learning speed benefits compared to vanilla training. Figure 4a shows an example where primitive pre-training (solid line) is much faster than vanilla training (dashed line). Turning in more detail to the effect of primitive difficulty, we can see that increasing T leads to exponential growth of training time in both vanilla learning and primitive pre-training (Figure 4b right), while their growth rate differs. Furthermore, as we can infer from Eq. 12 and Eq. 18, having larger T (higher task complexity) and smaller cosine similarity between V^0 and V^* significantly increases the learning speed boost from the primitives pre-training curriculum (Figure 4b left). See Appendix D and Figure 6 for a similar tendency in $K = 4$ case and under an extended learning rate hyperparameter search. Hence primitives pretraining confers a learning speed advantage in our compositional task setting.

4.2. Improved Noise Robustness

In this section we investigate another potential benefit of primitives pre-training curriculum: robustness to the noise in training. In the real world, learning is not noise-free. In the presence of noise during learning, we study the robustness of curriculum learning in comparison to 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 compare the average deviation of $\tilde{R}_{\text{noise}}$ from the $\tilde{R}_0$ which is reachable without noise during the training. We allow both training protocols enough

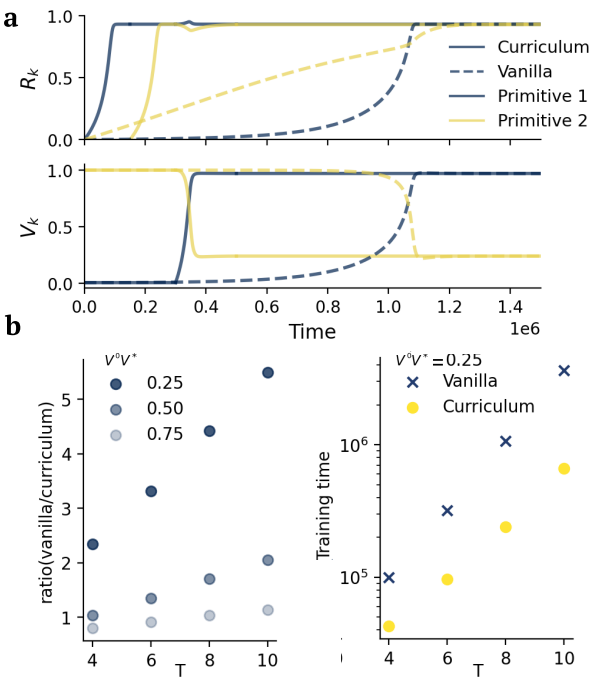

Figure 4. **Speed boost from curriculum learning. a)** Example of faster learning under the primitives pre-training strategy (solid) compared to vanilla learning (dashed) ($K = 2$, $V^0V^* = 0.25$, $\eta = 1$, $N = 1000$). **b)** Left) The ratio between training time of vanilla learning and primitive pre-training curriculum learning from numerical simulation. Different colors indicate the initial cosine similarity of the context weights V^0V^* ($K = 2$). Right) The numerical learning time from different strategies ($V^0V^* = 0.25$).

training iterations to reach $\tilde{R}_0$ in the noise-free case. We add the noise term to the updates throughout training. Figure 5b shows the difference in learning performance of the training protocols as we vary the noise scales σ_w and σ_v. When σ_w is large, both vanilla learning and curriculum learning suffer and fail to learn the compositional task ($\sigma_w \geq 0.05$ in the heatmap) irrespective of σ_v. When both σ_w and σ_v are small ($\sigma_w < 0.05, \sigma_v < 0.05$ in the heatmap), both learning protocols successfully learn the task. An interesting range is where σ_w is small but σ_v is relatively large ($\sigma_w \leq 0.05, \sigma_v \geq 0.05$ in the heatmap). In this range, primitives pre-training provides significantly better learning than vanilla training. In Figure 5a ($\sigma_w = 0.01$, $\sigma_v = 0.1$), we illustrate an example case where primitives pre-training curriculum enables learning of all primitives and the composition context while vanilla learning fails.

The robustness to noisy training in structured compositional tasks under curriculum learning comes from separating the two sources of noise: primitives learning and compositional generalization. In the primitives pre-training curriculum, the student initially only needs to handle the noise in the primitives. After learning the primitives, the effective noise only comes from the context learning, since the primitives are already well-aligned to the teacher and the fluctuations for the primitives are small. In contrast, in vanilla learning, the student needs to learn the primitive task and context learning simultaneously with mixed noise sources which makes learning more challenging.

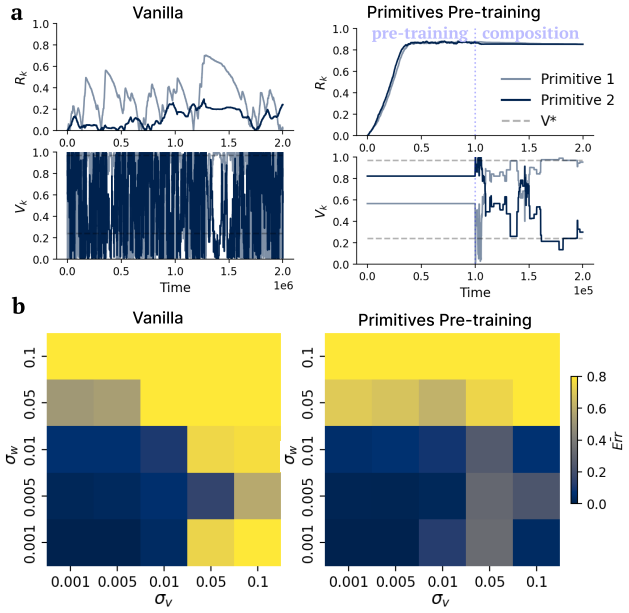

Figure 5. **Robustness to noise in curriculum learning. a)** Example of R_k and V_k dynamics during noisy learning ($K = 2$, $V^0V^* = 0.75$, $\sigma_w = 0.01$, $\sigma_v = 0.1$), in vanilla training and primitives pre-training. **b)** Average difference of $\tilde{R}_0 - \tilde{R}_{\text{noise}}$ during the last 5000 iterations of training in two different training protocols across a range of σ_w and σ_v($K = 2$, $V^0V^* = 0.25$). Primitives pre-training is more robust compared to vanilla training especially when noise in the context update σ_v is large.

5. Why do animals need shaping?

In psychology and neuroscience, there has been a longstanding appreciation of the importance of *behaviour shaping* for animal learning (Pavlov & Anrep, 1927; Skinner, 2019). In behaviour shaping, an animal learns multiple components of a complex task incrementally. It is colloquially known that shaping is critical for animals to learn complex tasks: different shaping strategies lead to different training times and sometimes animals completely fail to learn a task without shaping.

In our model, the primitives pre-training curriculum is reminiscent of shaping procedures. During the primitives pre-training phase, the student learns sub-components independently before combining them in the right context to solve the compositional task. Our analysis of the learning dynamics shows that the learning time for curriculum learning is significantly shorter than vanilla learning when the primi-

tive task complexity (T) is high and the initial and target composition context alignment V^0V^* is poor. This observation addresses several interesting findings from psychology and neuroscience. Many works in human psychology report feature-level biases on which features of objects or tasks draw attention and thus are easier to learn (Landau et al., 1988; Diesendruck & Bloom, 2003; Hudson Kam & Newport, 2005). In rodent studies, similar notions are known as attentional bias or rule bias (Smith et al., 2012; Vermaercke et al., 2014; Goltstein et al., 2021). These studies indicate that finding the right combination of importance or attention (context) poses a great challenge for learning the composition of multiple tasks, especially when it is largely misaligned to the initial bias or the search space is large. In this context, our analysis stresses that shaping or curriculum learning helps acquire the relevant primitives independently from the initial bias or attention toward certain features. Studies in humans also found that primitives pre-training can dramatically affect generalisation, allowing subjects to generalise after few trials under the right curriculum protocol (Dekker et al., 2022; Eckstein & Collins, 2021). Furthermore, our finding and analysis of the *multiplicative effect* implies that prior knowledge is important to further learning dynamics and behaviour. This aspect of learning in humans and animals has been observed in many works (Vergnaud, 1994; Block, 1982; Akrami et al., 2018). Although in our curriculum learning setting the training order among the primitives does not matter because all primitives are independent and orthogonal, in more realistic settings where the sub-tasks are correlated, our results suggest that the order of pre-training will impact the overall result of the shaping procedure.

6. Discussion

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 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.

While the scope of this study is limited to a particular type of compositionality, namely the linear combination of sub-tasks, many complex tasks can be described with different kinds of compositionality including sequential chaining of sub-tasks or recursive composition in tree-like task structures. The analytical study of learning dynamics in different kinds of compositional structure is an important avenue for future work. Our study further assumes strict modularisation of the primitives. By relaxing this constraint, we might observe more diverse behaviours of the model. One interesting future direction is to introduce a notion of task similarity either in terms of the primitives rules (Lee et al., 2021; 2022) or their input data (Mannelli et al., 2022). In our current model, the primitives are independently modularised and orthogonal to each other. We speculate that the benefits of curriculum learning that our current model shows is a lower bound on the actual potential effect. For example, correlation between the primitives might ease pre-training since one primitive might be able to leverage already acquired knowledge from a correlated primitive. Furthermore, non-modularized primitives and a compositional task built on them can possibly find alternative solutions during vanilla learning, which might harm compositional generalization performance on a series of different composition contexts. A broader study on different configurations of compositional tasks and primitives is an exciting future direction.

Acknowledgement

We thank Athena Akrami, Ann Chunyu Duan, Maria Eckstein, Kishore Kuchibhotla and Nishil Patel for useful discussions. SSM acknowledges Bocconi University for hospitality during the final stage 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).

Impact Statement

This paper focuses on theoretical research, and currently, we do not envision direct societal impact emerging from its conclusions.

References

Advani, M. S., Saxe, A. M., and Sompolinsky, H. High-dimensional dynamics of generalization error in neural networks. *Neural Networks*, 132:428–446, 2020.

Aguillon-Rodriguez, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., Cazettes, F., Chapuis, G., Churchland, A. K., Dan, Y., et al. Standardized and reproducible measurement of decision-making in mice. *Elife*, 10, 2021.

Ahissar, M. and Hochstein, S. Task difficulty and the specificity of perceptual learning. *Nature*, 387(6631):401–406, 1997.

Akrami, A., Kopec, C. D., Diamond, M. E., and Brody,

C. D. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, 554(7692): 368–372, 2018.

Andreas, J., Klein, D., and Levine, S. Modular multitask reinforcement learning with policy sketches. In *International conference on machine learning*, pp. 166–175. PMLR, 2017.

Bacon, P.-L., Harb, J., and Precup, D. The option-critic architecture. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Bahri, Y., Kadmon, J., Pennington, J., Schoenholz, S. S., Sohl-Dickstein, J., and Ganguli, S. Statistical mechanics of deep learning. *Annual Review of Condensed Matter Physics*, 11:501–528, 2020.

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., and Mordatch, I. Emergent tool use from multi-agent autocurricula. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=SkxpxJBKwS.

Barreto, A., Hou, S., Borsa, D., Silver, D., and Precup, D. Fast reinforcement learning with generalized policy updates. *Proceedings of the National Academy of Sciences*, 117(48):30079–30087, 2020.

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.

Biehl, M. and Schwarze, H. Learning by on-line gradient descent. *Journal of Physics A: Mathematical and general*, 28(3):643, 1995.

Block, J. Assimilation, accommodation, and the dynamics of personality development. *Child development*, pp. 281–295, 1982.

Bordelon, B., Masset, P., Kuo, H., and Pehlevan, C. Loss dynamics of temporal difference reinforcement learning. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Chomsky, N. *Aspects of the Theory of Syntax*. Number 11. MIT press, 2014.

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

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

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

Devin, C., Gupta, A., Darrell, T., Abbeel, P., and Levine, S. Learning modular neural network policies for multi-task and multi-robot transfer. In *2017 IEEE international conference on robotics and automation (ICRA)*, pp. 2169–2176. IEEE, 2017.

Diesendruck, G. and Bloom, P. How specific is the shape bias? *Child development*, 74(1):168–178, 2003.

Driscoll, L., Shenoy, K., and Sussillo, D. Flexible multi-task computation in recurrent networks utilizes shared dynamical motifs. *bioRxiv*, pp. 2022–08, 2022.

Eckstein, M. K. and Collins, A. G. How the mind creates structure: Hierarchical learning of action sequences. In *CogSci... Annual Conference of the Cognitive Science Society. Cognitive Science Society (US). Conference*, volume 43, pp. 618. NIH Public Access, 2021.

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

Engel, A. *Statistical mechanics of learning*. Cambridge University Press, 2001.

Gardner, E. and Derrida, B. Three unfinished works on the optimal storage capacity of networks. *Journal of Physics A: Mathematical and General*, 22(12):1983, 1989.

Gerace, F., Saglietti, L., Mannelli, S. S., Saxe, A., and Zdeborová, L. Probing transfer learning with a model of synthetic correlated datasets. *Machine Learning: Science and Technology*, 3(1):015030, 2022.

Goldt, S., Advani, M., Saxe, A. M., Krzakala, F., and Zdeborová, L. Dynamics of stochastic gradient descent for two-layer neural networks in the teacher-student setup. *Advances in neural information processing systems*, 32, 2019.

Goltstein, P. M., Reinert, S., Bonhoeffer, T., and Hübener, M. Mouse visual cortex areas represent perceptual and semantic features of learned visual categories. *Nature neuroscience*, 24(10):1441–1451, 2021.

Goyal, A., Lamb, A., Hoffmann, J., Sodhani, S., Levine, S., Bengio, Y., and Schölkopf, B. Recurrent independent mechanisms. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=mLcmdlEUxy-.

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

Hocker, D. L., Constantinople, C. M., and Savin, C. Curriculum learning inspired by behavioral shaping trains neural networks to adopt animal-like decision making strategies. *bioRxiv*, pp. 2024–01, 2024.

Hornsby, A. N. and Love, B. C. Improved classification of mammograms following idealized training. *Journal of Applied Research in Memory and Cognition*, 3(2):72–76, 2014.

Hudson Kam, C. L. and Newport, E. L. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*, 1(2):151–195, 2005.

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

Ito, T., Klinger, T., Schultz, D., Murray, J., Cole, M., and Rigotti, M. Compositional generalization through abstract representations in human and artificial neural networks. *Advances in Neural Information Processing Systems*, 35:32225–32239, 2022.

Jarvis, D., Klein, R., Rosman, B., and Saxe, A. On the specialization of neural modules. ICLR, 2023.

Karpathy, A. and Van De Panne, M. Curriculum learning for motor skills. In *Advances in Artificial Intelligence: 25th Canadian Conference on Artificial Intelligence, Canadian AI 2012, Toronto, ON, Canada, May 28-30, 2012. Proceedings 25*, pp. 325–330. Springer, 2012.

Lake, B. M. Compositional generalization through meta sequence-to-sequence learning. *Advances in neural information processing systems*, 32, 2019.

Lake, B. M., Linzen, T., and Baroni, M. Human few-shot learning of compositional instructions. In Goel, A. K., Seifert, C. M., and Freksa, C. (eds.), *Proceedings of the 41th Annual Meeting of the Cognitive Science Society, CogSci 2019: Creativity + Cognition + Computation, Montreal, Canada, July 24-27, 2019*, pp. 611–617. cognitivesciencesociety.org, 2019. URL https://mindmodeling.org/cogsci2019/papers/0123/index.html.

Landau, B., Smith, L. B., and Jones, S. S. The importance of shape in early lexical learning. *Cognitive development*, 3(3):299–321, 1988.

Lee, S., Goldt, S., and Saxe, A. Continual learning in the teacher-student setup: Impact of task similarity. In *International Conference on Machine Learning*, pp. 6109–6119. PMLR, 2021.

Lee, S., Mannelli, S. S., Clopath, C., Goldt, S., and Saxe, A. Maslow's hammer in catastrophic forgetting: Node re-use vs. node activation. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S. (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 12455–12477. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/lee22g.html.

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

Mannelli, S. S., Krzakala, F., Urbani, P., and Zdeborova, L. Passed & spurious: Descent algorithms and local minima in spiked matrix-tensor models. In *international conference on machine learning*, pp. 4333–4342. PMLR, 2019.

Mannelli, S. S., Gerace, F., Rostamzadeh, N., and Saglietti, L. Unfair geometries: exactly solvable data model with fairness implications. *arXiv preprint arXiv:2205.15935*, 2022.

Mendez, J. A., van Seijen, H., and EATON, E. Modular lifelong reinforcement learning via neural composition. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=5XmLzdslFNN.

Mittal, S., Lamb, A., Goyal, A., Voleti, V., Shanahan, M., Lajoie, G., Mozer, M., and Bengio, Y. Learning to combine top-down and bottom-up signals in recurrent neural networks with attention over modules. In *International Conference on Machine Learning*, pp. 6972–6986. PMLR, 2020.

Montero, M. L., Ludwig, C. J., Costa, R. P., Malhotra, G., and Bowers, J. The role of disentanglement in generalisation. In *International Conference on Learning Representations*, 2020.

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

Nachum, O., Gu, S. S., Lee, H., and Levine, S. Data-efficient hierarchical reinforcement learning. *Advances in neural information processing systems*, 31, 2018.

Narvekar, S., Peng, B., Leonetti, M., Sinapov, J., Taylor, M. E., and Stone, P. Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, 21(181):1–50, 2020.

Okawa, M., Lubana, E. S., Dick, R. P., and Tanaka, H. Compositional abilities emerge multiplicatively: Exploring diffusion models on a synthetic task. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=frVo9MzRuU.

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

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

Pavlov, I. P. and Anrep, G. V. *Conditioned Reflexes. An Investigation of the Physiological Activity of the Cerebral Cortex... Translated and Edited by GV Anrep*. London, 1927.

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

Saglietti, L., Mannelli, S., and Saxe, A. An analytical theory of curriculum learning in teacher-student networks. *Advances in Neural Information Processing Systems*, 35: 21113–21127, 2022.

Saxe, A. M., Earle, A. C., and Rosman, B. Hierarchy through composition with multitask lmdps. In *International Conference on Machine Learning*, pp. 3017–3026. PMLR, 2017.

Schug, S., Kobayashi, S., Akram, Y., Wolczyk, M., Proca, A. M., Oswald, J. V., Pascanu, R., Sacramento, J., and Steger, A. Discovering modular solutions that generalize compositionally. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=H98CVcX1eh.

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

Seung, H. S., Sompolinsky, H., and Tishby, N. Statistical mechanics of learning from examples. *Physical review A*, 45(8):6056, 1992.

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

Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., Spiering, B., Beran, M. J., Church, B. A., Ashby, F. G., et al. Implicit and explicit categorization: A tale of four species. *Neuroscience & Biobehavioral Reviews*, 36(10):2355–2369, 2012.

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

Sutton, R. S., McAllester, D., Singh, S., and Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12, 1999a.

Sutton, R. S., Precup, D., and Singh, S. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2): 181–211, 1999b.

Tessler, C., Givony, S., Zahavy, T., Mankowitz, D., and Mannor, S. A deep hierarchical approach to lifelong learning in minecraft. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Tong, W. L., Iyer, A., Murthy, V. N., and Reddy, G. Adaptive algorithms for shaping behavior. *bioRxiv*, pp. 2023–12, 2023.

Vergnaud, G. Multiplicative conceptual field: what and why. *The development of multiplicative reasoning in the learning of mathematics*, pp. 41–59, 1994.

Vermaercke, B., Cop, E., Willems, S., D'Hooge, R., and Op de Beeck, H. P. More complex brains are not always better: rats outperform humans in implicit category-based generalization by implementing a similarity-based strategy. *Psychonomic bulletin & review*, 21:1080–1086, 2014.

Wang, X., Chen, Y., and Zhu, W. A survey on curriculum learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4555–4576, 2021.

Wiedemer, T., Mayilvahanan, P., Bethge, M., and Brendel, W. Compositional generalization from first principles. *Advances in Neural Information Processing Systems*, 36, 2024.

Wu, X., Dyer, E., and Neyshabur, B. When do curricula work? In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=tW4QEInpni.

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., and Wang, X.-J. Task representations in neural networks trained to perform many cognitive tasks. *Nature neuroscience*, 22(2):297–306, 2019.

Yang, R., Xu, H., Wu, Y., and Wang, X. Multi-task reinforcement learning with soft modularization. *Advances in Neural Information Processing Systems*, 33:4767–4777, 2020.

Zdeborová, L. and Krzakala, F. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

Zhang, Y., Abbeel, P., and Pinto, L. Automatic curriculum learning through value disagreement. *Advances in Neural Information Processing Systems*, 33:7648–7659, 2020.

A. Summary of RL perceptron and Derivation of the Learning Rules

The RL perceptron (Patel et al., 2023) is a particular instantiation of the teacher-student framework designed to capture the learning of RL agents. In this case, the labels provided by the teacher are interpreted as the correct choice that the student must make to receive a reward. More precisely, given an input x_t (randomly sampled vector) at time step t, the teacher network W^* will generate the correct binary choice $y_t^* = \text{sign}(W^* \cdot x_t/\sqrt{N})$. The student network W will receive the same input x_t and make its prediction $y_t = \text{sign}(W \cdot x_t/\sqrt{N})$. If all T decisions in the episode are correct, the student will receive a reward and update its weights W according to the REINFORCE policy gradient update rule (Sutton et al., 1999a). The equation for the update is given by Eq. 1. Patel et al.(2023) showed that this perceptron-like learning rule is like policy gradient learning and it is amenable to analytical characterisation.

In more detail, consider a probabilistic policy function π_W such that π_W samples a binary decision ($y_t = \pm 1$) based on input x_t at each time step $t(t = 1 \dots T)$ following the policy $\pi_W(y_t|x_t) = \frac{1}{1+\exp(-y_t(W \cdot x_t)/\sqrt{N})}$, where N is input dimension. The reward is only available when decisions over all time steps ($t = 1 \dots T$) are correctly given so the expected reward becomes $J = \sum^T \log(\pi_W(y_t|x_t))$. Following the update rule of Patel et al. (2023), we take the gradient of the expected reward J with respect to policy parameter W, and with large input dimension N, we get the learning rule Eq. 1.

Similarly, Eq. 2-3 are results of policy gradient learning on compositional policy. Instead of a single policy for one primitive, the compositional policy becomes $\pi_{W,V_{1\dots K}}(y_t|x_t) = \frac{1}{1+\exp(-y_t \sum_K V_k W_k \cdot x_k/\sqrt{N})}$ with K primitives. Replace this policy for the expected reward $J = \sum^T \log(\pi_W(y_t|x_t))$ and taking gradient with respect to the policy parameters W_k and V_k leads to the update rules in Eq. 2-3.

B. Derivations of ODEs

In the thermodynamic limit, the stochastic updates of the student's dynamics concentrate to a deterministic limit, leading to effective dynamics characterised by ODEs. In this section, we show how to derive Eqs. 7-9 described in the main text.

We define the alignment fields of teacher and student weights of each primitive task k as $\nu_k = \frac{W_k^* \cdot x_k}{\sqrt{N}}$ and $\lambda_k = \frac{W_k \cdot x_k}{\sqrt{N}}$, and the compositional alignment as follows,

$$\tilde{\nu} = \sum_{i=1}^{K} V_i^* \frac{W_i^* \cdot x_i}{\sqrt{N}} = \sum_{i=1}^{K} V_i^* \nu_i \tag{19}$$

$$\tilde{\lambda} = \sum_{i=1}^{K} V_i \frac{W_i \cdot x_i}{\sqrt{N}} = \sum_{i=1}^{K} V_i \lambda_i \tag{20}$$

notice that the primitive decision and compositional task decision can be expressed with the alignment fields: $y_k^* = \text{sign}(\nu_k), y_k = \text{sign}(\lambda_k), \tilde{y}^* = \text{sign}(\tilde{\nu})$ and $\tilde{y} = \text{sign}(\tilde{\lambda})$. Note that these quantities are time-dependent and we will add the subscript t to indicate the time step when necessary.

These fields are key stochastic variables of the problem, indeed the fundamental idea behind the ODE limit is to track their distribution at all times. It is easy to notice that they are centered Gaussian variables with covariance given by the order parameters (Eq. 4) defined in the paper.

Redefining the expectations in terms of these fields allows us to derive fundamental observables of the problems, for instance the average probability of success $\tilde{P}$ (i.e. $\tilde{y_t} = \tilde{y_t^*}$). This quantity is equivalent to the probability of having both, the ground truth and the student choices, lying on the same hypersector defined as $\tilde{\nu}$ and $\tilde{\lambda}$,

$$\tilde{P} = \left\langle 1 - \frac{\cos^{-1}(\tilde{\nu}\tilde{\lambda})}{\pi} \right\rangle. \tag{21}$$

Given these preliminary definitions, the goal is to derive the updates for Q_k, R_k, and V_k.

Useful equalities. The following analysis will need to evaluate some Gaussian averages. It is therefore convenient to compute some useful relations:

$$\langle \nu_k sign(\tilde{\nu}) \rangle = \sqrt{\frac{2}{\pi}} \frac{V_k^* S_k}{\sqrt{\sum_k^K V_k^{*2} S_k}}, \tag{22}$$

$$\langle \lambda_k sign(\tilde{\nu}) \rangle = \sqrt{\frac{2}{\pi}} \frac{V_k^* R_k}{\sqrt{\sum_k^K V_k^{*2} S_k}}, \tag{23}$$

$$\langle \lambda_k sign(\tilde{\lambda}) \rangle = \sqrt{\frac{2}{\pi}} \frac{V_k Q_k}{\sqrt{\sum_k^K V_k^2 Q_k}}. \tag{24}$$

Notice that the correct decision, $sign\tilde{\nu} = sign\tilde{\lambda}$, is equivalent to $\theta(\tilde{\nu}\tilde{\lambda})$ with θ the Heaviside function. It is useful to notice that following relations hold

$$sign(\tilde{\lambda})\theta(\tilde{\nu}\tilde{\lambda}) = sign(\tilde{\nu})\theta(\tilde{\nu}\tilde{\lambda}) = \frac{1}{2}(sign(\tilde{\lambda}) + sign(\tilde{\nu})). \tag{25}$$

The average of these equalities can be evaluated in different ways, an easy approach is to evaluate the two fields independently by conditioning and using the properties of Guassian variables.

ODEs. Starting from the definition of the gradient updates Eqs. 1-3, it has been shown that the updates for the order parameters concentrate to their typical value as $N \to \infty$ (Goldt et al., 2019). Therefore, to derive dynamics of V_k, we evaluate the average Eq. 3 over the random fields, we use the angular brackets to denote the average over all the random variables.

$$\begin{aligned} \frac{dV_k}{dz} &= \eta_v \left\langle \frac{1}{T} \sum_{i,t=1}^{K,T} \tilde{y}_t \frac{W_i \cdot x_{i,t}}{\sqrt{N}} I(\Phi) \right\rangle \\ &= \eta_v \left\langle \lambda_k sign(\tilde{\lambda}) \theta(\tilde{\nu}\tilde{\lambda}) \right\rangle \tilde{P}^{T-1} \qquad (26) \\ &= \frac{\eta_v}{\sqrt{2\pi}} \left(\frac{V_k Q_k}{\sqrt{\sum_{i=1}^K V_i^2 Q_i}} + \frac{V_k^* R_k}{\sqrt{\sum_{i=1}^K V_i^{*2} S_i}} \right) \tilde{P}^{T-1}. \qquad (27) \end{aligned}$$

From the definition of R_k in Eq.4, we take the dot product by W_k^*/N on both sides of Eq. 2 and average at the mth update.

$$R_k^{m+1} - R_k^m = \frac{\eta_w}{N} \left\langle \frac{1}{T} \sum_{t=1}^{T} V_k sign(\tilde{\lambda}) \frac{W_k^* \cdot x_{t,k}}{\sqrt{N}} \right\rangle \tag{28}$$

finally in the thermodynamic limit

$$\begin{aligned} \frac{dR_k}{dz} &= \eta_w \left\langle \frac{1}{T} \sum_{t=1}^{T} V_k sign(\tilde{\lambda}) \nu_k I(\Phi) \right\rangle = \frac{\eta_w}{2} V_k \left\langle \nu_k sign(\tilde{\lambda}) + \nu_k sign(\tilde{\nu}) \right\rangle \tilde{P}^{T-1} \\ &= \frac{\eta_w}{\sqrt{2\pi}} \left(\frac{V_k^2 R_k}{\sqrt{\sum_{i=1}^K V_i^2 Q_i}} + \frac{V_k V_k^* S_k}{\sqrt{\sum_{i=1}^K V_i^{*2} S_i}} \right) \tilde{P}^{T-1}. \qquad (29) \end{aligned}$$

Since $Q_k = W_k \cdot W_k/N$, we take the square of Eq. 2 and, after taking the thermodynamic limit, we obtain

$$\frac{dQ_k}{dz} = 2\eta_w \left\langle \frac{1}{T} \sum_{t=1}^{T} V_k sign(\tilde{\lambda}) \lambda_k \right\rangle \tilde{P}^{T-1} + \frac{\eta_w^2}{T^2} \left\langle \sum_{t,t'}^{T,T} V_k V_k \tilde{y}_t \tilde{y}_{t'} x_t^\intercal x_{t'} \right\rangle \tilde{P}^T.$$

$$= 2\eta_w \left\langle \frac{1}{T}\sum_{t=1}^{T} V_k sign(\tilde{\lambda})\lambda_k \right\rangle \tilde{P}^{T-1} + \frac{\eta_w^2}{T} V_k^2 \tilde{P}^T$$
$$= \frac{2\eta_w}{\sqrt{2\pi}} \left(\frac{V_k^2 Q_k}{\sqrt{\sum_{i=1}^{K} V_i^2 Q_i}} + \frac{V_k V_k^* R_k}{\sqrt{\sum_{i=1}^{K} V_i^{*2} S_i}} \right) \tilde{P}^{T-1} + \frac{\eta_w^2}{T} V_k^2 \tilde{P}^T.$$

where the last steps follows from the fact the correlation between x and the fields becomes negligible to the leading order and we can average them following the Gaussianity.

Finally, for the *spherical case* (ie. $||W_k|| = 1\ \forall k$), we obtain the following expression that can be derived by using a Lagrange multiplier to enforce the spherical constraint

$$\frac{dR_k}{dz} = \frac{\eta_w}{\sqrt{2\pi}} V_k V_k^* (1 - R_k^2)(1 - \cos^{-1}(\tilde{R}))^{T-1} + \frac{\eta_w^2}{T} V_k^2 (1 - \cos^{-1}(\tilde{R}))^T. \tag{30}$$

C. Asymptotic Analysis of Learning Dynamics

C.1. Primitives learning

From the ODE of a single RL perceptron (Patel et al., 2023)

$$\frac{dR_k}{dz} = \frac{\eta_w}{\sqrt{2\pi}}(1 - R_k^2)\left(1 - \frac{1}{\pi}\cos^{-1}(R_k)\right)^{T-1} - \frac{\eta_w^2}{2T} R_k \left(1 - \frac{1}{\pi}\cos^{-1}(R_k)\right)^T, \tag{31}$$

we approximate the function around the point $R_k \approx 0$ to capture the early dynamics and get

$$\frac{dR_k}{dz} \approx \frac{\eta_w}{\sqrt{2\pi}}\left(1 - \frac{1}{\pi}\cos^{-1}(R_k)\right)^{T-1} \approx \frac{\eta_w}{\sqrt{2\pi}}\left(\frac{1}{2} + \frac{R_k}{\pi}\right)^{T-1}.$$

Solving the ODE above, we get a closed form expression on R_k dynamics as follows,

$$R_k = -\frac{\pi}{2} + \pi\left(\frac{\eta_w}{\sqrt{2\pi}\pi}(2 - T)z + 2^{(T-2)}\right)^{\frac{1}{2-T}}. \tag{32}$$

C.2. Compositional Generalisation after pre-training

Since we consider a case where $|V| = 1$, we use a Lagrange multiplier to impose the constraint on our V_k update

$$\frac{dV_k}{dz} = g_k - \lambda V_k \tag{33}$$
$$\sum_k V_k g_k - \lambda = 0 \Rightarrow \lambda = \sum_k V_k g_k. \tag{34}$$

Plugging Eq. 34 into Eq. 33 and rearranging the equation for g_k:

$$g_k = \frac{\eta_v}{\sqrt{2\pi}}\Big(V_k + V_k^* R_k\Big)\Big(1 - \frac{1}{\pi}\cos^{-1}(\tilde{R})\Big)^{T-1}. \tag{35}$$

We obtain V_k update with the constraint satisfied by plugging in Eq. 35 into Eq. 33 and $\tilde{R}$, and by multiplying $R_k V_k^*$ on both sides:

$$\frac{dV_k}{dz} = \frac{\eta_v}{\sqrt{2\pi}}\Big(V_k + V_k^* R_k\Big)\Big(1 - \frac{1}{\pi}\cos^{-1}(\tilde{R})\Big)^{T-1} - \frac{\eta_v}{\sqrt{2\pi}} V_k \Big(1 + \tilde{R}\Big)\Big(1 - \frac{1}{\pi}\cos^{-1}(\tilde{R})\Big)^{T-1} \tag{36}$$
$$\frac{d\tilde{R}}{dz} = \frac{\eta_v}{\sqrt{2\pi}}\Big(1 - \frac{1}{\pi}\cos^{-1}(\tilde{R})\Big)^{T-1}\Big(\sum_k V_k^2 R_k^2 - \tilde{R}^2\Big). \tag{37}$$

After the pre-training of all primitives, we can assume $R_k \approx R_0$ to be a constant, where R_0 is a maximum R which can be achieved with the episode length T. This allows us to have the initial value of $\tilde{R}_0 \approx \sum_k V_k^* V_k R_0$ at the beginning of compositional generalisation $z_{compositional} = 0$. We linearize Eq. 37 around $\tilde{R} = \tilde{R}_0$ treating V_k as constant and get

$$\frac{d\tilde{R}}{dz} = \frac{\eta_v}{\sqrt{2\pi}} \tilde{P}_0^{T-1} \Big(1 + \frac{\tilde{R}_0(T-1)}{\tilde{P}_0 \pi \sqrt{1-\tilde{R}_0^2}}\Big) \Big(R_0^2 + \tilde{R}_0{}^2 - 2\tilde{R}_0 \tilde{R}\Big) \Big(1 - (\frac{\tilde{P}_0 \pi \sqrt{1-\tilde{R}_0^2}}{(T-1)} + \tilde{R}_0)^{-1} \tilde{R}\Big). \tag{38}$$

For convenience, we define the below variables

$$\alpha = \frac{(T-1)}{\tilde{P}_0 \pi \sqrt{1-\tilde{R}_0^2}} \tag{39}$$

$$a = -\frac{2\tilde{R}_0 \eta_v}{\sqrt{2\pi}} \tilde{P}_0^{T-1} (1 - \alpha \tilde{R}_0) \tag{40}$$

$$b = -\frac{1}{2\tilde{R}_0} \big(R_0^2 + \tilde{R}_0^2\big) \tag{41}$$

$$c = \alpha (1 - \alpha \tilde{R}_0)^{-1}. \tag{42}$$

With the above variables, we rearrange Eq. 38

$$\frac{d\tilde{R}}{dz} = a(b + \tilde{R})(1 + c\tilde{R}). \tag{43}$$

Solving this, we get approximated closed form solution of $\tilde{R}(z')$ after primitive pre-training where $z' = z - z_{\text{pre-training}}$,

$$\tilde{R}(z') = \frac{Ae^{a(1-bc)z'} - b}{1 - Ace^{a(1-bc)z'}}, \quad A = \frac{\tilde{R}_0 + b}{1 + c\tilde{R}_0}. \tag{44}$$

Since we are looking for a time z' where V and V^* gets aligned ($V \cdot V^* = 1$) and $\tilde{R}$ recovers possible maximum value,, we can solve $\tilde{R}(z') = R_0$ and get

$$z'_{(\tilde{R}=R_0)} = \frac{\sqrt{2\pi}}{\eta_v} \tilde{P}_0^{1-T} \Big(\alpha(\tilde{R}_0^2 - R_0^2) - 2\tilde{R}_0\Big)^{-1} \log(\frac{R_0 + b}{A(cR_0 + 1)}). \tag{45}$$

We can characterise the critical timescale of compositional learning phase after primitives pre-training $z'_{(\tilde{R}=R_0)}$ with the dominant factor $\tilde{P}_0^{2-T}$ with respect to the task difficulty T.

C.3. Vanilla Learning Time

We approximate the dynamics of R_k by considering $R_{k' \neq k}$ components to account for the multiplicative learning effect. We modify Eq. 15 by adding the contribution of the other primitives $k \neq k'$ as follows for R_1 and R_2

$$\frac{dR_1}{dz} \approx \frac{\eta_w}{\sqrt{2\pi}} V_1^0 V_1^* \Big(\frac{1}{2} + \frac{V_1^0 V_1^* R_1 + V_2^0 V_2^* R_2}{\pi}\Big)^{T-1} \tag{46}$$

$$\frac{dR_2}{dz} \approx \frac{\eta_w}{\sqrt{2\pi}} V_2^0 V_2^* \Big(\frac{1}{2} + \frac{V_2^0 V_2^* R_2 + V_1^0 V_1^* R_1}{\pi}\Big)^{T-1}. \tag{47}$$

Finally, we get

$$R_1(z) = \frac{1}{A}\Big(-\frac{\pi}{2} + \pi\Big(\frac{\eta_w}{\sqrt{2\pi}\pi}(2-T)(V_1^0 V_1^*)Az + 2^{T-2}\Big)^{\frac{1}{2-T}}\Big), \tag{48}$$

$$R_2(z) = \frac{1}{B}\Big(-\frac{\pi}{2} + \pi\Big(\frac{\eta_w}{\sqrt{2\pi}\pi}(2-T)(V_2^0V_2^*)Bz + 2^{T-2}\Big)^{\frac{1}{2-T}}\Big), \tag{49}$$

with $A = V_1^0V_1^* + CV_2^0V_2^*$, $B = \frac{1}{C}V_1^0V_1^* + V_2^0V_2^*$, and $C \approx \frac{R_2}{R_1}$. We can use a binomial approximation on Eqs. 48-49 and divide them to get C, which lead us to $R_2/R_1 \approx V_2^0V_2^*/V_1^0V_1^*$. We plug this into Eq. 49 and get the asymptotic critical time for the learning scale of the second primitive considering the multiplicative effect from the first primitive as following:

$$\tau_{\text{vanilla},k=2} = \frac{\sqrt{2\pi}\pi}{\eta_w(T-2)}2^{(T-2)}\frac{1}{(V_1^0V_1^*)^2 + (V_2^0V_2^*)^2}. \tag{50}$$

D. Additional Experiments

In section 4.2 and Figure 4, we showed that for $K = 2$ case where the learning rates $\eta_{w,\text{pretrain}}, \eta_{w,\text{compositional}}$ and η_v are all fixed to 1. Here we extended our experiments with a hyperparameter search on learning rates in each of the training protocols and $K = 4$ case to make the curriculum learning effect on learning speed concrete, shown in Figure 6.

In addition to Figure 5b with $V^0V^* = 0.25$ case, we show the results for different V^0V^* values, shown in Figure 7.

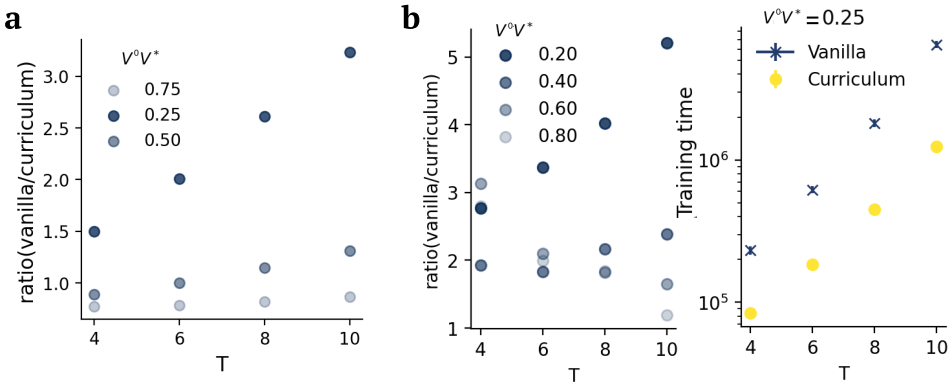

Figure 6. **a)** The ratio between vanilla learning and curriculum learning in $K = 2$ after a hyperparameter search on the learning rates in each of the training protocols. We search over the range [0.01, 10] in logarithmic scale for $\eta_{w,\text{pretrain}}, \eta_{w,\text{composition}}$ and η_v individually. **b)** The speed boost effect of the curriculum learning in $K = 4$.

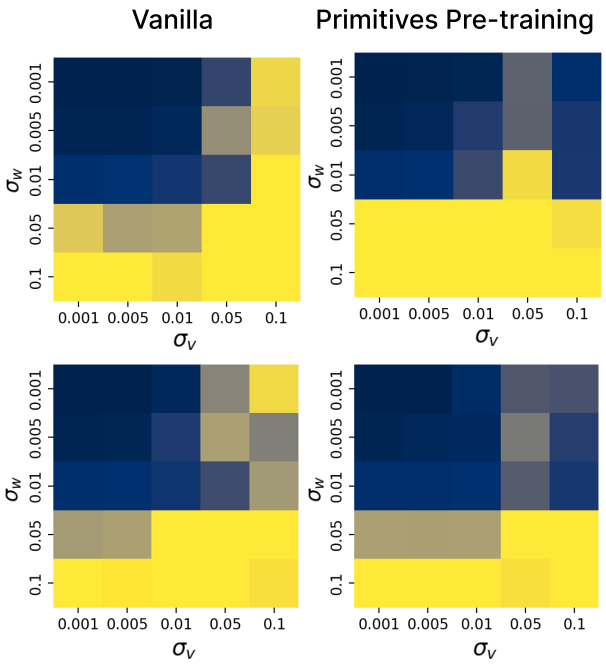

Figure 7. Results from different V^0V^* for Figure. 5b experiment. Top: $K = 2$, $V^0V^* = 0.5$, Bottom: $K = 2$, $V^0V^* = 0.75$