Environmental Statistics Shape Learning Dynamics in Meta-Reinforcement Learning

Monami Nishio[1*], Zhenglong Zhou[2*], Anna Schapiro[1], Allyson Mackey[1]
[1]University of Pennsylvania, Philadelphia, US.
[2]New York University, New York, US.
*Co-first authors

Abstract

Adverse environments have been proposed to influence multiple aspects of cognition, including learning and adaptation. However, causal tests of long-term effects of adversity on human cognition are extremely difficult. Moreover, different types of adversity often co-occur in real life, making it challenging to disentangle their individual effects. Computational models provide a way to independently manipulate specific environmental statistics and examine their causal impact on learning. Here, we integrate meta-learning with reinforcement learning to investigate the mechanisms by which environmental statistics shape learning processes. Specifically, we examine how two core environmental dimensions of adversity, volatility and controllability, shape trajectories of learning rates in deep neural networks. We find that both high volatility and low controllability reduce the maximum learning rate, but only controllability affects the initial slope of learning, leading to a flattening of early learning dynamics. Notably, deeper layers of the model are more strongly influenced by environmental factors than earlier layers, leading to reduced hierarchical differentiation in learning rates under conditions of high volatility and low controllability. These results suggest that a meta–reinforcement learning framework may provide a simplified yet useful approach for gaining mechanistic insight into how distinct types of adverse environments affect learning.

Keywords: meta-learning, reinforcement learning; learning rate; adversity; volatility; controllability.

Introduction

Early-life adversity is one of the strongest predictors of well-being across the lifespan (Draper & Yousafzai, 2024). Exposure to adversity has been associated with a range of learning strategies and outcomes, including risk-taking (Courson et al., 2025) and explore–exploit tradeoffs (Humphreys, et al., 2015; Lloyd et al., 2022; Xu et al., 2023). These differences in learning, in turn, contribute to long-term educational and economic disparities (Rakesh et al., 2025). However, much of this work is correlational, as it is not possible to causally manipulate specific aspects of children's early environments. Moreover, although different dimensions of childhood adversity show distinct associations with outcomes (Sheridan et al., 2017), multiple forms of adversity often co-occur in real life, making it difficult to disentangle their individual effects and timing in a clean way.

Two environmental statistics that have been particularly well-studied in the context of adversity are controllability and volatility (Piray & Daw, 2020, Nussenbaum & Hartley, 2024). Broadly, controllability is low in environments where stressful events occur frequently and individual actions do not reliably change outcomes (Ligneul, 2021). In contrast, volatility is high in environments where the statistical structure of the world changes unpredictably. In a parenting context, for example, low controllability corresponds to situations in which a child's actions rarely influence parental responses, whereas high volatility corresponds to situations in which parental responses vary unpredictably across similar contexts. Although the ability to accurately assess environmental controllability improves over development, controllability already shapes learning in childhood, with children using it to guide exploration and behavior (Raab et al., 2022).

Computational modeling of human behavior offers an opportunity to manipulate the learning environment and causally test its impacts on learning (Collins et al., 2022). Models also provide an opportunity to develop specific mechanistic accounts of these behavioral effects. Previous computational work suggests that controllability and volatility differentially modulate the speed of learning, often referred to as the learning rate (Frankenhuis & Gopnik, 2023). Adults increase their learning rates in environments with high volatility (Nassar et al., 2010; McGuire et al., 2014), but decrease their learning rates in environments with low controllability (Dorfman & Gershman, 2019; Csifcsák et al., 2019).

Another approach to causally examine the effects of environment is through computational simulations using deep neural networks. Although such simulations necessarily rely on simplified abstractions, they provide a powerful way to study long-term environmental effects on learning systems—for example, how learning rates evolve over extended experience, which is not feasible in experimental studies of human behavior. Furthermore, by studying the effects of the environment on distinct components of the model, we can gain insights into how environmental factors differentially influence different parts of a learning system. Previous work suggests that deep feedforward neural network can offer a rough correspondence between model layers and brain regions, with earlier layers mapping onto primary sensory areas and deeper layers corresponding to higher-order regions across cortical hierarchies (Yamins & DiCarlo, 2016). Recent studies have further shown that deep neural network simulations using meta-learning can produce a cascading pattern of learning-rate development across

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

layers, in which earlier layers plateau sooner than later layers in image classification tasks (Zhou & Schapiro, 2025). This pattern is consistent with findings in the human brain, where sensory cortices mature earlier than higher-order association regions (Sydnor et al., 2023), further supporting the correspondence between model layers and cortical hierarchies. Because environmental effects unfold through continuous interaction with the environment, we adopt a meta–reinforcement learning framework in which agents act, receive feedback, and adapt their learning processes over time (Nussenbaum & Hartley, 2024).

In this study, we present a proof of concept using the meta-reinforcement learning framework to investigate how volatility and controllability influence both overall performance (reward accumulation) and learning dynamics (learning rates) across different layers. This approach provides a simplified yet potentially useful framework for exploring how distinct environmental structures may shape learning dynamics across different components of a learning system.

Methods

Model

We employed a meta-reinforcement learning (meta-RL) model based on Advantage Actor–Critic (A2C) augmented with Meta-Stochastic Gradient Descent (Meta-SGD), implemented using the Learn2Learn library (Arnold et al., 2020). The agent learned a policy $\pi_\theta(a \mid s)$ that maps visual observations of the environment s to discrete actions a, where θ denotes the policy parameters optimized via meta-learning. Observations consisted solely of the RGB image representation of Minigrid environment (Chevalier-Boisvert et al., 2018). The original observation was flattened into a vector $s \in R^d$.

To parameterize the policy, we defined an Actor network as a feedforward neural network with two hidden layers of 100 units each and ReLU nonlinearities, followed by a linear output layer with 4 units corresponding to the four available actions (move up, down, left, and right; Figure 1A). The weights of the network layers are denoted as w_1, w_2, and w_O, corresponding to the first hidden layer, the second hidden layer, and the output layer, respectively. The policy defines a categorical distribution:

$$\pi_\theta(a|s) = Categorical(softmax(f_\theta(s)))$$

where $f_\theta(\cdot)$ denotes the neural network mapping from state to action logits. Actions were sampled stochastically from this distribution during training.

In addition to the Actor, a Critic network was used to estimate the value of each state. The critic consisted of a single linear layer mapping the state representation directly to a scalar output representing the state-value function $V_\phi(s)$.

Value Function and Advantage Estimation

A linear value function $V_\phi(s)$ was used as a baseline to reduce variance in policy gradient estimation. Advantages were computed using generalized advantage estimation (GAE):

$$\hat{A}_t = \sum_{l=0}^{\infty} (\gamma\tau)^l \delta_{t+l}, \delta_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t),$$

with discount factor $\gamma = 0.99$ and trace decay parameter $\tau = 0.95$. The A2C policy loss for a trajectory was defined as:

$$L_{policy} = -E_t[log\ \pi_\theta(a_t \mid s_t)\hat{A}_t].$$

Meta-learning with Meta-SGD

Meta-learning optimizes a set of meta-parameters consisting of (i) a shared initialization of the Actor parameters, denoted by θ, and (ii) a vector of parameter-specific learning rates α, defined for all Actor parameters, including weights, biases, and σ. For each task T_i, fast adaptation (the inner loop) starts from the shared initialization θ and produces task-specific

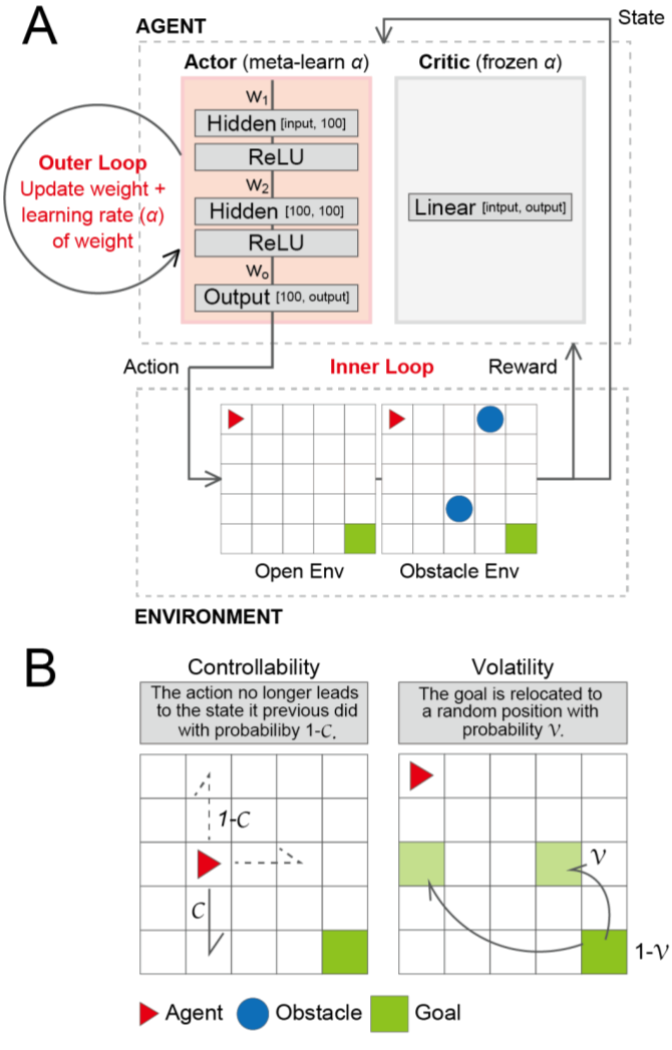

Figure 1. Model architecture and task modulation. (A) Structure of the meta–reinforcement learning model. We used two task contexts: *Open* and *Obstacle* environments. **(B)** Schematic of controllability and volatility: ***C*** indicates controllability, and ***V*** indicates volatility.

adapted parameters θ_i' by performing a gradient-based update on task-specific training trajectories:

$$\theta_i' = \theta - \alpha \odot \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(\theta),$$

where $\odot$ denotes element-wise multiplication and $\mathcal{L}_{T_i}$ is the loss function for task T_i. The adapted parameters θ_i' are used only within task T_i and are not shared across tasks.

In the outer loop, the meta-parameters θ and α are updated using gradients computed from post-adaptation validation trajectories, evaluated with the adapted parameters θ_i' across a batch of tasks. Importantly, the outer loop does not directly update θ_i'; instead, it updates the shared initialization θ and the learning rates α so as to improve the effectiveness of future inner-loop adaptations.

The Critic network is trained using standard A2C updates with a fixed learning rate, ensuring that meta-learning is restricted to the Actor. This design allows the meta-learned learning rates α to be interpreted as a proxy for plasticity in the Actor's perceptual and action-selection processes.

Task Paradigm

Agents were trained and evaluated in two types of Minigrid environments (Figure 1A). In the open environment (Open Env), the grid world contained no obstacles and rewards were sparse. The agent's objective was to navigate to a goal, receiving a reward of $+1$ upon success and zero otherwise. Episodes terminated upon reaching the goal or exceeding a maximum step limit proportional to the grid size 5×5. The agent's starting position changes in each trial. In the obstacle environment (Obstacle Env), static internal walls were added, requiring the agent to navigate around obstacles to reach the goal. For each experimental condition, the meta-learner was trained only on task instances sampled from a single environment type (Open Env or Obstacle Env). Within each type, task instances varied in layout (e.g., start positions or obstacle configurations) but shared the same observation format and discrete action space, allowing a single policy architecture to be used across tasks. Each meta-training iteration consisted of an inner-loop adaptation phase on a sampled task instance, followed by an outer-loop update based on the adapted policy's performance on validation trajectories from that same task instance.

Task Modulation: Volatility and Controllability

To model environmental dimensions associated with adversity, we parametrically manipulated volatility and controllability within the environments (Figure 1B).

Volatility. Volatility governed the stability of the goal location over time. At each time step, with probability v, the goal was removed and re-placed at a random valid location on the grid:

$$\Pr(\text{goal relocation at } t) = v,$$

where $v \in [0,1]$ denotes volatility. High volatility produces unstable contingencies, as the correct goal can change unpredictably even within an episode, whereas low volatility yields a relatively stable goal location across time.

Controllability. Controllability determined the degree to which the agent's intended action influenced the executed action. On each time step, the executed action $\tilde{a}_t$ followed:

$$\tilde{a}_t = \begin{cases} a_t, & \text{with probability } c \\ \text{Uniform}(\mathcal{A}), & \text{with probability } 1-c \end{cases}$$

where $c \in [0,1]$ denotes controllability. When $c = 1$, the environment was fully controllable; when $c = 0$, actions were entirely random and independent of the agent's choice.

Statistical Analysis

For the learning rate, we averaged across all nodes within each layer to obtain a layer-wise estimate. We ran the model with five different random seeds (i.e., different initializations of the random number generator) and calculated the mean and standard deviation across seeds for visualization. For each variable of interest, we performed an ANOVA on the seed-level values across conditions, in addition to post-hoc Tukey tests for pairwise comparisons.

We examined reward accumulation trajectories and the maximum reward the model can achieve as measures of learning performance. The reward accumulation trajectory refers to the time course of cumulative reward obtained by the agent across training or interaction steps. At each time point, it reflects the total reward collected up to that step, providing a dynamic measure of how efficiently the agent learns and improves its performance across different task conditions. To better understand learning dynamics, we also analyzed the maximum learning rate as well as the initial slope of the learning curve. The initial slope was defined as the slope of the value over the first 50 iterations. This metric captures the early speed of learning and complements the maximum performance achieved over time.

Results

Effects of Volatility and Controllability on Reward Accumulation

First, as a measure of learning performance, we examined how reward accumulation trajectories and the maximum reward attainable by the model varied across task paradigms and environmental conditions. In the Open environment, increasing volatility led to monotonic decreases of the maximum reward attainable (Figure 2A; ANOVA $p < 0.001$). Decreasing controllability also produced a graded reduction in the maximum reward, with a stronger effect than that of volatility (Figure 2B; ANOVA $p < 0.001$). This hierarchical decline in maximum reward was consistent across different combinations of volatility and controllability (Figure 2C). Similarly, in the Obstacle environment, both increasing volatility and decreasing controllability resulted in graded reductions in maximum reward (Figure 2D,E; ANOVAs $p < 0.001$). Overall, the maximum reward attainable in the Obstacle environment was lower than in the open environment (Figure 2F).

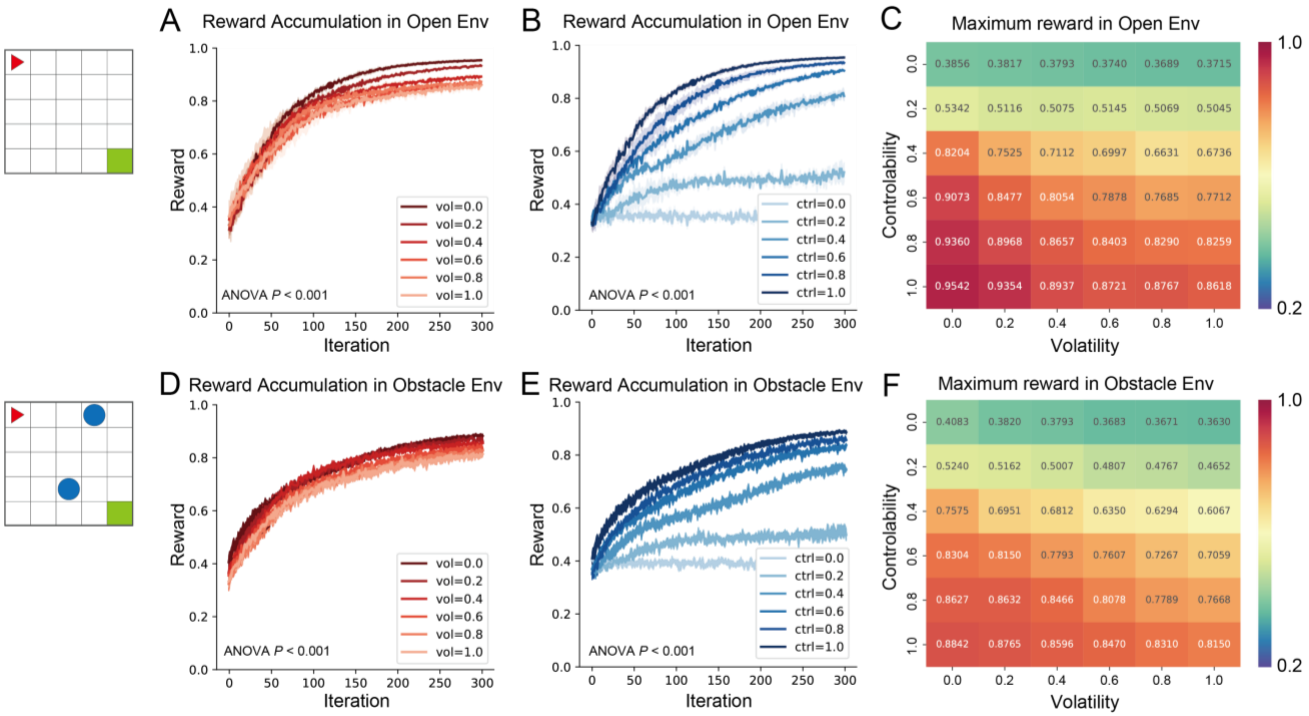

Figure 2. Reward accumulation trajectories across varying levels of task complexity and environmental conditions. (A, B) Reward accumulation trajectories in the Open environment under varying levels of **(A)** volatility and **(B)** controllability. **(C)** Maximum reward attained by the model for each combination of volatility and controllability in the open environment. **(D, E)** Reward accumulation trajectories in the Obstacle environment under varying levels of **(D)** volatility and **(E)** controllability. **(F)** Maximum reward attained by the model for each combination of volatility and controllability in the obstacle environment.

Attenuation of Layer-wise Hierarchical Differences under Higher Volatility or Low Controllability

Next, we investigated how volatility and controllability affect the layer-wise learning rates in the Actor network. Previous work in supervised learning models trained on image classification tasks has shown that learning rates in earlier layers tend to plateau earlier and at lower values than those in later layers (Zhou & Schapiro., 2025). Our meta–reinforcement learning models replicated this hierarchical pattern under stable conditions (controllability = 1, volatility = 0), learning rates in earlier layers plateaued earlier and reached lower asymptotic values, whereas learning rates in higher layers continued to increase and converged to higher levels (Figure 3A, controllability =1, volatility = 0, layer-wise ANOVA $p < 0.001$).

We next examined whether environmental modulation altered this hierarchical organization by computing the standard deviation across layers for the maximum learning rate. Compared with the zero-volatility condition, increasing volatility produced a graded reduction in layer-wise differences of maximum learning rate (Figure 3B; controllability = 1, volatility = 1; layer-wise ANOVA, $P = 0.021$). A similar but more pronounced effect was observed with reduced controllability, which led to a stronger reduction in layer-wise differences in maximum learning rate (Figure 3C; controllability = 0, volatility = 0; layer-wise ANOVA, $P = 0.853$). We calculated the standard deviation across layers for all combinations of volatility and controllability and found that layer-wise differences in maximum learning rates decreased progressively as volatility increased or controllability decreased (Figure 3D). Overall, these findings suggest that highly volatile or poorly controllable environments reduce the differentiation of learning rates across layers.

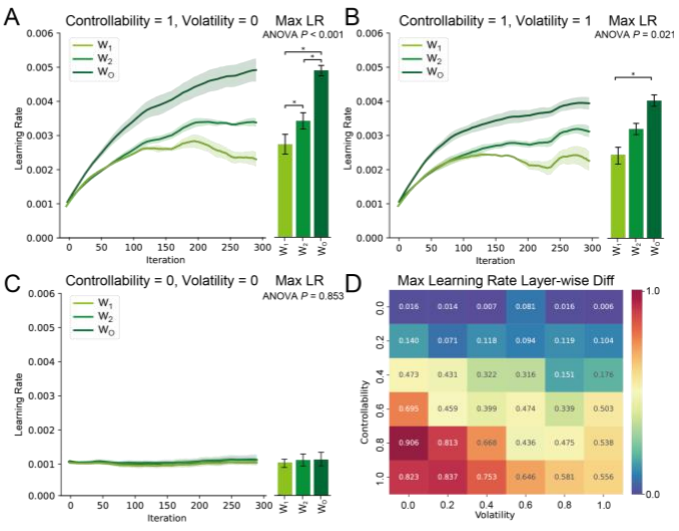

Figure 3. Layer-wise differences across different environmental conditions. (A–C) Developmental trajectories and peak learning rates of weights across layers under (A) stable conditions (controllability = 1, volatility = 0), (B) highly volatile conditions (controllability = 1, volatility = 1), and (C) low-controllability conditions (controllability = 0, volatility = 0). **(D)** Layer-wise standard deviation of maximum learning rate.

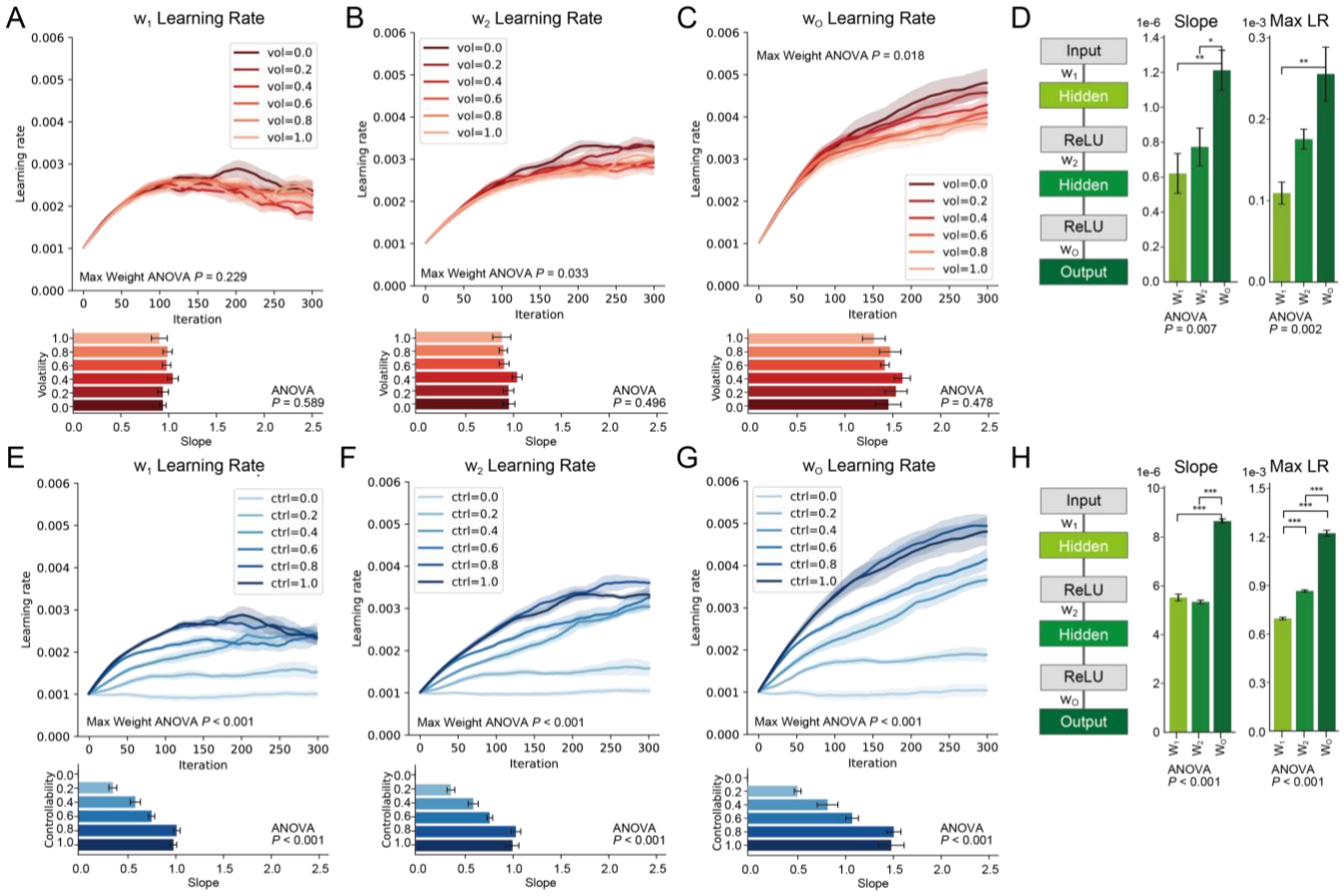

Figure 4. Learning-rate dynamics of linear layers across different environmental conditions. (A–C, E–G) Learning-rate trajectories and initial slope for the weights of hidden layers 1 (w_1; A, E), 2 (w_2; B, F), and the output layer (w_O; C, G) in the obstacle environment under different levels of (A–C) volatility and (E–G) controllability. **(D, H)** Standard deviation across conditions for each layer in initial slope (Slope) and maximum learning rate (Max LR) under varying (D) volatility (H) and controllability.

Greater Impact of Volatility and Controllability on Later Layers

Next, we investigated how volatility and controllability differentially affect learning rates across layers. We found that increasing volatility reduced the maximum learning rate in the later layers (w_2 and w_O), but not in the first layer (w_1) (Figure 4A–C; ANOVA, P = 0.229, 0.033, and 0.018, respectively). Volatility did not affect the initial slope in any layer (Figure 4A–C; ANOVA, P = 0.589, 0.496 and 0.478, respectively). In contrast, reduced controllability substantially decreased the maximum learning rate across all layers and flattened the initial slope throughout the network (Figure 4E–G; all ANOVAs, $P < 0.001$).

To test which layers were most susceptible to environmental modulation, we compared the effects of volatility and controllability on initial slope and maximum learning rate across layers. For both volatility and controllability, the initial slope and the maximum learning rate were more strongly affected in later layers than in earlier layers (Figure 4D, volatility: initial slope P = 0.007, maximum learning rate P = 0.002; Figure 4H, controllability: initial slope $P < 0.001$, maximum learning rate $P < 0.001$). Together, these results indicate that later layers are more sensitive to environmental modulation, both volatility and controllability, than earlier layers.

Discussion

In this study, we employed a meta–reinforcement learning framework to investigate how environmental volatility and controllability shape the trajectories of learning rates. Our results show that, in both simple open environments and more complex obstacle environments, higher volatility and lower controllability reduce the maximum attainable reward, which aligns with generally worse learning outcomes observed in adverse environments in humans (Rakesh et al., 2024). By examining the trajectories of meta-learning rates, we further found that both high volatility and low controllability reduce the maximum learning rate; however, only controllability affects the initial slope of learning, leading to a flattening of early learning dynamics. Notably, later model layers were more strongly influenced by both volatility and controllability than earlier layers, and the layer-wise

hierarchy of maximum learning rate was attenuated under highly volatile or poorly controllable conditions.

The reduced learning rate in low-controllability environments is consistent with previous computational work (Dorfman & Gershman, 2019; Csifcsák et al., 2019). However, the learning rate was also slightly but significantly reduced in high-volatility environments (Nassar et al., 2020; McGuire et al., 2014), which appears to contradict previous findings. A key difference between controllability and volatility is that controllability reduces the learning rate from the very beginning of learning (i.e., a lower initial slope), even before performance reaches a plateau, whereas volatility primarily affects the learning rate after performance has stabilized. Most previous studies have examined learning rates during the early, active learning phase, prior to plateauing. Therefore, the discrepancy between previous and current findings may be explained by differences in the phase of training being analyzed. Another possible confound is that reward accumulation and learning rate changes are closely aligned across volatility and controllability conditions, making it unclear whether the observed effects on learning rate are independent or simply reflect differences in reward accumulation dynamics. This issue could be addressed in future work by aligning reward accumulation trajectories across conditions and testing whether learning-rate differences persist. Alternatively, a train–test paradigm could be used, in which the effects of environmental perturbations are evaluated based on performance in a test environment that is independent of the training environment.

Increases in volatility and reductions in controllability had stronger effects on the later layers of the model. Drawing a loose analogy between earlier versus later layers and more sensory versus higher-order association systems, this pattern is broadly consistent with the idea that higher-order regions, often characterized by more prolonged plasticity, may be particularly sensitive to environmental perturbations (Sydnor et al., 2023). The observed attenuation of layer-wise hierarchical organization may also provide a mechanistic explanation for empirical links between adversity and reduced regional specialization (Vedechkina et al., 2024). However, any mapping between model layers and cortical hierarchy remains approximate. Additional evidence, such as systematic shifts in representations across layers (e.g., from more concrete to more abstract features, or from sensory-related to action-related information), would be needed to better support this correspondence. Furthermore, because volatility and controllability were held constant throughout training, the present results most closely reflect chronic environmental exposure. It is possible that temporally specific environmental modulation could produce different outcomes. For example, early exposure may affect earlier layers more strongly (Tooley et al., 2025).

Several limitations should be noted. First, the study relies solely on meta-learned learning rate as a measure of learning paradigm, which captures only one aspect of adaptive change. Future work could strengthen this framework by incorporating representation-level metrics, such as changes in internal feature structure or geometry, to provide a more comprehensive account of learning dynamics. Second, the correspondence between model layers and specific brain regions is necessarily approximate. Future work could incorporate more biologically realistic components, such as recurrent networks with both feedforward and feedback connectivity (Spoerer et al., 2017). Third, we only considered volatility and controllability as proxies for environmental adversity. Real-world adversity involves additional dimensions, such as threat and deprivation (McLaughlin et al., 2014; Humphreys & Zeanah, 2015). Previous studies have also reported that overall reward rate or the prevalence of positive versus negative rewards can influence learning rates in computational models, providing another potential proxy for adversity (Nussenbaum & Hartley, 2024). Expanding the framework to include a broader range of environmental variables could enhance its ecological validity.

Overall, our results show the potential of a computational framework to disentangle the distinctive effects of different dimensions of adversity, which often co-occur and are difficult to separate in empirical data. Our approach also allows us to examine how different environmental stressors can have distinct effects on learning dynamics, including their variation across layers of the model. While any correspondence between model layers and brain systems remains tentative, the observed pattern is broadly consistent with empirical hypotheses that higher-order association regions may be particularly sensitive to environmental perturbations. In summary, our findings suggest that a meta–reinforcement learning paradigm may provide a simplified yet potentially useful framework for modeling how environmental factors influence the learning dynamics across different components of a learning system.

Acknowledgments

This work was supported by the National Science Foundation CAREER award (to A.P.M. under Grant No. 2045095), the Nakajima Foundation Scholarship (to M.N.) and the Quad Fellowship (to M.N.). The authors disclose the use of the large language model for assistance with coding and for revising and improving the clarity of the manuscript.

References

Arnold, S. M. R., Mahajan, P., Datta, D., Bunner, I., & Zarkias, K. S. (2020). *learn2learn: A Library for Meta-Learning Research*. doi:10.48550/arXiv.2008.12284

Chevalier-Boisvert, M., Dai, B., Towers, M., De Lazcano, R., Willems Miple, L., Lahlou, S., Castro, P. S., Deepmind, G., & Terry, J. (2023). Minigrid & Miniworld: Modular & Customizable Reinforcement Learning Environments for Goal-Oriented Tasks. *Advances in Neural Information Processing Systems*, *36*. doi:10.48550/arXiv.2306.13831

Collins, A.G.E., Shenhav, A. (2022). Advances in modeling learning and decision-making in neuroscience.

Neuropsychopharmacol. 47. doi:10.1038/s41386-021-01126-y

Csifcsák, G., Melsæter, E., & Mittner, M. (2019). Intermittent absence of control during reinforcement learning interferes with pavlovian bias in action selection. *Journal of Cognitive Neuroscience*, *32*(4). doi:10.1162/jocn_a_01515

De Courson, B., Frankenhuis, W. E., & Nettle, D. (2025). Poverty is associated with both risk avoidance and risk taking: empirical evidence for the desperation threshold model from the UK and France. *Proceedings of the Royal Society B: Biological Sciences*, *292*(2040). doi:10.1098/rspb.2024.2071

Dorfman, H. M., & Gershman, S. J. (2019). Controllability governs the balance between Pavlovian and instrumental action selection. *Nature Communications*, *10*(1). doi:10.1038/s41467-019-13737-7

Draper, C., & Yousafzai, A. (2024). The next 1000 days: building on early investments for the health and development of young children. *Lancet*. doi: 10.1016/S0140-6736(24)01389-8

Humphreys, K. L., Lee, S. S., Telzer, E. H., Gabard-Durnam, L. J., Goff, B., Flannery, J., & Tottenham, N. (2015). Exploration-exploitation strategy is dependent on early experience. *Developmental Psychobiology*, *57*(3). doi:10.1002/dev.21293

Humphreys, K. L., & Zeanah, C. H. (2015). Deviations from the Expectable Environment in Early Childhood and Emerging Psychopathology. *Neuropsychopharmacology*. doi:10.1038/npp.2014.165

Ligneul, R. (2021). Prediction or Causation? Towards a Redefinition of Task Controllability. *Trends in Cognitive Sciences* (Vol. 25, Number 6). doi:10.1016/j.tics.2021.02.009

Lloyd, A., McKay, R. T., & Furl, N. (2022). Individuals with adverse childhood experiences explore less and underweight reward feedback. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(4). doi:10.1073/pnas.2109373119

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron*, *84*(4). doi:10.1016/j.neuron.2014.10.013

McLaughlin, K. A., Sheridan, M. A., & Lambert, H. K. (2014). Childhood adversity and neural development: Deprivation and threat as distinct dimensions of early experience. *Neuroscience and Biobehavioral Reviews*. doi:10.1016/j.neubiorev.2014.10.012

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*(37). doi:10.1523/JNEUROSCI.0822-10.2010

Nussenbaum, K., & Hartley, C. A. (2024). Understanding the development of reward learning through the lens of meta-learning. *Nature Review Psychology*. doi:10.1038/s44159-024-00304-1

Piray P., Daw N. D. (2020). A simple model for learning in volatile environments. *PLoS Computational Biology*, *16*(7). doi:10.1371/journal.pcbi.1007963

Raab, H. A., Foord, C., Ligneul, R., & Hartley, C. A. (2022). Developmental shifts in computations used to detect environmental controllability. *PLoS Computational Biology*, *18*(6). doi:10.1371/journal.pcbi.1010120

Rakesh, D., Lee, P. A., Gaikwad, A., & McLaughlin, K. A. (2025). Annual Research Review: Associations of socioeconomic status with cognitive function, language ability, and academic achievement in youth: a systematic review of mechanisms and protective factors. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. doi:10.1111/jcpp.14082

Sheridan, M. A., Peverill, M., Finn, A. S., & McLaughlin, K. A. (2017). Dimensions of childhood adversity have distinct associations with neural systems underlying executive functioning. *Development and Psychopathology*, *29*(5). doi:10.1017/S0954579417001390

Spoerer, C. J., McClure, P., & Kriegeskorte, N. (2017). Recurrent convolutional neural networks: A better model of biological object recognition. *Frontiers in Psychology*. doi:10.3389/fpsyg.2017.01551

Sydnor, V. J., Larsen, B., Seidlitz, J., Adebimpe, A., Alexander-Bloch, A. F., Bassett, D. S., Bertolero, M. A., Cieslak, M., Covitz, S., Fan, Y., Gur, R. E., Gur, R. C., Mackey, A. P., Moore, T. M., Roalf, D. R., Shinohara, R. T., & Satterthwaite, T. D. (2023). Intrinsic activity development unfolds along a sensorimotor–association cortical axis in youth. *Nature Neuroscience*, *26*(4). doi:10.1038/s41593-023-01282-y

Tooley, U. A., Latham, A., Kenley, J. K., Alexopoulos, D., Smyser, T. A., Nielsen, A. N., Gorham, L., Warner, B. B., Shimony, J. S., Neil, J. J., Luby, J. L., Barch, D. M., Rogers, C. E., & Smyser, C. D. (2024). Prenatal environment is associated with the pace of cortical network development over the first three years of life. *Nature Communications*, *15*(1). doi:10.1038/s41467-024-52242-4

Vedechkina, M., Astle, D. E., & Holmes, J. (2024). Dimensions of early life adversity and their associations with functional brain organisation. *Imaging Neuroscience*, *2*. doi:10.1162/imag_a_00145

Xu, Y., Harms, M. B., Green, C. S., Wilson, R. C., & Pollak, S. D. (2023). Childhood unpredictability and the development of exploration. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(49). doi:10.1073/pnas.2303869120

Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience, 19*(3). doi:10.1038/nn.4244

Zhou, Z., & Schapiro, A. (2025). A gradient of complementary learning systems emerges through meta-learning. *BioArxiv*. doi:10.1101/2025.07.10.664201