Meta-learning local learning rules for structured credit assignment with sparse feedback

Anonymous Author(s)
Affiliation
Address
email

Abstract

Biological neural networks learn complex behaviors from sparse, delayed feedback using local synaptic plasticity, yet the mechanisms enabling structured credit assignment remain elusive. In contrast, artificial recurrent networks solving similar tasks typically rely on biologically implausible global learning rules or hand-crafted local updates. The space of local plasticity rules capable of supporting learning from delayed reinforcement remains largely unexplored. Here, we present a meta-learning framework that discovers local learning rules for structured credit assignment in recurrent networks trained with sparse feedback. Our approach interleaves local neo-Hebbian-like updates during task execution with an outer loop that optimizes plasticity parameters via **backpropagation through learning**. The resulting three-factor learning rules enable long-timescale credit assignment using only local information and delayed rewards, offering new insights into biologically grounded mechanisms for learning in recurrent circuits.

1 Introduction

Learning in biological organisms involves changes in synaptic connections (synaptic plasticity) between neurons [1, 2]. Synaptic changes are believed to underlie memory formation and are essential for adaptive behaviour [3]. Experimental evidence suggests that synaptic changes depend on the co-activation of pre- and postsynaptic activity [4, 5], and possibly other local variables available at the synaptic site [6, 7]. These unsupervised synaptic modifications have explained activity-dependent circuit refinement during development such as the emergence of functional properties like receptive field formation based on naturalistic input statistics [8].

Yet, most organisms routinely solve complex tasks that require feedback through explicit supervisory or reinforcement signals. These signals are believed to gate or modulate plasticity, acting in the form of a third factor that scales and also probably imposes the direction of the synaptic modifications [9, 10]. How error- or reward-related information is propagated through the recurrent interactions is not yet clear. While prior work has largely focused on hand-crafted synaptic updates for unsupervised self-organization, or biologically plausible approximations of backpropagation [11], the space of plasticity rules capable of supporting structured credit assignment from delayed feedback remains vastly underexplored.

Backpropagation through time (BPTT), the standard approach for training recurrent neural networks (RNNs), is biologically implausible since it requires symmetric forward and backward connections and non-local information [12, 13]. Although recent work has reformulated BPTT into more biologically plausible variants using random feedback [14], truncated approximations [15], or by learning feedback pathways [16], these methods require continuous error signals to refine recurrent connections.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

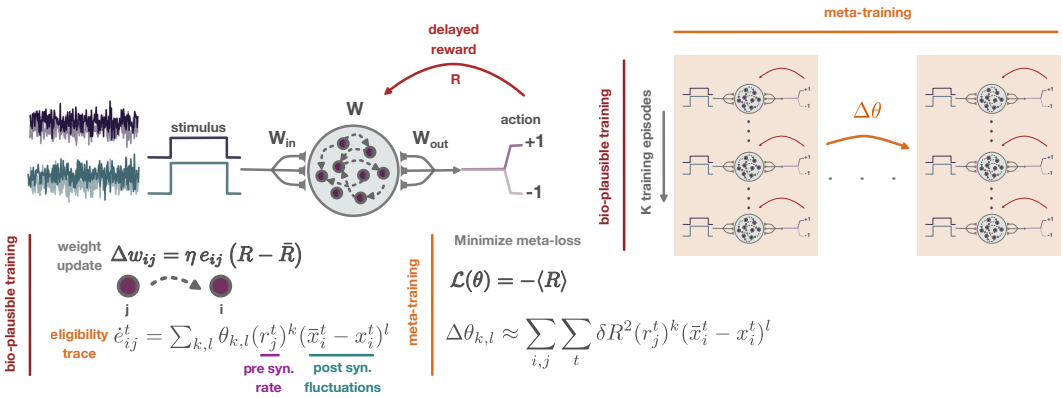

Figure 1

Outline of the proposed meta-learning framework.

Here, we adopt a bottom-up approach: instead of imposing hand-designed synaptic rules, we discover biologically plausible plasticity rules that support learning through delayed reinforcement signals via meta-optimisation. Building on recent work [17], we parameterise plasticity rules as functions of local signals (presynaptic activity, postsynaptic activity, and synapse size) and meta-learn their parameters within a second reinforcement learning loop. With that, our present work tackles the following questions:

- **Which local learning rules can implement structured credit assignment under biological constraints?**
- **Do different forms of plasticity give rise to different computational regimes and representations as observed with gradient based training (e.g., "lazy" vs. "rich" learning)?**

Recent theory distinguishes between lazy and rich regimes of learning in RNNs: in the lazy regime, representations remain fixed while output weights adapt; in the rich regime, the network reorganises its internal dynamics to encode task structure. While these regimes are well-characterised for gradient-trained networks, it remains unclear whether biologically plausible learning rules can support either or both, and what synaptic mechanisms underlie each regime. Here we demonstrate that different forms of plasticity naturally lead to qualitatively different learning trajectories and internal representations, akin to their gradient-based learning rules.

2 Method

Network dynamics. We consider recurrent neural networks (RNNs) of firing rate neurons coupled through a synaptic matrix $W \in \mathbb{R}^{N\times N}$ [18], with additional input and output matrices $W_{\text{in}} \in \mathbb{R}^{N_{\text{in}}\times N}$ and $W_{\text{out}} \in \mathbb{R}^{N\times N_{\text{out}}}$ that route task-relevant input into the recurrent circuit and read out network activation to generate task-specific outputs (actions). The equations governing the network dynamics are

$$\frac{\mathrm{d}\mathbf{x}^t}{\mathrm{d}t} = -\mathbf{x}^t + W\boldsymbol{\phi}(\mathbf{x}^t) + W_{\text{in}}\mathbf{u}^t, \tag{1}$$

$$\mathbf{r}^t = \phi(\mathbf{x}^t) \doteq \tanh(\mathbf{x}^t), \tag{2}$$

where $\mathbf{x}^t \in \mathbb{R}^N$ is the vector of pre-activations (or input currents) to each neuron in the network, $\boldsymbol{\phi}(\cdot) : \mathbb{R}^N \to \mathbb{R}^N$ denotes the single-neuron transfer functions, $\mathbf{r}^t \in \mathbb{R}^N_+$ is the vector of instantaneous firing rates, $\mathbf{u}^t$ stands for the activity of the N_{in} input neurons. In the terms above, the $\cdot^t$ superscript indicates time dependence. Network outputs $\mathbf{z}^t$ are obtained from linear read-out neurons as

$$\mathbf{z}^t = W_{\text{out}}\mathbf{r}^t. \tag{3}$$

Sparse feedback and parametrized learning rules. We consider networks that learn context-dependent cognitive tasks using biologically plausible local learning rules, guided by sparse reinforcement signals R provided only at the end of each training episode. To enable learning from

such delayed and global signals, each synapse between a pre-synaptic unit j and a post-synaptic unit i maintains an eligibility trace e_{ij} [19], which integrates the history of (co-)activation during the episode. We define the evolution of eligibility traces with differential equations of the form

$$\frac{\mathrm{d}e_{ij}^t}{\mathrm{d}t} = \mathcal{H}_\theta(r_j^t, x_i^t) - \frac{e_{ij}}{\tau_e} = \sum_{0 \le k,l \le d;} \theta_{k,l} \ \left(r_j^t\right)^k \left(\bar{x}_i - x_i^t\right)^l - \frac{e_{ij}}{\tau_e}, \tag{4}$$

where τ_e is a decay time-scale, $\bar{x}_i$ is a running average of the pre-activation of neuron i, and $\theta_{k,l} \in \mathbb{R}$ are learnable coefficients. In contrast to eligibility traces based solely on pairwise correlations [20], we use here a polynomial expression that captures richer interactions between pre- and post-synaptic activity. Each coefficient $\theta_{k,l}$ can be construed as a term-specific learning rate, which may be positive (Hebbian), negative (anti-Hebbian). This parameterization allows individual terms to modulate synaptic eligibility based on pre-synaptic activity, post-synaptic activity, co-activity, or deviations from a homeostatic set point. In our experiments, we set $d = 3$.

The recurrent weight matrix W gets updated at the end of each training episode according to a reward-modulated learning rule

$$\Delta w_{ij} = \ e_{ij} \ \left(R - \bar{R}\right) - \frac{w_{ij}}{\tau_w}, \tag{5}$$

where τ_w denotes the time scale of weight decay, e_{ij} stands for the eligibility trace accumulated during the episode, while R, $\bar{R}$ stand for the obtained and the expected reward. Here, we model reward expectations for each type of trial independently as a running average of past rewards for this trial type [21]. This update rule enables credit assignment through the interaction between synaptic eligibility and trial-specific reward prediction error, consistent with neo-Hebbian three-factor learning rules hypothesized to operate in biological circuits [20]. In principle the weight updates happen due to (slow) weight decay or due to reward prediction errors.

Meta-learning plasticity rules. While previous work has relied on hand-crafted eligibility trace dynamics and synaptic update rules to train recurrent neural networks with sparse feedback [21], we instead adopt a meta-learning approach to learn the parameters of the plasticity rules. Our framework consists of two nested training loops: **(i)** an inner loop in which the recurrent network is trained over several episodes using local learning rules and sparse reinforcement signals provided at the end of each episode **(bio-plausible training)**, as described above; and **(ii)** an outer loop that optimizes the plasticity meta-parameters $\mathbf{\Theta} = \{\{\theta_{k,l}\}_{k,l=0}^{3}, \tau_w, \tau_e\}$ via gradient descent using **backpropagation through learning** on a meta-loss computed over K training episodes (trials) **(meta-training)** . This approach allows the learning rules themselves to be adapted to the task, rather than be fixed a priori.

Backpropagation through learning. Our goal is to optimise the learning rule parameters θ to maximise task performance, measured as the expected cumulative reward $\langle R \rangle$ obtained after a fixed number of learning episodes. However, the reward R obtained by the agent depends on the network's output, which in turn is determined by its synaptic weights $\mathcal{W} = \{W_{in}, W, W_{out}\}$. The weights are dynamically updated according to the employed synaptic update rule (Eq. 5). This plasticity rule, depends on the eligibility traces e_{ij}, which themselves are parameterised by $\mathbf{\Theta}$. This establishes a dependency chain over the network parameters: $R \leftarrow W \leftarrow e \leftarrow \mathbf{\Theta}$. Thus directly computing the gradient $\nabla_\theta \langle R \rangle$ by backpropagating through the entire network dynamics over learning is computationally challenging.

To address this, we employ a REINFORCE-inspired approximation [22] to estimate the gradient $\nabla_\theta \langle R \rangle$. Recall that the REINFORCE gradient formula involves computing the gradient of an expected value by observing outcomes and scaling a measure of what elicited that outcome with the associated reward. Or more formally, scaling the gradient of the log-probability of an outcome with the reward associated with that outcome

$$\nabla_\Theta \langle R \rangle = \langle (R - \bar{R}) \ \nabla_\Theta \log \pi(R \mid \mathbf{\Theta}) \rangle. \tag{6}$$

Here, since we consider deterministic weight updates, we do not have a stochastic policy π, as is common in policy gradient methods in reinforcement learning. However, we can consider the final weight configuration $\mathcal{W}(\mathbf{\Theta})$ as an *implicit policy* with parameters Θ, that determine the learned network behaviour. We then use the **reward prediction error**, defined as $\delta R = R - \bar{R}$ (where $\bar{R}$ is a running average of the reward), as a scaling factor to adapt the parameters Θ

$$\nabla_\Theta \langle R \rangle \approx (R - \bar{R}) \, \frac{\mathrm{d}W}{\mathrm{d}\Theta}. \tag{7}$$

Since the weight updates depend linearly on the eligibility trace (Eq. 5), we have for the plasticity parameters

$$\frac{\mathrm{d}W_{ij}}{\mathrm{d}\theta_{kl}} = \delta R \frac{\mathrm{d}e_{ij}}{\mathrm{d}\theta_{kl}}. \tag{8}$$

To relate this to the gradient of the reward with respect to θ, we sum over all synapses, resulting in the approximation

$$\nabla_\theta \langle R \rangle \approx \sum_{i,j} \delta R \frac{\mathrm{d}e_{ij}}{\mathrm{d}\theta_{kl}} = \sum_{i,j} \delta R \left(r_j^t\right)^k \left(\bar{x}_i^t - x_i^t\right)^l. \tag{9}$$

The eligibility trace e_{ij} is a function of neural activity, and its dependency on the parameters θ is explicitly defined by the model (Eq. 4). For the eligibility trace parametrised in the polynomial form of Eq. 4, the term $\frac{\mathrm{d}e_{ij}}{\mathrm{d}\theta}$ has an explicit expression in terms of neural activations and firing rates (Eq. 9). This expression is fully analytic and requires no gradient propagation through the network or the learning episodes. The plasticity parameters θ are then updated using gradient ascent based on this estimated gradient.

To enforce sparsity on the identified rules in order to minimise the number of active terms in the identified rule to render it interpretable.

3 Results

4 Related work

Decades of research on synaptic plasticity have focused on hand-crafted learning rules designed to replicate experimentally observed changes in post-synaptic potentials from single-neuron recordings. However, the recent explosion in large-scale functional recordings, particularly longitudinal data collected across learning, has sparked growing interest in identifying the types of plasticity rules that may underlie observed changes in neural activity and behavioural performance. Despite this interest, the task remains extremely challenging: current experimental techniques do not allow direct measurement of synaptic interactions across large neural populations, making it difficult to infer the underlying synaptic mechanisms at play. Thus an increasing number of frameworks have emerged that aim to discover plasticity rules from indirect signatures such as changes in neural activity distributions, recorded trajectories, or behavioural performance. These approaches differ in what kind of observations they use, and in the assumptions they make about the network structure, plasticity rule parameterization, and underlying task.

Matching rate distributions. One line of work focuses on inferring synaptic plasticity rules from pre- and post-learning firing rate distributions. Lim et al. [23] jointly infer neuron transfer functions and synaptic updates from observed rate distributions, under assumptions of Poisson firing statistics and linearized plasticity. This approach was later extended using Gaussian process priors over plasticity functions [24], improving flexibility but still restricted to feedforward networks and ignoring temporal dynamics.

These approaches do not model the full trajectory of activity during learning, instead identify plasticity rules that explain cumulative changes across learning. As a result, they cannot constrain rule parameters based on how learning unfolded in time.

Inference by conditioning on neural trajectories. A second group of methods exploits neural activity trajectories recorded over learning. Ramesh et al. [25] use a generative adversarial framework to infer plasticity rules that generate neural trajectories similar to empirical ones. While highly expressive, this method requires extensive data and computational resources, and suffers from known instability issues in GAN training. Confavreux et al. [17] proposed a meta-learning framework to discover plasticity rules that produce desired temporal coding properties in rate-based networks. While insightful, their approach optimises for a fixed synthetic objective (e.g., encoding elapsed time), rather than learning from observed data or behaviour.

Behavior-based plasticity inference. A third set of studies use behavioural performance trajectories to constrain synaptic plasticity. Ashwood et al.[26] fit learning rule parameters in rodent

decision tasks using a Bayesian model, requiring approximation of the full posterior over synaptic weights. Rajagopalan et al.[27] reformulate the plasticity inference problem as logistic regression by assuming presynaptic activity and reward as the only inputs. These frameworks remain limited in flexibility, often neglecting dependencies on postsynaptic activity or synapse strength, which are essential for biologically grounded learning.

Most of these approaches assume feed-forward structure of the underlying network [24, 28], and consider plasticity evolving network dynamics in an unsupervised setting. Only the recent work of [28] considers a reward term in the plasticity rule, that effectively puts the learning framework under a reinforcement learning and thus closer to how biological organisms learn.

5 Limitations

Despite its strengths, our work has several limitations that point to opportunities for future improvement and extension. One limitation is that the proposed meta-learning procedure must be run multiple times independently to discover multiple plasticity rules that satisfy the same task constraints. Recent advances using simulation-based inference [17] provide a promising alternative for sampling entire distributions over plasticity rules that solve a given cognitive task, potentially offering a more efficient and principled exploration of solution space. Yet, simulation based inference is easy to incorporate in our setting.

Another limitation is that our current framework is purely exploratory and does not explicitly incorporate constraints from experimentally recorded neural activity. While this allows for a broad and flexible search over possible learning mechanisms, it limits the biological specificity of the discovered rules. Extending our framework to incorporate such constraints, for instance, by biasing the meta-optimisation toward activity trajectories consistent with recorded data, could yield more realistic models of synaptic updates.

References

[1] Craig H Bailey and Eric R Kandel. Structural changes accompanying memory storage. *Annual review of physiology*, 1993.

[2] Mark Mayford, Steven A Siegelbaum, and Eric R Kandel. Synapses and memory storage. *Cold Spring Harbor perspectives in biology*, 4(6):a005751, 2012.

[3] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[4] Guo-qiang Bi and Mu-ming Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of neuroscience*, 18(24):10464–10472, 1998.

[5] Per Jesper Sjöström, Gina G Turrigiano, and Sacha B Nelson. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32(6):1149–1164, 2001.

[6] Michael Graupner and Nicolas Brunel. Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proceedings of the National Academy of Sciences*, 109(10):3991–3996, 2012.

[7] Victor Pedrosa and Claudia Clopath. Voltage-based inhibitory synaptic plasticity: network regulation, diversity, and flexibility. *bioRxiv*, pages 2020–12, 2020.

[8] Stephen J Martin, Paul D Grimwood, and Richard GM Morris. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual review of neuroscience*, 23(1):649–711, 2000.

[9] Łukasz Kuśmierz, Takuya Isomura, and Taro Toyoizumi. Learning with three factors: modulating hebbian plasticity with errors. *Current opinion in neurobiology*, 46:170–177, 2017.

[10] Baram Sosis and Jonathan E Rubin. Distinct dopaminergic spike-timing-dependent plasticity rules are suited to different functional roles. *bioRxiv*, 2024.

[11] Thomas Miconi, Kenneth Stanley, and Jeff Clune. Differentiable plasticity: training plastic neural networks with backpropagation. In *International Conference on Machine Learning*, pages 3559–3568. PMLR, 2018.

[12] Timothy P Lillicrap, Daniel Cownden, Douglas B Tweed, and Colin J Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications*, 7(1):13276, 2016.

[13] Jordan Guerguiev, Timothy P Lillicrap, and Blake A Richards. Towards deep learning with segregated dendrites. *Elife*, 6:e22901, 2017.

[14] Navid Shervani-Tabar and Robert Rosenbaum. Meta-learning biologically plausible plasticity rules with random feedback pathways. *Nature Communications*, 14(1):1805, 2023.

[15] James M Murray. Local online learning in recurrent networks with random feedback. *Elife*, 8: e43299, 2019.

[16] Jack Lindsey and Ashok Litwin-Kumar. Learning to learn with feedback and local plasticity. *Advances in Neural Information Processing Systems*, 33:21213–21223, 2020.

[17] Basile Confavreux, Poornima Ramesh, Pedro J Goncalves, Jakob H Macke, and Tim Vogels. Meta-learning families of plasticity rules in recurrent spiking networks using simulation-based inference. *Advances in Neural Information Processing Systems*, 36:13545–13558, 2023.

[18] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. Chaos in random neural networks. *Physical review letters*, 61(3):259, 1988.

[19] Eugene M Izhikevich. Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cerebral cortex*, 17(10):2443–2452, 2007.

[20] Wulfram Gerstner, Marco Lehmann, Vasiliki Liakoni, Dane Corneil, and Johanni Brea. Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in neural circuits*, 12:53, 2018.

[21] Thomas Miconi. Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks. *Elife*, 6:e20899, 2017.

[22] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.

[23] Sukbin Lim, Jillian L McKee, Luke Woloszyn, Yali Amit, David J Freedman, David L Sheinberg, and Nicolas Brunel. Inferring learning rules from distributions of firing rates in cortical neurons. *Nature neuroscience*, 18(12):1804–1810, 2015.

[24] Shirui Chen, Qixin Yang, and Sukbin Lim. Efficient inference of synaptic plasticity rule with gaussian process regression. *Iscience*, 26(3), 2023.

[25] Poornima Ramesh, Basile Confavreux, Pedro J Goncalves, Tim P Vogels, and Jakob H Macke. Indistinguishable network dynamics can emerge from unalike plasticity rules. *bioRxiv*, pages 2023–11, 2023.

[26] Zoe Ashwood, Nicholas A Roy, Ji Hyun Bak, and Jonathan W Pillow. Inferring learning rules from animal decision-making. *Advances in Neural Information Processing Systems*, 33: 3442–3453, 2020.

[27] Adithya E Rajagopalan, Ran Darshan, Karen L Hibbard, James E Fitzgerald, and Glenn C Turner. Reward expectations direct learning and drive operant matching in drosophila. *Proceedings of the National Academy of Sciences*, 120(39):e2221415120, 2023.

[28] Yash Mehta, Danil Tyulmankov, Adithya Rajagopalan, Glenn Turner, James Fitzgerald, and Jan Funke. Model based inference of synaptic plasticity rules. *Advances in Neural Information Processing Systems*, 37:48519–48540, 2024.