Biologically-plausible backpropagation through arbitrary timespans via local neuromodulators

Anonymous Author(s)
Affiliation
Address
email

Abstract

The spectacular successes of recurrent neural network models where key parameters are adjusted via backpropagation-based gradient descent have inspired much thought as to how biological neuronal networks might solve the corresponding synaptic credit assignment problem [1, 2, 3]. There is so far little agreement, however, as to how biological networks could implement the necessary backpropagation through time, given widely recognized constraints of biological synaptic network signaling architectures. Here, we propose that extra-synaptic diffusion of local neuromodulators such as neuropeptides may afford an effective mode of backpropagation lying within the bounds of biological plausibility. Going beyond existing temporal truncation-based gradient approximations [4, 5, 6], our approximate gradient-based update rule, ModProp, propagates credit information through *arbitrary* time steps. ModProp suggests that modulatory signals can act on receiving cells by convolving their eligibility traces via causal, time-invariant and synapse-type-specific filter taps. Our mathematical analysis of ModProp learning, together with simulation results on benchmark temporal tasks, demonstrate the advantage of ModProp over existing biologically-plausible temporal credit assignment rules. These results suggest a potential neuronal mechanism for signaling credit information related to recurrent interactions over a longer time horizon. Finally, we derive an *in-silico* implementation of ModProp that could serve as a low-complexity and causal alternative to backpropagation through time.

1 Introduction

Recurrent connectivity is a hallmark of neuronal circuits. While this feature enables rich and flexible computation, mechanisms enabling efficient task learning in large circuits remain a central problem in neuroscience and artificial intelligence research. Fundamentally, the problem stems from the fact that potentially all history of all neurons, including synaptically far away ones, can affect neuronal activity and contribute to task output. Motivated by the success of gradient descent learning, several biological learning models approximate the exact gradient in recurrent neural networks using known biological processes and gain insights into computational principles of how the brain might learn [4, 5, 6].

By ignoring dependencies beyond a few recurrent steps — thus severely truncating the gradient computational graph — these existing models succeed in representing their approximate gradient-based update rule as a combination of terms that resemble known synaptic physiological processes: "eligibility trace", which maintains a fading memory of coincidental activation of presynaptic and postsynaptic neurons [7, 8, 9, 10, 11], combined with a third “modulatory” factor — top-down learning or “reward” signals [12, 13, 14, 15], for which dopamine is a prominent candidate [16]. Despite the impressive performance of such approximations on a variety of tasks, truncating relevant credit information results in a significant performance gap compared to algorithms using exact gradient

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

information; backpropagation-through-time (BPTT) and real time recurrent learning (RTRL) [4, 5, 6, 17].

How might neural circuits account for long-term recurrent interactions in order to assign credit (or blame) to neuronal firing that happened arbitrary steps before the presentation of reward? Given the rich repertoire of dynamical and signaling elements in the brain, one avenue could be to examine biological processes that have been underexplored in existing models. Dopamine — whose cellular actions are exerted by activation of G protein-coupled receptors (GPCRs), which can greatly impact STDP — is not the only neuromodulator involved in learning [18, 19, 20, 21]. More recently, transcriptomic studies have uncovered strong evidence for many other neuromodulatory pathways throughout the brain that also act via the activation of GPCRs, leading to similar downstream actions as dopamine [22, 23]. This suggests that, similar to dopamine, they could also play a role in shaping synaptic credit assignment. A conspicuous family within these pathways is neuropeptide signaling because peptidergic genes are densely and abundantly expressed in the forebrains of divergent species, including the human, in a cell-type-specific fashion [24], suggesting widespread interaction between synaptic and peptidergic modulatory networks for synaptic credit assignment [25, 6]. Moreover, intra-cortical expression allows neuropeptides to potentially carry information *local* to the cortical network, cell type specificity enables sculpted signals for different recipient cells, and their diffusive nature could enable communication between neurons that are not synaptically connected. Finally, peptidergic signals have time-scales much longer than the time scales for axonal propagation of action potentials or synaptic delays [26]. Taken together, these properties make cell-type-specific local neuromodulation seem promising for propagating credit signal over multiple recurrent steps. Developing explicit computational principles of how these local modulatory elements could propagate credit signal over arbitrary recurrent steps could advance our understanding of biological learning and may inspire more efficient low complexity bio-inspired learning rules.

Motivated by the question above as well as shortcomings of gradient approximations based on severe temporal truncations, we investigate how biological credit signals could be propagated through *arbitrary* recurrent steps via widespread cell-type-specific neuropeptidergic signaling [23, 25]. While Ref. [6] recently introduced a framework exploiting properties of neuropeptidergic signaling for temporal credit assignment, similar to [4] and [5], their approach performs severe temporal truncation of the error gradient and does not consider credit propagation beyond disynaptic connections. Our main contributions are summarized as follows:

- We derive a theory that provides mechanism and intuition for the effectiveness of synapse-type-specific modulatory backpropagation (through time) weights (Theorem 1).
- We develop a model predicting how modulatory signaling could be the basis for biological backpropagation through *arbitrary* time spans (Figures 1 and 2). Unlike temporal truncation-based approximations [4, 5, 6], our model enables each neuron to receive filtered (rather than precise) credit signals regarding its contribution to the outcome via neuromodulation.
- We demonstrate the effectiveness of modulatory signaling via synapse-type-specific, rather than synapse-specific, modulatory weights on learning tasks that involve temporal credit assignment (Figure 4). In particular, we demonstrate an *online* learning setting where weights are updated causally and in real-time (Figure 5).
- We also derive a low-complexity *in silico* implementation of our algorithm suitable for *online learning* (Proposition 1).

2 Related works

2.1 Neuromodulatory factors in synaptic plasticity

One of the most fundamental learning rules, Hebbian plasticity, attributes lasting change in synaptic strength and memory formation to correlations of spike timing between particular presynaptic and postsynaptic neurons [27, 28]. However, multiple experimental and theoretical investigations now indicate that the Hebbian rule alone is insufficient. First, there have been numerous suggestions that some persistent "eligibility trace" must exist to bridge the temporal gap between correlated firings at millisecond timescales and behavioral timescales lasting seconds [7, 8, 10, 9, 29, 30]. Moreover, impacts of correlated spike timing must be augmented by one or more additional modulatory factors to steer weight updates toward desired outcome [31, 16, 32, 15, 33, 7, 1, 34, 35, 36, 8, 37]. This

is commonly known as learning with three factors. A prominent candidate for such a modulatory factor is dopamine [13]. Dopamine influences receiving neurons via activation of G protein-coupled receptors (GPCRs), which can regulate membrane excitability and key parameters of synaptic plasticity rules.

Besides dopamine, recent transcriptomic evidence has uncovered genes encoding a plethora of other neuromodulatory pathways throughout the brain, including neuropeptide signaling genes that are abundantly expressed in the forebrains of tetrapods, including the human [23, 24]. Like dopamine, their cellular actions are exerted by the activation of G protein-coupled receptors (GPCRs), which can persistently modulate Hebbian synaptic plasticity [15, 38, 39]. This suggests that they too, could play a role in shaping cortical learning and synaptic credit assignment. Moreover, nearly all neurons express one or more neuropeptide signaling gene [23], which suggests a dense interplay between synaptic and peptidergic modulatory networks to shape synaptic credit assignment [25].

2.2 Approximate gradient descent learning

Standard algorithms for gradient descent learning in recurrent neural networks (RNNs), real time recurrent learning (RTRL) and backpropagation through time (BPTT), are not biologically-plausible [40] and have vast computational and storage demands [41]. However, multiple studies have shown that learning algorithms that approximate the gradient, while mitigating some of the problems of exact gradient computation, can lead to satisfactory learning outcomes [40, 42, 43]. In feedforward networks, plenty of biologically-plausible learning rules have been proposed and demonstrated impressive performance that rival backpropagation on many different tasks [33, 40, 14, 43, 34, 44, 45, 46, 47, 48].

For efficient online learning in RNNs, approximations to RTRL have been proposed [17, 49, 50, 51, 4, 52, 53]. For instance, a recent influential study [5] conceived how the three-factor learning framework could approximate the gradient. Among the biologically-plausible proposals [4, 5, 6], approximations have mainly been based on temporal truncations of the gradient computation graph and because of that, their ability to learn dependencies across arbitrary recurrent steps has been limited. Lastly, non-truncation based approximations have been proposed outside of bio-plausible research [49, 50].

3 Results

3.1 Learning rule overview

We consider a discrete-time rate-based RNN similar to the form in [55] with observable states, i.e. firing rates, as z_t at time t, and the corresponding internal states as s_t. W denotes the recurrent weight matrix with $(pq)^{\text{th}}$ entry W_{pq} representing the synaptic connection strength between presynaptic neuron q and postsynaptic neuron p. See Supp. Section A for the neuron model.

RNNs are typically trained by gradient descent learning on the task error (or negative reward) E. However, the two equivalent factorizations of error gradient in RNNs, BPTT and RTRL, both involve nonlocal information that is inaccessible to neural circuits. This is due to recurrent connectivity: a synaptic weight W_{pq} can affect loss E through many other neurons in the other network in addition to its pre- and postsynaptic neurons. To see this more concretely, the following is the exact gradient at time t using RTRL factorization, on which we base our approximation for online learning:

$$\frac{\mathrm{d}\,E}{\mathrm{d}\,W_{pq}}|_t = \sum_j \frac{\partial E}{\partial s_{j,t}} \frac{\mathrm{d}\,s_{j,t}}{\mathrm{d}\,W_{pq}} \tag{1}$$

$$\begin{aligned} \frac{\mathrm{d}\,s_{j,t}}{\mathrm{d}\,W_{pq}} &= \frac{\partial s_{j,t}}{\partial W_{pq}} + \sum_l \frac{\partial s_{j,t}}{\partial s_{l,t-1}} \frac{\mathrm{d}\,s_{l,t-1}}{\mathrm{d}\,W_{pq}} \\ &= \frac{\partial s_{j,t}}{\partial W_{pq}} + \frac{\partial s_{j,t}}{\partial s_{j,t-1}} \frac{\mathrm{d}\,s_{j,t-1}}{\mathrm{d}\,W_{pq}} + \underbrace{\textstyle\sum_{l \neq j} W_{jl} \frac{\partial z_{l,t-1}}{\partial s_{l,t-1}} \frac{\mathrm{d}\,s_{l,t-1}}{\mathrm{d}\,W_{pq}}}_{\text{depends on all weights } W_{jl}}, \end{aligned} \tag{2}$$

where ∂ denotes direct dependency and d accounts for all (direct and indirect) dependencies, following the notation in [5]. While $\frac{\partial E}{\partial s_{j,t}}$, which considers only the direct contribution of the internal state of neuron j at time t to the loss, is easy to compute, the factor $\frac{\mathrm{d}\,s_{j,t}}{\mathrm{d}\,W_{pq}}$ is a memory trace of all

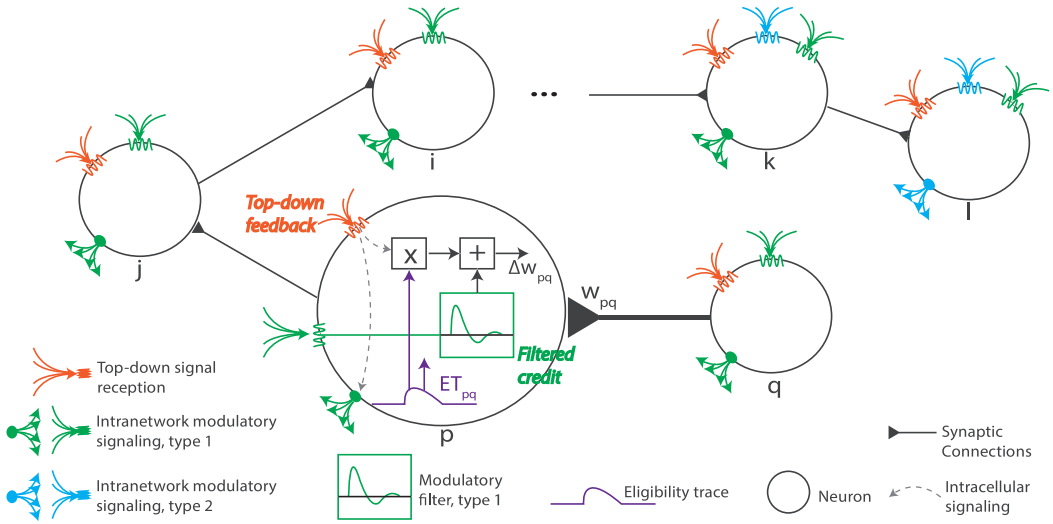

Figure 1: **Biologically-plausible temporal credit assignment via modulatory and synaptic message passing.** In addition to established biological learning ingredients (eligibility traces and a top-down learning signal [54, 8]), synapse-type-specific local modulatory networks may also be involved in weight updates [25]. Our learning rule, ModProp, conceives the action of participating modulators on receiving cells as a convolution of the eligibility trace with causal, time-invariant, cell-type specific filters. Each circle represents a neuron and the synaptic weight of interest is W_{pq}; we illustrate the cellular processes of postsynaptic neuron p. Our derivation (Supp. Section A and Theorem 1) predicts that the modulatory signal each neuron receives can represent a filtered credit signal regarding how its past firing (arbitrary steps back) contributes to the task outcome.

inter-cellular dependencies and it requires $O(N^3)$ memory and $O(N^4)$ computations. This makes RTRL expensive to implement for large networks. Moreover, this last factor $\frac{\mathrm{d}\, s_{j,t}}{\mathrm{d}\, W_{pq}}$ poses a serious problem for biological plausibility: it involves nonlocal terms in Eq. 14, so that knowledge of all other weights in the network is required in order to update the weight W_{pq}. Existing biologically-plausible solutions to this problem apply severe truncations: references [4] and [5] completely ignore the third nonlocal term (Figure 2), whereas reference [6] restores terms within one recurrent step through local modulatory signaling but truncates further terms.

To move beyond severe truncations of the exact gradient while retaining biological plausibilty, we propose two approximations to Eq. 14. **Approximation 1** replaces the activation derivative with a constant:

$$\frac{\mathrm{d}\, z_{j,t}}{\mathrm{d}\, s_{j,t}} \approx \mu \quad \text{(approximation 1)}, \tag{3}$$

where μ represents average activity of neurons. This approximation assumes stationarity in neuron activity and uncorrelatedness of such activity with the weight, as explained in derivations leading up to Eq. 20 and Eq. 21 (Methods in Appendix A). In our simulations, we treat μ as a hyperparameter and explain how this is tuned in Appendix E. This approximation enables the filter taps (Eq. 6) to be *time-invariant*, a property likely required for biological plausibility. We also define S, the arbitrary number of credit propagation steps, and it will become clear later that this corresponds to the number of modulatory filter taps. With this, the estimated gradient becomes:

$$\frac{\mathrm{d}\, E}{\mathrm{d}\, W_{pq}}|_t \approx \frac{\partial E}{\partial z_{p,t}} e_{pq,t} + \sum_j \frac{\partial E}{\partial s_{j,t}} \sum_{s=1}^{S} (W^s)_{jp}\, \mu^{s-1}\, e_{pq,t-s} \tag{4}$$

where we defined $e_{pq,t} := \frac{dz_{p,t}}{dW_{pq}}$, which can be interpreted as a persistent Hebbian "eligibility trace" [10, 9, 8] that keeps a fading memory of past coincidental pre- and postsynaptic activity [5]. Here, $(W^s)_{jp}$ represents the $(jp)^{\text{th}}$ entry of W^s, W raised to the power of s, and $(W^1)_{jp} = W_{jp}$ Details of the derivation can be found in Appendix A.

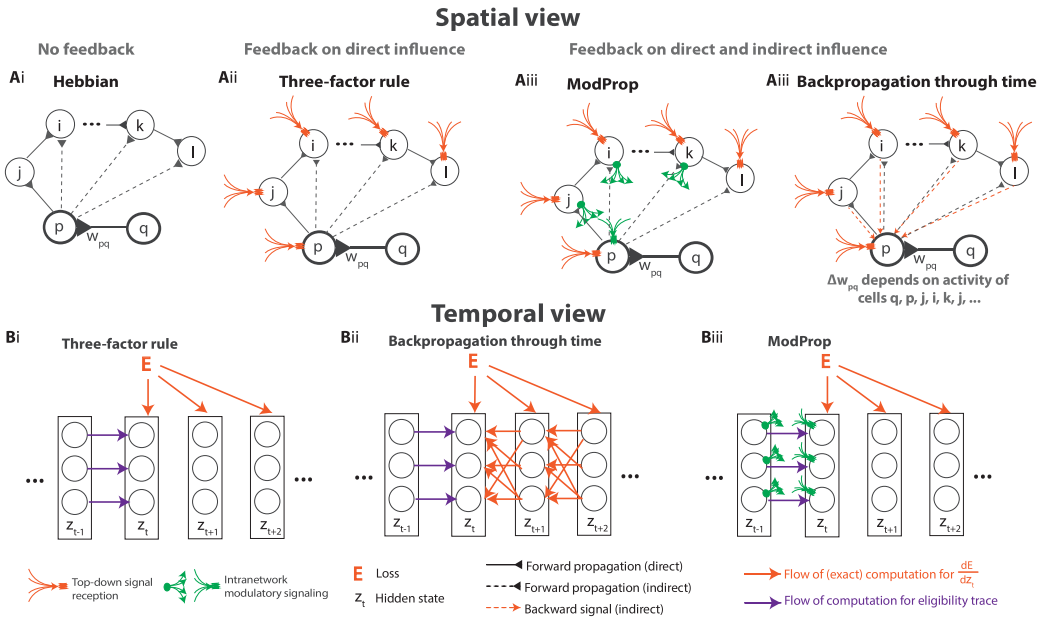

Figure 2: **Local modulatory signaling for gradient estimation.** A) A spatial view of learning rules for updating weight W_{pq}. (i) Hebbian learning, where weight update depends only on pre-/post-synaptic activities. (ii) Three-factor learning [54, 4, 5], which updates weights using additional top-down learning signals, severely truncates the exact gradient. (iii) ModProp also accounts for (filtered) distant feedback information delivered through synapse-type-specific neuromodulation; (iv) BPTT computes the exact gradient: weight update involves nonlocal information, i.e. activities of indirectly connected units. B) A temporal view. Bi) Three factor learning rule neglects all the intercellular dependencies in the temporal propagation of the credit signal. Bii) BPTT propagates the precise intercellular dependencies in an acausal manner. Biii) ModProp *approximates* such spatiotemporal dependencies through local neuromodulatory signals (Eq. 6). ModProp approximates the exact gradient by assuming similar connectivity among cells of the same type, and filtering the indirect effects on loss from neurons that are potentially many synapses away.

Approximation 2 replaces *synapse-specific* feedback weights with *type-specific* weights:

$$(W^s)_{jp} \approx (W^s)_{\alpha\beta} \quad \text{(approximation 2)}. \tag{5}$$

Here, cell j belongs to type α, cell p is of type β and C denotes the set of cell types. (e.g., $W_{\alpha\beta} = \mathbb{E}_{j\in\alpha, p\in\beta}[W_{jp}]$, $\alpha, \beta \in C$.) This approximation is due to the type-specific nature of modulatory channels [23]. We call these modulatory weights synapse-type-specific (as opposed to cell-type-specific) to emphasize connectivity-based grouping. Details of how these modulatory weights were obtained can be found in Appendix A.

Substituting **Approximation 1** and **Approximation 2** into Eq. 14 leads to the ModProp update:

$$\Delta W_{pq}|_{ModProp} \propto L_p \times e_{pq} + \left(\sum_{\alpha\in C} \left(\sum_{j\in\alpha} \mathrm{L}_j \text{ activity}_j \right) \times F_{\alpha\beta} \right) * \mathrm{e}_{pq},$$
$$F_{\alpha\beta,s} = \mu^{s-1}(W^s)_{\alpha\beta}, \tag{6}$$

where L and e denote top-down learning signal and eligibility trace, respectively. We postulate that $F_{\alpha\beta}$ represents type-specific filter taps of GPCRs expressed by cells of type β to precursors secreted by cells of type α; $*$ is the convolution operation with S as the number of filter taps. Details of biological interpretation can be found in Appendix C.1.

In summary, we have proposed a synaptic weight update rule, where the eligibility trace is compounded not only with top-down learning signals —- as in modern biologically-plausible learning rules [7, 8] —- but also with local modulatory pathways through convolution (Figure 1). This modulatory mechanism allows the propagation of credit signal through arbitrary number of recurrent steps.

3.2 Properties of ModProp

Remark 1. *The biologically plausible implementation (Eq. 6) of ModProp complexity scales as $O(SN^2)$ per time step t.*

Here, N and S denote the number of recurrent units and number of filter taps. As seen in Eq. 4, the number of filter taps corresponds to the number of recurrent steps for which the credit information is propagated. We also present an alternative implementation with potentially lower cost later in Proposition 1. Details for cost analysis and biological interpretation can be found in Appendix C.1.

We next show through Theorem 1 that learning with synapse-type-specific weights leads to reduction in loss at every step on average. The intuition is that by assuming a statistical connection between synaptic weights and feedback modulatory weights, ModProp should have a positive component in the gradient direction. Effectiveness of our approximation is further demonstrated by comparing the alignment angle with the exact gradient across different learning algorithms (Figure 8).

In the following, we define the "residual weights" away from the cell-type averages as $\epsilon_{ij} := W_{ij} - W_{\alpha\beta}$ and $(\epsilon_s)_{ij} := (W^s)_{ij} - (W^s)_{\alpha\beta}$, where $i \in \alpha, j \in \beta$, $\alpha, \beta \in C$. We consider these terms as stochastic so that the circuit output and the eligibility traces are also stochastic, as functions of ϵ_s. We consider the cell type averages and the output connection strengths as deterministic. (Merely modifying the uncorrelatedness assumption below would extend our results to stochastic output connection strengths.) Below, $\mathbb{E}$ is short for $\mathbb{E}_\epsilon$, $\Delta E|_t$ denotes the change in the task error at time t and $\Delta E|_{pq,t}$ denotes the contribution of the synapse (pq) to it.

Theorem 1. *Consider linear RNNs $\mathcal{N}$, $\widehat{\mathcal{N}}$ with weight matrices W, $\widehat{W}$, respectively and identical architectures. Let $\widehat{W}_{ij} = W_{\alpha\beta}$, $\forall i \in \alpha, j \in \beta$. Assume a small enough learning rate such that the remainder term of the first order Taylor expansion of loss is negligible. For network $\mathcal{N}$, if $\mathbb{E}[\epsilon_{ij}] = 0$, $(\epsilon_s)_{ik}$ and ϵ_{kj} are uncorrelated for $i \neq j$, and ϵ_s and $(y_t - y_t^*)^2 e_{pq,t'} e_{pq,t-s}$ are uncorrelated for any s, t, t', then $\mathbb{E}[\Delta E|_{pq,t}] \leq 0$ and $\mathbb{E}[\Delta E|_t] \leq 0$. Moreover, $\mathbb{E}[\Delta E|_t] < 0$ if gradient descent is possible for network $\widehat{\mathcal{N}}$.*

Proof. Proof is in Appendix D. □

Note that the uncorrelatedness assumption above is relatively mild because any single connection strength is typically a very poor predictor of network activity, especially as the network size grows. Note also that gradient update can introduce a drift in residual weights, requiring a similar drift in cell-type averages for the strict inequality to be applicable over multiple update steps.

3.3 Simulation results

To test the ModProp formulation, we study its performance in well-known tasks involving temporal processing: pattern generation, delayed XOR, and sequential MNIST. We compare the learning performance of ModProp with the state-of-the-art biologically-plausible learning rules (MDGL [6] and e-prop [5] (labelled as "three-factor"), as well as BPTT to provide a lower bound on task error. Consistent with [6], we also found MDGL to confer little advantage over e-prop when all neurons are connected to the readout (Figure 3), so we focus our analysis on that case.

We first study pattern generation with RNNs, where the aim is to produce a one-dimensional target output, generated from the sum of five sinusoids, given a fixed random Gaussian input realization. We change the target output and the random input along with the initial weights for different training runs, and illustrate the learning curve in Figure 3A across seven such runs. By using a densely connected network, we observe that MDGL confers little advantage over the three-factor rule (e-prop), as reported in the original MDGL paper [6]. Moreover, communicating longer indirect effects, despite being filtered, leads to improved learning outcomes, as demonstrated by the superior performance of ModProp over MDGL.

Next, to study how RNNs can learn to process discrete cues that impact delayed outcomes, we consider a delayed XOR task. Here, two cue alternatives, 1 or 0, are encoded by the presence/absence of input. The network is trained to remember the first cue and learn to compare it with the second cue delivered at a later time to determine if the two cues match or mismatch. Learning curves for this task are illustrated in Figure 3B, and we observe the same general conclusion as the previous task.

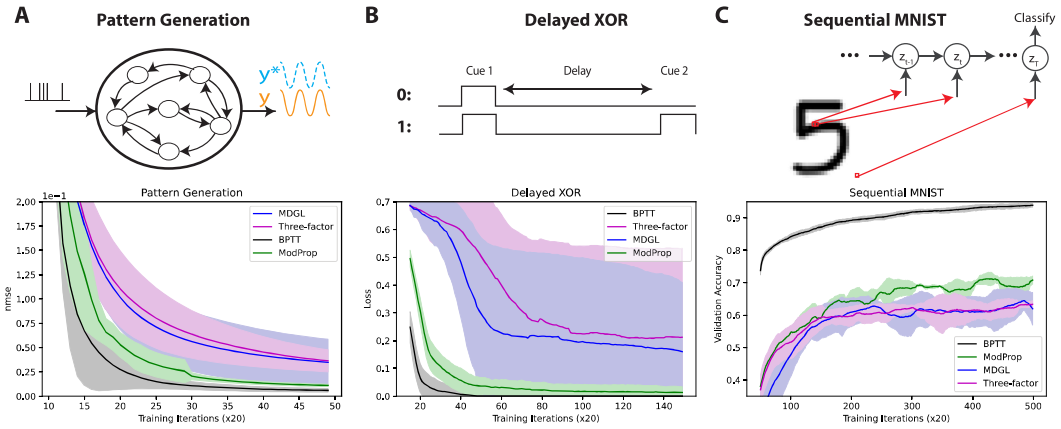

Figure 3: **Modulatory signaling of credit information on long-term recurrent interactions can improve learning outcomes.** ModProp improves the learning performance over existing bio-plausible rules. This experiment examines the performance due to Approximation 1 (Eq. 3) before any cell-type approximation of modulatory weights (Eq. 5). See Fig. 4 for the effect of cell-type approximation. A) Learning to produce a time-resolved target output pattern. B) A delayed XOR task, where the network determines if two cue alternatives — the presence or absence of input represented by 1 or 0 — match or mismatch after a delay, requiring memory via recurrent activity. c) Pixel-by-pixel MNIST task [56]. Note that this task is unlikely to be solved effectively by humans. (See text.) Consistent with the original MDGL paper [6], we also find that MDGL confers little advantage over the three-factor rule (e-prop) under dense connectivity. Solid lines/shaded regions: mean/standard deviation of loss curves across five runs.

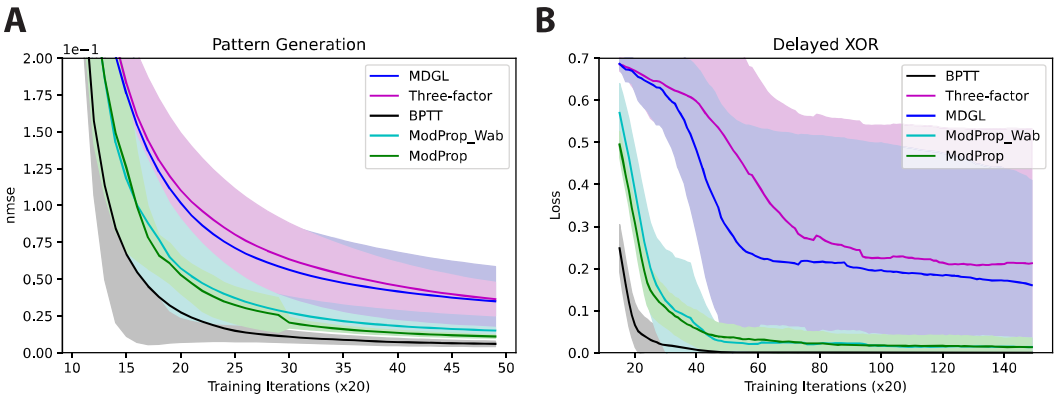

Figure 4: **Efficient learning with type-specific modulatory weights.** In addition to Approximation 1 studied in Figure 3, this figure investigates the effect of Approximation 2 (labeled as ModProp_Wab), which uses type-specific, rather than synapse-specific feedback weights for signaling credit information (Eq. 5). Here, ModProp_Wab uses only two modulatory types mapped to the two main cell classes. The cell type approximation does not result in any significant performance degradation in A) the pattern generation task and B) the delayed XOR task. This analysis is not done for the sequential MNIST task, where neurons were not divided into E and I types. Solid lines/shaded regions: mean/standard deviation of loss curves across five runs.

Some learning curves have a wide region of standard deviation, which indicates that the network struggled to learn the task with these rules for some seeds. Interestingly, the performance of ModProp approaches that of BPTT for this task and the previous one, further closing the gap between artificial network training and biological learning mechanisms in performing credit assignment over a long period.

Finally, we study the pixel-by-pixel MNIST [56] task, which is a popular machine learning benchmark task. Although sequential MNIST is not a task that the brain would solve well (i.e., humans would struggle to predict the digit with only one pixel presented at a time), we investigate it to test the limits of spatiotemporally filtered credit signals for tasks that demand temporally precise input integration. Learning curves for this task are illustrated in Figure 3C. While the performance ordering of learning rules is still the same as previous tasks, we observe a wider gap between ModProp and BPTT. Since the time-invariant filter approximation (Eq 3) restricts the spatiotemporal resolution of the credit signal, this is in line with our expectation that ModProp will struggle with tasks that demand highly precise spatiotemporal integration, such as the pixel-by-pixel MNIST task that even the brain would struggle to solve.

As a proof-of-concept study, we initially focused our analysis on the approximation performance by imposing time invariant filter taps (Eq. 3). Next, in Figure 4, we investigate the learning performance using type-specific rather than synapse-specific feedback modulatory weights (Eq. 5). These type-specific weights were calculated using weight averages as in [6] (we also investigate using fixed random modulatory weights later in Appendix Figure 6). Little performance degradation is observed for only two modulatory types mapped onto the two main cell classes, indicating the effectiveness of cell-type discretization. Other than the sequential MNIST task, cells are divided into two main cell classes — with 80% of the cells being excitatory (E) and 20% being inhibitory (I) — and obeyed connection sign constraints.

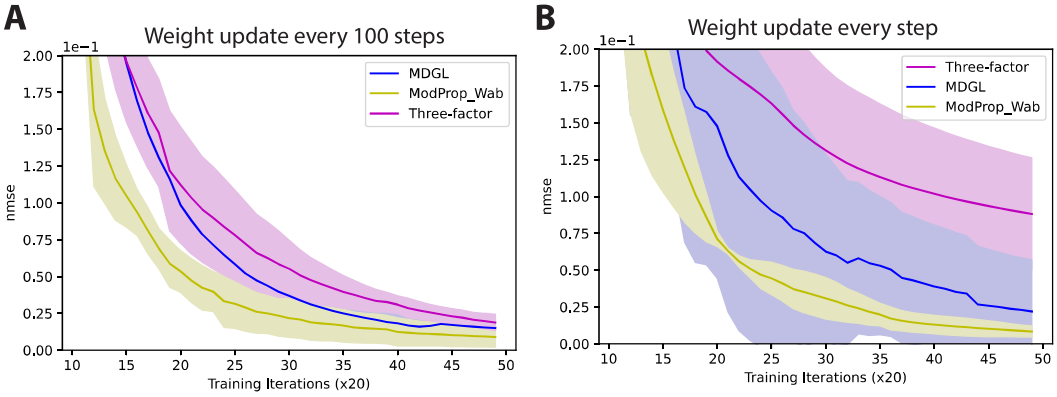

Figure 5: **Superior performance of ModProp in an online learning setting**. We investigate an online learning version of the pattern generation task, where weights are updated either A) every 100 time steps or B) every single step. Here, ModProp uses the efficient online learning implementation derived in Proposition 1. We omit comparisons with the exact gradient update because online learning requires the computation of the RTRL factorization, whose complexity is prohibitive. Plotting conventions follow those of previous figures.

Lastly, we demonstrate the advantage of our weight update rule in an online learning setting of the pattern generation task, where weights are updated in real time (Figure 5). For that, we also derive a cost and storage efficient *in silico* implementation of ModProp in Proposition 1.

Proposition 1. *ModProp has an online in-silico (not necessarily biologically-plausible) implementation with $O(CN^2)$ storage and $O(C^2N^2)$ computational complexity, where C is the number of cell types.*

Proof. Proof is in Appendix C.2. □

We note that the specific implementation outlined in the proof of Proposition 1 can significantly reduce the implementation cost, but is likely biologically-implausible, because each synaptic weight update requires the knowledge of all modulatory weights in the network (Appendix C). Moreover, it reduces the cost compared to RTRL ($O(N^3)$ storage and $O(N^4)$ computational complexity) as well as SnAP-2 ($O(d^2N^3)$ storage and $O(d^2N^4)$ computational complexity for connection density d) [52] significantly if only a few cell types are used. In this work, we used only two cell types ($C = 2$) that map onto the two main cell classes: excitatory and inhibitory.

4 Discussion

A central question in the study of biological and artificial intelligence is how the temporal credit assignment problem is solved in neuronal circuits [3]. Motivated by recent genomic evidence on the widespread presence of local modulatory networks [23, 24], we demonstrated how such signaling could complement Hebbian learning and global neuromodulation (e.g., via dopamine) to achieve biologically plausible propagation of the credit signal through *arbitrary* time steps. Going beyond the scalar feedback provided by global top-down modulation, our study proposes how detailed vector feedback information [1] can be delivered in neuronal circuits. From an approximation perspective, instead of the severe temporal truncations of the gradient information proposed by the state-of-the-art [4, 5, 6], ModProp offers a framework where the full temporal feedback signal can be received, albeit via low-pass filtering at the post-synaptic neuron due to specificity at the level of neuronal types, and not individual neurons.

While feedback alignment [43] addresses the weight transport problem in *feedforward* networks, it is not clear in RNNs which biological pathways would constitute the *temporal* feedback pathways. Our model suggests that such pathways could come from synapse-type-specific local neuromodulation. In addition to improved performance compared to the state-of-the-art across all of our simulations, our theoretical and experimental results show that ModProp can be implemented efficiently for online learning. Taken together, these findings suggest that synapse-type-specific local modulatory signaling could be a neural mechanism for propagating credit information over more than just a few recurrent steps.

Among the many future directions, a natural extension could be to investigate the performance of ModProp across a broader range of tasks. This could include finding situations where the assumptions in deriving the rule (e.g., stationarity of activity) are severely violated. Over time, this might improve the accuracy of the approximations introduced here. Along this line, we observed a significant gap between ModProp and BPTT for the pixel-by-pixel MNIST task that demands precise temporal integration of input (Figure 3C); but we have discussed earlier that sequential MNIST is challenging for the brain to solve. Additionally, this study focuses on dense network with ReLU activation; future directions include investigating two biologically relevant paradigms: sparse connectivity and diverse set of activation functions (spike-based in particular). Although ModProp can be applied in theory to temporal credit assignment over an arbitrary duration, the presented learning rule accounts for dependencies at every single step (as for BPTT). This means that, similar to BPTT, it is ill-suited for *very* long-term credit assignment [57]. An interesting line of research attends to the issue of extracting relevant, rather than full, memories [57, 58]. Investigating how our approximations could potentially be combined with memory sparsification techniques to perform very long-term biologically-plausible credit assignment can be fruitful [57].

While our paper advances the basic science of learning and we do not foresee immediate societal impacts of our work, its benefits to both neuroscience and deep learning research could have long term (positive or negative) societal impacts. It is hard to overstate the philosophical implications of understanding how the brain works (and, in particular, learns). Moreover, such understanding could guide us towards curing certain diseases of the brain. Nevertheless, in the same vein, such future tools could also enable abuse if left unregulated. On the machine learning side, our method can be considered as a low-cost alternative to the ubiquitous BPTT algorithm. In this sense, our algorithm or a future method that builds on it can make tasks that are currently approachable by just a few wealthy entities available to many more practitioners. On the flip side, as with many other capable machine learning tools, such entities are also capable of utilizing low-cost learning to conquer even harder tasks, which is potentially a reason for concern. Finally, data-driven tools will ultimately reflect the various biases in their training data and our method is no exception.

References

[1] Timothy P Lillicrap, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. Backpropagation and the brain. *Nature Reviews Neuroscience*, pages 1–12, 2020.

[2] Blake A Richards, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, et al. A deep learning framework for neuroscience. *Nature neuroscience*, 22(11):1761–1770, 2019.

[3] Timothy P. Lillicrap and Adam Santoro. Backpropagation through time and the brain. *Current Opinion in Neurobiology*, 55:82–89, apr 2019.

[4] James M. Murray. Local online learning in recurrent networks with random feedback. *eLife*, 8, may 2019.

[5] Guillaume Bellec, Franz Scherr, Anand Subramoney, Elias Hajek, Darjan Salaj, Robert Legenstein, and Wolfgang Maass. A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 11(1):1–15, dec 2020.

[6] Yuhan Helena Liu, Stephen Smith, Stefan Mihalas, Eric Shea-Brown, and Uygar Sümbül. Cell-type–specific neuromodulation guides synaptic credit assignment in a spiking neural network. *Proceedings of the National Academy of Sciences*, 118(51), 2021.

[7] Jeffrey C. Magee and Christine Grienberger. Synaptic Plasticity Forms and Functions. *Annual Review of Neuroscience*, 43(1):95–117, jul 2020.

[8] 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, jul 2018.

[9] Sho Yagishita, Akiko Hayashi-Takagi, Graham C.R. Ellis-Davies, Hidetoshi Urakubo, Shin Ishii, and Haruo Kasai. A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science*, 345(6204):1616–1620, sep 2014.

[10] Stijn Cassenaer and Gilles Laurent. Conditional modulation of spike-timing-dependent plasticity for olfactory learning. *Nature*, 482(7383):47–51, feb 2012.

[11] Magdalena Sanhueza and John Lisman. The CaMKII/NMDAR complex as a molecular memory. *Molecular Brain*, 6(1):10, feb 2013.

[12] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Computational Neuroscience Series, 2001.

[13] Wolfram Schultz. Neuronal reward and decision signals: from theories to data. *Physiological reviews*, 95(3):853–951, 2015.

[14] Jonathan E Rubin, Catalina Vich, Matthew Clapp, Kendra Noneman, and Timothy Verstynen. The credit assignment problem in cortico-basal ganglia-thalamic networks: A review, a problem and a possible solution. *European Journal of Neuroscience*, 53(7):2234–2253, 2021.

[15] Zuzanna Brzosko, Susanna B Mierau, and Ole Paulsen. Neuromodulation of spike-timing-dependent plasticity: past, present, and future. *Neuron*, 103(4):563–581, 2019.

[16] Wolfram Schultz. Dopamine reward prediction-error signalling: a two-component response. *Nature Reviews Neuroscience*, 17(3):183, 2016.

[17] Owen Marschall, Kyunghyun Cho, and Cristina Savin. A unified framework of online learning algorithms for training recurrent neural networks. *Journal of Machine Learning Research*, 21(135):1–34, 2020.

[18] Jean-Christophe Cassel. The role of serotonin in learning and memory: a rich pallet of experimental studies. In *Handbook of Behavioral Neuroscience*, volume 31, pages 549–570. Elsevier, 2020.

[19] Amjad H Bazzari and H Rheinallt Parri. Neuromodulators and long-term synaptic plasticity in learning and memory: A steered-glutamatergic perspective. *Brain sciences*, 9(11):300, 2019.

[20] Éva Borbély, Bálint Scheich, and Zsuzsanna Helyes. Neuropeptides in learning and memory. *Neuropeptides*, 47(6):439–450, dec 2013.

[21] Trevor J. Hamilton, Sara Xapelli, Sheldon D. Michaelson, Matthew E. Larkum, and William F. Colmers. Modulation of distal calcium electrogenesis by neuropeptide Y1 receptors inhibits neocortical long-term depression. *Journal of Neuroscience*, 33(27):11184–11193, jul 2013.

[22] Bosiljka Tasic, Zizhen Yao, Lucas T. Graybuck, Kimberly A. Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, Emma Garren, Michael N. Economo, Sarada Viswanathan, Osnat Penn, Trygve Bakken, Vilas Menon, Jeremy Miller, Olivia Fong, Karla E. Hirokawa, Kanan Lathia, Christine Rimorin, Michael Tieu, Rachael Larsen, Tamara Casper, Eliza Barkan, Matthew Kroll, Sheana Parry, Nadiya V. Shapovalova, Daniel Hirschstein, Julie Pendergraft, Heather A. Sullivan, Tae Kyung Kim, Aaron Szafer, Nick Dee, Peter Groblewski, Ian Wickersham, Ali Cetin, Julie A. Harris, Boaz P. Levi, Susan M. Sunkin, Linda Madisen, Tanya L. Daigle, Loren Looger, Amy Bernard, John Phillips, Ed Lein, Michael Hawrylycz, Karel Svoboda, Allan R. Jones, Christof Koch, and Hongkui Zeng. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729):72–78, nov 2018.

[23] Stephen J. Smith, Uygar Sümbül, Lucas T. Graybuck, Forrest Collman, Sharmishtaa Seshamani, Rohan Gala, Olga Gliko, Leila Elabbady, Jeremy A. Miller, Trygve E. Bakken, Jean Rossier, Zizhen Yao, Ed Lein, Hongkui Zeng, Bosiljka Tasic, and Michael Hawrylycz. Single-cell transcriptomic evidence for dense intracortical neuropeptide networks. *eLife*, 8, nov 2019.

[24] Stephen J Smith. Transcriptomic evidence for dense peptidergic networks within forebrains of four widely divergent tetrapods. *Current opinion in neurobiology*, 71:100–109, 2021.

[25] Stephen J. Smith, Michael Hawrylycz, Jean Rossier, and Uygar Sümbül. New light on cortical neuropeptides and synaptic network plasticity. *Current Opinion in Neurobiology*, 63:176–188, aug 2020.

[26] Anthony N. van den Pol. Neuropeptide Transmission in Brain Circuits. *Neuron*, 76(1):98–115, oct 2012.

[27] Yang Dan and Mu-Ming Poo. Spike timing-dependent plasticity: from synapse to perception. *Physiological reviews*, 86(3):1033–1048, 2006.

[28] Sen Song, Kenneth D Miller, and Larry F Abbott. Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature neuroscience*, 3(9):919–926, 2000.

[29] Marco P Lehmann, He A Xu, Vasiliki Liakoni, Michael H Herzog, Wulfram Gerstner, and Kerstin Preuschoff. One-shot learning and behavioral eligibility traces in sequential decision making. *Elife*, 8:e47463, 2019.

[30] Aparna Suvrathan. Beyond STDP — towards diverse and functionally relevant plasticity rules. *Current Opinion in Neurobiology*, 54:12–19, feb 2019.

[31] Michael A. Farries and Adrienne L. Fairhall. Reinforcement Learning With Modulated Spike Timing–Dependent Synaptic Plasticity. *Journal of Neurophysiology*, 98(6):3648–3665, dec 2007.

[32] Verena Pawlak, Jeffery R Wickens, Alfredo Kirkwood, and Jason ND Kerr. Timing is not everything: neuromodulation opens the stdp gate. *Frontiers in synaptic neuroscience*, 2:146, 2010.

[33] Pieter R. Roelfsema and Anthony Holtmaat. Control of synaptic plasticity in deep cortical networks. *Nature Reviews Neuroscience*, 19(3):166–180, feb 2018.

[34] Alexandre Payeur, Jordan Guerguiev, Friedemann Zenke, Blake A Richards, and Richard Naud. Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nature neuroscience*, pages 1–10, 2021.

[35] Johnatan Aljadeff, James D'amour, Rachel E Field, Robert C Froemke, and Claudia Clopath. Cortical credit assignment by hebbian, neuromodulatory and inhibitory plasticity. *arXiv preprint arXiv:1911.00307*, 2019.

[36] Nicolas Frémaux and Wulfram Gerstner. Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9(JAN2016):85, jan 2015.

[37] Răzvan V Florian. Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural computation*, 19(6):1468–1502, 2007.

[38] Nicolas X. Tritsch and Bernardo L. Sabatini. Dopaminergic Modulation of Synaptic Transmission in Cortex and Striatum. *Neuron*, 76(1):33–50, oct 2012.

[39] Eve Marder. Neuromodulation of neuronal circuits: back to the future. *Neuron*, 76(1):1–11, 2012.

[40] Blake A. Richards and Timothy P. Lillicrap. Dendritic solutions to the credit assignment problem. *Current Opinion in Neurobiology*, 54:28–36, feb 2019.

[41] R. J. Williams and D. Zipser. Gradient-based learning algorithms for recurrent networks and their computational complexity. In Y. Chauvin and D. E. Rumelhart, editors, *Back-propagation: Theory, Architectures and Applications*, chapter 13, pages 433–486. Hillsdale, NJ: Erlbaum, 1995.

[42] Drew Linsley, Alekh Karkada Ashok, Lakshmi Narasimhan Govindarajan, Rex Liu, and Thomas Serre. Stable and expressive recurrent vision models. *arXiv preprint arXiv:2005.11362*, 2020.

[43] 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):1–10, 2016.

[44] Isabella Pozzi, Sander Bohté, and Pieter Roelfsema. A biologically plausible learning rule for deep learning in the brain. *arXiv preprint arXiv:1811.01768*, 2018.

[45] João Sacramento, Rui Ponte Costa, Yoshua Bengio, and Walter Senn. Dendritic cortical microcircuits approximate the backpropagation algorithm. *arXiv preprint arXiv:1810.11393*, 2018.

[46] Axel Laborieux, Maxence Ernoult, Benjamin Scellier, Yoshua Bengio, Julie Grollier, and Damien Querlioz. Scaling equilibrium propagation to deep convnets by drastically reducing its gradient estimator bias. *Frontiers in neuroscience*, 15:129, 2021.

[47] Yali Amit. Deep learning with asymmetric connections and hebbian updates. *Frontiers in computational neuroscience*, 13:18, 2019.

[48] Beren Millidge, Alexander Tschantz, Anil K Seth, and Christopher L Buckley. Activation relaxation: A local dynamical approximation to backpropagation in the brain. *arXiv preprint arXiv:2009.05359*, 2020.

[49] Asier Mujika, Florian Meier, and Angelika Steger. Approximating Real-Time Recurrent Learning with Random Kronecker Factors. In *32nd Conference on Neural Information Processing Systems*, pages 6594–6603, 2018.

[50] Corentin Tallec and Yann Ollivier. Unbiased Online Recurrent Optimization. In *ICLR*, feb 2018.

[51] Christopher Roth, Ingmar Kanitscheider, and Ila Fiete. Kernel RNN Learning (KERNL). In *ICLR*, sep 2019.

[52] Jacob Menick, Erich Elsen, Utku Evci, Simon Osindero, Karen Simonyan, and Alex Graves. A practical sparse approximation for real time recurrent learning. *arXiv preprint arXiv:2006.07232*, 2020.

[53] Friedemann Zenke and Emre O Neftci. Brain-inspired learning on neuromorphic substrates. *arXiv preprint arXiv:2010.11931*, 2020.

[54] Nicolas Frémaux and Wulfram Gerstner. Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in neural circuits*, 9:85, 2016.

[55] Daniel B Ehrlich, Jasmine T Stone, David Brandfonbrener, Alexander Atanasov, and John D Murray. Psychrnn: An accessible and flexible python package for training recurrent neural network models on cognitive tasks. *Eneuro*, 8(1), 2021.

[56] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

[57] Guangyu Robert Yang and Manuel Molano Mazon. Next-generation of recurrent neural network models for cognition. 2021.

[58] Timothy P Lillicrap and Adam Santoro. Backpropagation through time and the brain. *Current opinion in neurobiology*, 55:82–89, 2019.

[59] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *ICLR*. International Conference on Learning Representations, ICLR, dec 2015.

[60] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. {TensorFlow}: A system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pages 265–283, 2016.

[61] Guillaume Bellec, Darjan Salaj, Anand Subramoney, Robert Legenstein, and Wolfgang Maass. Long short-term memory and learning-to-learn in networks of spiking neurons. In *32nd Conference on Neural Information Processing Systems*, pages 787–797, 2018.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] We have supported our key contribution points (see Introduction) with simulations and mathematical proofs.
   (b) Did you describe the limitations of your work? [Yes] Please see discussion on future direction in the Discussion section.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] Please see the Discussion section.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes] We have carefully read the ethics review guidelines and confirm that this paper conforms to them.
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] We have made sure that all assumptions are precisely stated in the Theorem statement. Details for where these assumptions come into the proof can be found in Appendix D.
   (b) Did you include complete proofs of all theoretical results? [Yes] We have provided complete and detailed proofs in Appendix D.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Anonymized code link is provided in Appendix E.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] Please see Appendix E.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] We ran experiments across random seeds and reported uncertainty for all applicable figures. We explained the plotting convention in the caption.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Please see Appendix E.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [Yes] Please see Appendix E.

(b) Did you mention the license of the assets? [Yes] Please see Appendix E.

(c) Did you include any new assets either in the supplemental material or as a URL? [No]

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]