Brains learn tasks via experience-driven differential adjustment of their myriad individual synaptic connections, but the mechanisms that target appropriate adjustment to particular connections remain deeply enigmatic. While Hebbian synaptic plasticity, synaptic eligibility traces, and top-down feedback signals surely contribute to solving this synaptic credit-assignment problem, alone, they appear to be insufficient. Inspired by new genetic perspectives on neuronal signaling architectures, here, we present a normative theory for synaptic learning, where we predict that neurons communicate their contribution to the learning outcome to nearby neurons via cell-type-specific local neuromodulation. Computational tests suggest that neuron-type diversity and neuron-type-specific local neuromodulation may be critical pieces of the biological credit-assignment puzzle. They also suggest algorithms for improved artificial neural network learning efficiency.

Significance

Synaptic connectivity provides the foundation for our present understanding of neuronal network function, but static connectivity cannot explain learning and memory. We propose a computational role for the diversity of cortical neuronal types and their associated cell-type-specific neuromodulators in improving the efficiency of synaptic weight adjustments for task learning in neuronal networks.

Author contributions: U.S. conceived initial theory; Y.H.L., S.S., S.M., E.S.-B., and U.S. performed research; Y.H.L. analyzed data; and Y.H.L., S.S., E.S.-B., and U.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1 To whom correspondence may be addressed. Email: hyliu24@uw.edu or uygars@alleninstitute.org.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2111821118/-/DCSupplemental.
that neurons utilize modulatory networks to actively broadcast their own contribution to the network performance to nearby neurons via cell-type–specific local neuromodulation (Figs. 1D and 2C)—specifically, each cell broadcasts its own direct contribution to the overall task “error” signal. This is a major departure from more global roles for modulators previously proposed, such as carrying error or reward signals. From a neuroscience perspective, our study proposes a model of cortical learning shaped by the interplay of local modulatory signaling carrying credit-assignment information and synaptic transmission and potentially brings us closer to understanding biological intelligence. From a computer science perspective, our method offers a significantly smaller number of interconnects for on-chip neuro-inspired artificial intelligence.
A conventional three-factor local learning rule models action of a “third,” TD GPCR-activating ligand (e.g., dopamine) that governs synapse reweighting ($\Delta w$) in proportion to temporal coincidence of the two Hebbian factors (presynaptic and postsynaptic activity). Such models generally require a lingering ET to sustain information about Hebbian coincidence until arrival of the TD signal. 

Embracing new genetic evidence for local GPCR-based modulatory machinery, the MDGL theory introduces additional factors that allow spike-dependent secretion of NP-like local modulators (LMe from excitatory neurons and LMi from inhibitory neurons) to participate in governing synapse reweighting ($\Delta w$) (35). As indicated here and in Fig. 1, the present MDGL model comprises both directly TD-recipient cells (types D–F; B, Left) and non–TD-recipient cells (types A–C; B, Right). Synapse reweighting requires combined GPCR activation with a persistent ET for all cell types, but GPCRs are activated on non–TD-recipient cells only by the local modulatory ligands. 

Propagation of TD error/reward signal via spike-dependent secretion of local modulators from both excitatory and inhibitory cell types to cells lacking direct access to TD modulatory signal. For simplicity, this schema represents only the four subscripted synapses/weights, while the full model represents many more synaptic inputs per cell.

As illustrated in Fig. 2A, the resulting three-factor local learning rules represent this presynaptic and postsynaptic information as a time-dependent eligibility trace (ET) and combine it with TD signals to update the weight $\Delta w$ of each synapse (31–34) (detailed further in Methods; see Figs. 5B and 6 D, ii; see also ref. 7).

We propose a role for cell-type-based modulatory signals in recovering a key part of the error-gradient information that is lost by dropping nonlocal terms in such conventional three-factor rules. We describe this in terms of the update $\Delta w$ to synapse $p\mid q$ from neuron $q$ to neuron $p$ with strength $w$. To begin, we consider the contributions to the error gradient of those neurons that are one synapse away from the postsynaptic site (i.e., neurons $j$ such that synapse $j\mid p$ exists) (illustrated in Methods; see Fig. 5C). We find that this set of contributions can be realized by activity-dependent signals emitted by neurons $j$ and the ET for the synapse $p\mid q$ (Eq. 19). Intriguingly, this signal is precisely neuron $j$’s contribution to the task error, thereby taking into account the indirect contribution of the synapse $p\mid q$ to the network performance via neurons $j$ for more accurate synaptic credit assignment.
This initial form, however, still requires the knowledge of cell-specific signals from cells $j$ not participating in the synapse of interest. We further make the key observation that when just the contributions from cells up to two synapses away are considered, those terms only appear under a sum. The mechanism updating the synapse $p/q$ does not need to know the contributions from individual “indirect” neurons $j$, as their sum suffices. This observation is critical in elucidating a role for diffuse neuromodulatory signaling in carrying this summed, indirect signal and thus serving as an additional factor in synaptic plasticity.

To fully remove cell-specific dependencies in the indirect signal, we further approximate the cell-specific weights $w_{jp}$ that it contains with the cell-type–specific terms $w_{\alpha\beta}$ when postsynaptic cell $j$ belongs to type $\alpha$ and presynaptic cell $p$ belongs to type $\beta$. We postulate that $w_{\alpha\beta}$ represents the affinity of GPCRs expressed by cells of type $\beta$ to peptides secreted by cells of type $\alpha$ (Fig. 1C and Methods, Eq. 20; see Fig. 5D), and this type-cell-specific variable is genetically determined. The local diffusion assumption (35) suggests a further idealization, where this type of signaling is registered only by local synaptic partners and therefore preserves the connectivity structure of $w_{jp}$ (Eq. 23). It is also worth noting that the rich set of ligand and receptor types with different downstream actions (30) can support that $w_{\alpha\beta}$ is a signed term.

Bringing these together, we have the learning rule illustrated in Fig. 2B. The weight update is given as

$$\Delta w_{pq} \propto \left( TD_p + \sum_{\alpha \in C} \sum_{j \in C} w_{\alpha\beta} \right) \times ET_{pq} \times \sum_{j \in C} (activity_j) \times (modulatory signal_j) \tag{1}$$

where neuron $j$ is of type $\alpha$ and neuron $p$ is of type $\beta$, $p \rightarrow j$ denotes that synapse $j/p$ exists, $C$ denotes the set of neuronal cell types, $TD_p$ denotes the TD signal received by $p$, $ET_{pq}$ denotes the ET for $p/q$, $w_{\alpha\beta}$ denotes the effect of ligands secreted by class $\alpha$ neurons on class $\beta$ receptors, and $LM_{\alpha\beta}$ denotes the contribution of local modulation to synaptic plasticity that has been ignored so far. Thus, our update rule suggests a set of modulatory terms that combine with the ET in order to more accurately assign credit across a network when updating its synapses. Neurons that receive TD feedback regarding their role on the circuit goals propagate this information to nearby synaptic partners via cell-type–specific local modulatory signals (Fig. 2C; see also Eq. 24 for details). Specifically, the modulatory signal $j$ in Eq. 1 is precisely the contribution of cell $j$ to the task error, as measured by the (partial) derivative of the error with respect to cell $j$’s membrane potential. Moreover, this framework proposes that cell-type–specific GPCR affinities allow these local signals to be informative without the need to know precise synaptic weights. The ability to assess the indirect impact of neurons on the overall loss via such communication is critical to accurate synaptic reweighting and improved performance over existing biologically plausible rules, as we demonstrate next (Fig. 3 and SI Appendix, Fig. S2).

In summary, we have proposed a rule for updating a synapse $w_{pq}$, which we refer to as the multidigraph learning rule, or MDGL, where the Hebbian ET is compounded not only with TD learning signals—as in modern biologically plausible learning rules (31, 32)—but also with cell-type–specific, diffuse modulatory signals.

**Simulation Framework for Testing Multidigraph Learning in Recurrent Spiking Neural Networks.** To test the MDGL formulation, we study its performance in recurrent spiking neural networks (RSNNs) learning well-known tasks involving temporal processing: pattern generation, delayed match to sample, and evidence accumulation. We use two main cell classes, inhibitory (I) and excitatory (E) cells, and obey experimentally observed constraints (e.g., refractoriness, synaptic delay, and connection sparsity). We further endow a fraction of the E cells with threshold adaptation (37). This mimics the hierarchical structure of cell types (38) through the simple example of two main cell types, one of which has two subtypes (E cells with and without threshold adaptation). The existence of synaptic connections to output neurons further divides each population into two, thus bringing the cell type tally to six in our conceptual model (Fig. 1). Our implementation does not involve random and rapid formation of new synapses after each experience (39), further increasing its biological plausibility.

We compare the learning performance of MDGL (Fig. 2B) with the state-of-the-art biologically plausible learning rule [e-prop (22)] (Fig. 2A and SI Appendix, Fig. S1). As a three-factor rule, e-prop does not involve local cell-type–specific signaling and restricts the update to depend only on presynaptic and postsynaptic activity, as well as a TD instructive signal. To provide a lower bound on task error, we also compare performance with BPTT (see Fig. 5.4), which uses exact error gradients to update weights. These learning rules are further illustrated in Methods; see Figs. 5A, B, and D.

**Multidigraph Learning Guides Temporal Credit Assignment in Benchmark Tasks.** We first study pattern generation with RSNNs, where the aim is to produce a one-dimensional target output, generated from the sum of five sinusoids, given a fixed Poisson input realization (40). We change the target output and the Poisson input along with the initial weights for different training runs (SI Appendix, Fig. S3A) and illustrate the learning curve in Fig. S4A across five such runs. We observe that MDGL performs significantly better than e-prop.

Next, to study how RSNNs can learn to process discrete cues that impact delayed rewards, we consider a delayed match to sample task (41). Here, two cue alternatives are encoded by the presence/absence of input spikes. The RSNN is trained to remember the first cue and learn to compare it with the second cue delivered at a later time (SI Appendix, Fig. S3B). Fig. 3B and SI Appendix, Figs. S4B and S5 display the learning curve for novel inputs. We observe that the same general conclusions as for the pattern-generation task hold; introducing cell-type–specific neuromodulation significantly improves learning outcomes.

Finally, we study an evidence-accumulation task (29), which involves integration of several cues to produce the desired output at a later time: A simulated agent moves along a path while encountering a series of sensory cues presented either on the right or left side of a track (Fig. 3C and SI Appendix, Figs. S3C and S4C). When it reaches a T-junction, it decides if more cues were received on the left or right. We test our learning rule to see if the addition of diffuse modulatory signals can indeed bring the learning curve closer to BPTT, without relying on rapid and random rewiring (39). Fig. 3C demonstrates that the performance trends of the previous two experiments continue to hold. SI Appendix, Fig. S6 illustrates that without threshold adaptation and recurrent connectivity, the network cannot significantly decrease loss and thus is unable to learn this task. In line with these experiments, gradients approximated by MDGL are more similar to the exact gradients (SI Appendix, Fig. S2), shedding light on its superior performance. We also observe that MDGLs performance depends only weakly on the hypothesized link (Eq. 23) between abstract cell-type-based connectivities and modulatory receptor affinities (SI Appendix, Fig. S7), enabling flexible implementations in vivo and in silico.
Fig. 3. Cell-type-specific neuromodulation guides learning across multiple tasks. (A) Learning to produce a time-resolved target output pattern. (B) A delayed match to sample task, where two cue alternatives are represented by the presence/absence of input spikes. (C) An evidence-accumulation task (29, 36). (Lower) Addition of cell-type-specific modulatory signals improves learning outcomes across tasks. In line with these results, SI Appendix, Fig. S2 shows that gradients approximated by MDGL are more similar to the exact gradients than those approximated by e-prop. Solid lines/shaded regions: mean/SD of loss curves across runs (Methods).

Discussion

Here, we have presented a multidigraph theory and instantiated simple models based on this theory that explicitly represent diverse neuron types classified by their synaptic and neuromodulatory connections. Simulations based on these simple models show that diverse signaling modes can facilitate credit assignment and enhance learning. A wealth of new genetic data provide strong support for the biological plausibility of this array of signaling modes and furthermore argue strongly that most or all modern eumetazoans (all multicellular animals except sponges) comprise numbers of cell types and modulatory signals far in excess of those represented in our simulations (43). We believe therefore that conceiving of neuronal networks as multidigraphs, involving multiple modulatory and synaptic signals, integrated by discrete cell-type nodes, may offer fruitful paths toward the understanding of synaptic credit assignment in biological neuronal networks. This multidigraph theory may also lead to more computationally efficient local learning rules for neural-network-based artificial intelligence.

In addition to established elements of Hebbian plasticity, ETs, and reward feedback signals, our normative theory posits important roles for neuronal cell-type diversity and local neuromodulatory communication in enabling efficient synaptic credit assignment. In particular, our findings predict that neurons secrete information about TD feedback signals they receive to nearby neurons in an activity-dependent and cell-type–specific manner using local modulation. As a consequence, levels of local modulatory signals may reflect the learning process. Indeed, our computational experiments imply that the level of modulatory input decreases over training and sharply rises in response to changes in task condition (SI Appendix, Fig. S11). It
is also interesting to note that phylogenomic studies now suggest that peptidergic neuromodulation may evolutionarily predate dopamine signaling (43) and thus may have actually provided the foundation upon which dopaminergic TD signaling evolved.

The nature of “intermediate” cells (38), whose phenotypes appear to be a mixture of “pure” cell types, is a key problem in cell-types research. Our findings may explain the existence of such phenotypes from a connectivity perspective: While average connectivities between types remain relatively constant during training, connectivities of individual cells can deviate significantly from those averages (SI Appendix, Fig. S12). We hypothesize a link between abstract cell-type-based connectivities and modulatory receptor affinities, where the average synaptic connection weights between types are taken to be the cell-type–specific modulatory receptor affinities (Eq. 23). How tightly the individual synaptic weights and cell-type–specific receptor affinities should be coupled may be explored in future work. SI Appendix, Fig. S2 indicates that even with imprecise GPCR affinities, MDGL can still improve gradient approximation, while SI Appendix, Fig. S7 suggests that the effect of imprecise GPCR affinities on the performance is task-dependent.

Learning rules often explicitly minimize a loss function, and the error gradient, if available, tells exactly how much each network parameter should be adjusted to reduce this loss function. Rules that follow this gradient, RTRL and BPTT, are well established, but are not biologically plausible and have unmanageably vast memory storage demands. However, a growing body of studies have demonstrated that learning rules that only partially follow the gradient, while alleviating some of these problems of the exact rules, can still lead to desirable outcomes (44, 45).

An example is the seminal concept of feedback alignment (14), which rivals backpropagation on a variety of tasks, even using random feedback weights for credit assignment. In addition, approximations to RTRL have been proposed (4–9, 21) for efficient online learning in recurrent neural networks. Our learning rule has $O(N^2)$ complexity, where $N$ is the number of neurons, which is less expensive than SnAp-2 that has a storage cost of $O(N^3)$ (8) (for simplicity, connection sparsity factor is neglected here). It also outperforms biological learning rules with similar complexity scale (21, 22). Thus, our model further advances approximated gradient-based learning methods and continues the line of research in energy-efficient, on-chip learning through spike-based communications (10, 46). Such efficient approximations of the gradient computation can be especially important as artificial networks become ever larger and are used to tackle ever more complex tasks under both time and energy-efficiency constraints.

Examination of the learning capability of MDGL under a broader range of tasks and conditions represents a valuable future avenue. For instance, our preliminary simulations on modulating the delay period in the match to sample task (SI Appendix, Fig. S13) suggests that such studies can help reveal the reasons underpinning the observed animal behavior (47), as well as limitations of MDGL. In addition, brain cells are extremely diverse (38, 48) with a matching diversity in the expression of peptidergic genes (30). Further studies can also investigate the interplay of task complexity and cell diversity. A starting point for that could be further dividing inhibitory cells into subtypes with and without threshold adaptation.

Our work suggests that multiple cell-type–specific, diffuse, and relatively slow modulatory signals should be considered as possible bases for credit-assignment computations. Though inspiration for the present work came primarily from new transcriptomic data on local NP signaling in neocortex (30, 42), it is quite possible that other cell-type–specific neuromodulators could likewise contribute to credit assignment. Many of these alternative agents act, as do NPs, via GPCRs (e.g., the monoamines, amino acids, acetylcholine, and endocannabinoids), but our multidigraph template might even apply to other neuronally secreted neuromodulators, such as the neurotrophins and cytokines, that act via different classes of receptors (49, 50). While experimental
tests of such hypotheses have not seemed feasible up until now, emerging methods for genetically addressed measurement of various neuromodulatory signals in specific cell types (30, 51) are now bringing the necessary critical tests within reach (e.g., ref. 52).

Methods

Visual Summary of Learning Rules. An overview of our network model and the mathematical basis of the learning rules used in this work is given in the beginning of Results. Here, we first present an additional, more detailed visual illustration of these learning rules in Fig. 5, beginning with the exact gradient update (Fig. 5A), as for BPTT, and leading from its dramatic truncation in the e-prop rule (Fig. 5B), to MDGL (Fig. 5C–E), which partially recovers gradient information lost in this truncation.

Spiking neuron model. We consider a discrete-time implementation of RSNNs. The model, as shown in Fig. 6A, observes the measurable states, i.e., spikes, as \( z_t \) at time \( t \), and the corresponding hidden states as \( s_t \). For leaky integrate-and-fire (LIF) cells, the state \( s_t \) corresponds to membrane potential, and the dynamics of those states are governed by

\[
S_{j,t+1} = \eta S_{j,t} + (1 - \eta) \left( \sum_{j} w_{pq} z_{p,t} + \sum_{m} w_{m,n} x_{m,t+1} \right) - z_{j,t} V_{th},
\]

where \( S_j \) denotes the membrane potential for neuron \( j \) at time \( t \), \( V_{th} \) denotes the spiking threshold potential, \( \eta = e^{-\alpha t}/\tau_m \) denotes the leak factor for simulation time step \( \Delta t \) and membrane time constant \( \tau_m \), \( w_{pq} \) denotes the weight of the synaptic connection from neuron \( j \) to \( l \), \( w_{m,n} \) denotes the strength of the connection between the input neuron \( m \) and neuron \( j \), \( \chi_t \) denotes the external input spike at time \( t \), and \( H \) denotes the Heaviside step function.

Following ref. 22, which implemented adaptive threshold LIF (ALIF) units (37) and observed that this neuron model improves computing capabilities of RSNNs relative to networks with LIF neurons only, we also include ALIF cells in our model. In addition to the membrane potential, ALIF cells have a second hidden variable, \( b_t \), governing the adaptive threshold. The spiking dynamics of both LIF and ALIF cells can be characterized by the following set of equations:

\[
S_{j,t+1} = \eta S_{j,t} + (1 - \eta) \left( \sum_{j} w_{pq} z_{p,t} + \sum_{m} w_{m,n} x_{m,t+1} \right) - z_{j,t} V_{th},
\]

\[
z_{j,t} = H(S_{j,t} - A_{j,t}),
\]

\[
A_{j,t} = V_{th} + \beta b_{j,t},
\]

\[
b_{j,t} = \beta b_{j,t-1} + (1 - \beta) z_{j,t-1}.
\]

where the voltage dynamics in Eq. 3 is the same as Eq. 2. A spike is generated when the voltage \( S_j \) exceeds the dynamic threshold \( A_{j,t} \). Parameter \( \beta \) controls how much adaptation affects the threshold, and \( s_p \equiv b_j \) denotes the variable component of the dynamic threshold. The decay factor \( \beta \) is given by \( e^{-\alpha t}/\tau_m \) for simulation time step \( \Delta t \) and adaptation time constant \( \tau_m \), which is typically chosen on the behavioral task time scale. For regular LIF neurons without adaptive threshold, one can simply set \( \beta = 0 \).

Network output and loss function. Dynamics of leaky, graded readout neurons is implemented as

\[
y_{k,t} = \eta y_{k,t-1} + (1 - \eta) \sum_{j} w_{kj}^\text{OUT} z_{j,t} + b_{k,t}^\text{OUT},
\]

where \( w_{kj}^\text{OUT} \) denotes the strength of the connection from neuron \( j \) to output neuron \( k \), \( b_{k,t}^\text{OUT} \) denotes the bias of the k-th output neuron, \( k \in \{0, 1\} \) defines the leak, and \( \eta = e^{-\alpha t}/\tau_m \) for output membrane time constant \( \tau_m \).

We quantify how well the network output matches the desired target using error function \( E \):

\[
E = \left[ \frac{1}{2} \sum_{k} (y_{k,t} - y_{k,t}^\text{target})^2, \right. \text{for regression tasks} \]

\[
\left. - \sum_{k} \pi_k \log \pi_k, \text{for classification tasks} \right].
\]

where \( y_{k,t}^\text{target} \) is the time-dependent target, \( \pi_k \) is the one-hot encoded target, and \( \pi_k \equiv \text{softmax}(y_{k,t}) = \exp(y_{k,t})/\sum \exp(y_{k,t}) \), is the predicted category probability. We provide all simulation and training parameters in SI Appendix, Note 3.

While the tasks involving time-delayed rewards studied in this manuscript can be labeled as regression and classification tasks due to the nature of the objective function, we note that the theoretical development is general and allows for

Fig. 5. Cartoon summary of learning rules explored in this work. (A) The exact gradient: Updating weight \( w_{pq} \) the synaptic connection strength from presynaptic neuron \( q \) to postsynaptic neuron \( p \), involves nonlocal information inaccessible to neural circuits, i.e., the knowledge of activity (e.g., voltage \( s \)) for all distant neurons \( j \) and \( l \) in the network. This is because \( w_{pq} \) affects the activities of many other cells through indirect connections, which will then affect the network output at subsequent time steps (Eq. 17 in Methods). (B) E-prop, a state-of-the-art biologically plausible learning rule, restricts the weight update to depend only on presynaptic and postsynaptic activity and TD learning signal, as in a three-factor learning rule (Fig. 2A). (C) We allow the weight update to capture dependencies within one connection step, which are omitted in e-prop. The activity of neuron \( j \) could be delivered to \( p \) through local modulatory signaling. (D) For the signaling in C to be cell-type-specific, as consistent with experimental observation in ref. 42 and biologically plausible mechanisms, we approximate the cell-specific gain with cell-type-specific gain (Eq. 23), which leads to our MDGL. Effect of this cell-type approximation is explored in SI Appendix, Fig. 59. (E) NL-MDGL, where modulatory signal diffuses to all cells in the network without attenuation (Fig. 4).
Fig. 6. Computational graph and gradient propagation. (A) Schematic illustration of the recurrent neural network used in this study. (B) The mathematical dependencies of input \( x \), state \( s \), neuron spikes \( z \), and loss function \( E \) unwrapped across time. (C) The dependencies of state \( s \) and neuron spikes \( z \) unwrapped across time and cells. (D) The computational flow of \( ds/dw_{pq} \) is illustrated for exact gradients computed using exact calculation (Eq. 17) (i), e-prop (ii), and our truncation in Eq. 18, where dependency within one connection step has been captured (iii). Black arrows denote the computational flow of network states, output, and the loss; for instance, the forward arrows from \( z_t \) and \( s_t \) going to \( s_{t+1} \) are due to the neuronal dynamics equation in Eq. 2. Green arrows denote the computational flow of \( ds/dw_{pq} \) for various learning rules.

Gradient descent learning in RSNNs. We study iterative adjustment of all synaptic weights (input weights \( w^i \), recurrent weights \( w \), and output weights \( w^\text{OUT} \)) using gradient descent on loss \( E \):

\[
\begin{align*}
\text{w}_{pq,\text{new}} &= \text{w}_{pq,\text{old}} - \lambda \Delta \text{w}_{pq}, \\
\Delta \text{w}_{pq} &= \frac{\partial E}{\partial \text{w}_{pq,\text{old}}},
\end{align*}
\]

where \( \lambda \) denotes the learning rate, and the gradient of the error with respect to the synaptic weights must be calculated. This error gradient can be calculated with classical machine-learning algorithms, BPTT and RTRL, by unwrapping the RSNN dynamics over time (Fig. 6B). While these two algorithms yield equivalent results, their bookkeeping for chain rule differs. Gradient calculations in BPTT depend on future activity, which poses an obstacle for online learning and biological plausibility. Our learning-rule derivation follows the RTRL factorization because it is causal. Therefore, we focus our analysis on RTRL and factor the error gradient across time and space as

\[
\begin{align*}
\frac{\partial E}{\partial \text{w}_{pq}} &= \sum_{j,t'} \frac{\partial E}{\partial z_{j,t'}} \frac{\partial z_{j,t'}}{\partial \text{w}_{pq}}, \\
\frac{\partial z_{j,t'}}{\partial \text{w}_{pq}} &= \frac{\partial E}{\partial \text{w}_{pq}} \frac{\partial z_{j,t'}}{\partial \text{w}_{pq}},
\end{align*}
\]

following the derivative notation explained above. The factor \( \frac{\partial E}{\partial \text{w}_{pq}} \) in Eq. 13 is related to the TD learning signal \( \Delta \text{z}_{j,t} := \sum_{t'} \text{w}^{\text{OUT}}_{kj,t'} (y_{j,t'} - y'_{j,t'}) \) (22).

**Notation for derivatives.** There are two types of computational dependencies in RSNNs: direct and indirect dependencies. For example, variable \( w_{pq} \) can impact state \( s_{qt} \) directly through Eq. 2, as well as indirectly via its influence through other cells in the network. We distinguish direct dependencies vs. all dependencies (including indirect ones) using partial derivatives (d) vs. total derivatives (d).
that uses $L_{j\ell}$. We thus take TD learning signals to be cell-specific rather than global, which is justified in part by recent reports that dopamine signals (28, 29) and error-related neural firing (33) can be specific to a population of neurons (50). Moreover, in approximating the sum in $L_{j\ell}$ as in our main derivation below, following the argument on cognate receptors, or using the random feedback alignment theory (14) (on only the outgoing connections between spiking neurons and output units) suggest further biologically plausible implementations.

We now discuss the second factor in Eq. 13, i.e., $\frac{d z_{j\ell}}{d w_{pq}}$. This is expanded into two factors in Eq. 14. The first factor, $h_{j\ell} := \frac{\partial z_{j\ell}}{\partial w_{pq}}$, is problematic to compute for spiking neurons due to the discontinuous step function $H$ in Eq. 3, whose derivative is not defined at zero and is zero everywhere else. We overcome this issue by approximating the decay of the derivative using a piece-wise linear function (10, 22, 46, 54). Here, the pederivative $h_{j\ell}$ is defined as follows:

$$h_{j\ell} = \frac{d z_{j\ell}}{d w_{pq}} \approx \gamma \max \left( 0, 1 - \frac{s_{j\ell} - a_{j\ell}}{v_{n}} \right).$$  

The dampening factor $\gamma$ (typically set to 0.3) dampens the increase of backpropagated errors in order to improve the stability of training very deep (unrolled) RNNs (22). Throughout this study, neuronal firing displays refractoriness, where $h_{j\ell}$ and $z_{j\ell}$ are fixed at zero after each spike of neuron $j$ (SI Appendix, Note 31). Key problems that RTRL poses to biological plausibility and computational cost reside in the factor $\frac{d z_{j\ell}}{d w_{pq}}$, that arises during the factorization of the gradient (Eqs. 13 and 14). The factor $\frac{d z_{j\ell}}{d w_{pq}}$ keeps track of all direct and indirect dependencies of neuron state $j$ on weight $w_{pq}$. In other words, this factor accounts for both the spatial and temporal dependencies in RNNs: State dependencies across time $t$ as explained above, result from unrolling the temporal dependencies illustrated in Fig. 6B; state dependencies across space, however, are due to the indirect dependencies (of all $z_{j\ell}$ on $w$ and all $z_{j\ell}$) that arise from recurrent connections (Fig. 6C). These recurrent dependencies are all accounted for in the $\frac{d z_{j\ell}}{d w_{pq}}$ factor, which can be obtained recursively as follows:

$$\frac{d z_{j\ell}}{d w_{pq}} = \frac{\partial z_{j\ell}}{\partial w_{pq}} + \sum_{t} \frac{\partial z_{j\ell}}{\partial w_{pq}} \frac{d s_{j\ell-1}^t}{d w_{pq}} + \sum_{t} \frac{\partial z_{j\ell}}{\partial w_{pq}} \frac{d s_{j\ell-1}^t}{d w_{pq}} + \sum_{t} \frac{\partial z_{j\ell}}{\partial w_{pq}} \frac{d s_{j\ell-1}^t}{d w_{pq}}.$$  

Thus, the factor $\frac{d z_{j\ell}}{d w_{pq}}$ is a memory trace of all intercellular dependencies (Fig. 6D, j) and requires $O(N^3)$ memory and $O(N^3)$ computations. This makes RTRL expensive to implement for large networks. Moreover, this last factor poses a serious problem for biological plausibility: It involves nonlocal terms, so that knowledge of all other weights in the network is required in order to update the weight $w_{pq}$.

To address this, Murray (21) and Bellec et al. (22) (e-prop) dropped the nonlocal terms so that the updates to weight $w_{pq}$ would only depend on presynaptic and postsynaptic activity (Figs. 5B and 6 D, ii) and applied this truncation to train rate-based and spiking neural networks, respectively. While both works succeed in improving over previous biologically plausible learning rules, a significant performance gap with respect to the full BPTT/RTRL algorithms remains.

**Derivation of multitigraph learning in RNNs.** We continue from the previous section in giving a detailed derivation of our learning rule. To reveal a potential role for cell-type–specific modulatory signals in synaptic plasticity as well as improve upon the aforementioned biologically plausible gradient descent approximations, we begin by partially restoring nonlocal dependencies between cells—those within one connection step. This is the “truncated” RTRL framework (Figs. 5D and 6 D, iii), and the memory trace term $\frac{d z_{j\ell}}{d w_{pq}}$ becomes

$$\frac{d z_{j\ell}}{d w_{pq}} \approx \left( \frac{\partial z_{j\ell}}{\partial w_{pq}} + \frac{\partial z_{j\ell}}{\partial w_{pq}} \frac{d s_{j\ell-1}^t}{d w_{pq}} \right).$$  

Thus, when $j = p$, our truncation implements $\frac{d z_{j\ell}}{d w_{pq}} \approx \frac{\partial z_{j\ell}}{\partial w_{pq}} + \frac{\partial z_{j\ell}}{\partial w_{pq}} \frac{d s_{j\ell-1}^t}{d w_{pq}}$, which coincides with e-prop. Eq. 18 adds the case when $p \neq j$, for which $\frac{d s_{j\ell}}{d w_{pq}}$ was simply set to zero in e-prop. We note that the truncation in Eq. 18 resembles the n-step RTRL approximation recently proposed in ref. 8, known as SnAP-n, which stores $\frac{d s_{j\ell}}{d w_{pq}}$ only for $j$ such that parameter $\frac{\partial z_{j\ell}}{\partial w_{pq}}$ influences the activity of unit $j$ within $n$ time steps. The computations of SnAP-n converge to those of RTRL as $n$ increases, resonating with our improved performance when more terms of the exact gradient are included. Our truncation in Eq. 18 is similar to SnAP-n with $n = 2$ with two differences: 1) We apply it to spiking neural networks, and 2) we drop the previous time step's Jacobian term $\frac{\partial z_{j\ell}}{\partial w_{pq}}$, which would necessitate the maintenance of a rank-three (“3-d”) tensor with costly storage demands ($O(N^3)$) and for which no known biological mechanisms exist. Thus, the truncation in Eq. 18 requires the maintenance of only a rank-two (“2-d”) tensor specific to synapse $p/q$, which can be realized via an ET, as we explain next.

By substituting Eq. 18 into Eqs. 13 and 14, we approximate the overall gradient as

$$\frac{d E}{d w_{pq}} = \sum_{t} \frac{d E}{d z_{j\ell}} \frac{d z_{j\ell}}{d w_{pq}} + \sum_{t} \frac{d E}{d z_{j\ell}} \frac{d z_{j\ell}}{d w_{pq}} + \sum_{j \neq p} \frac{d E}{d z_{j\ell}} \frac{d z_{j\ell}}{d w_{pq}}.$$  

where $L_{j\ell} := \frac{d E}{d w_{pq}}$ is the TD learning signal to cell $p$, $a_{j\ell}$ (Eq. 24) denotes the activity-dependent modulatory signal emitted by neuron $j$ at time $t$, and $e_{p,q}$ (Eq. 25) is the ET maintained by postsynaptic cell $p$ to keep a memory of the preceding activity of presynaptic cell $q$ and postsynaptic cell $p$. In Eq. 19, the first term $L_{j\ell}a_{j\ell}$ alone gives exactly the e-prop synaptic update rule. The second term, which we define as $\Gamma_{p,q}$, is a synthetically nonlocal term due to contributions from local modulatory signals. As seen in Eq. 19, our truncation requires maintaining a $(p, q)$-dependent double tensor (for $e_{p,q}$) instead of a single triple, thereby reducing the memory cost of RTRL from $O(N^4)$ to $O(N^2)$.

Importantly, we observe that, for the update to synapse $w_{pq}$ in Eq. 19, the terms that depend on cells $j$ only appear under a sum. Therefore, the mechanism updating the synapse $p/q$ does not need to know the individual terms indexed by $j$. Rather, only their sum suffices.

While it is tempting to consider the first factors in $\Gamma_{p,q}$, $a_{j\ell}$, as the modulatory signal emitted by neuron $j$, the involvement of the synapse from neuron $p$ via $w_{pq}$ and a lack of known mechanisms in calculating this neuron-specific composite signal suggest that this is unlikely to be a biological solution. Instead, inspired by the cell-type–specific (rather than neuron-specific) affinities for peptidergic neuromodulation (30, 42), we propose to approximate the signaling gain $w_{pq}$ in Eq. 19 by the average value $w_{pq}$ across its presynaptic and postsynaptic cell types. More specifically, when postsynaptic cell $j$ belongs to type $\alpha$ and presynaptic cell $p$ belongs to type $\beta$, we approximate neuron-specific weight $w_{p,q}$ with cell-type–specific gain $w_{\alpha,\beta} = w_{\alpha,\beta} / \sum_{p} \sum_{q} w_{\alpha,\beta}$, where $w_{\alpha,\beta}$ represents the affinity of the GPCRs expressed by cells of type $\beta$ to the modulators secreted by cells of type $\alpha$ (Cell-type-specific receptor affinities).

Thus, the gradient estimate at time $t$ due to our learning rule involves compounding ET with both TD and local modulatory signals, thereby recovering the general form introduced in Eq. 1:

$$\frac{d E}{d w_{pq}} \approx \sum_{t} L_{j\ell}a_{j\ell} + \Gamma_{p,q},$$  

where neuron $p$ is of type $\beta$, $C$ denotes the set of neuronal cell types, $\rightarrow j$ denotes that there is a synaptic connection from neuron $p$ to $j$, and $\Gamma_{p,q}$ approximates the second term in Eq. 19 with cell-type–specific weight averages.

In summary, cell $p$ receives local modulatory input $\text{Mod}.\text{input}_p$, that gets combined with the ET (as per Eq. 20) in addition to synaptic input $\text{Syn}.\text{input}_p$, (as per Eq. 2):

$$\text{Mod}.\text{input}_p := \sum_{\alpha \in C} w_{\alpha,\beta} \sum_{j \in \rightarrow \alpha} a_{j\ell},$$  

$$\text{Syn}.\text{input}_p := \sum_{p} w_{pq} z_{p\ell} + \sum_{\alpha} w_{\alpha,\beta} x_{\alpha,\beta,\ell}.$$  

Liu et al.
Cell-type-specific neuromodulation guides synaptic credit assignment in a spiking neural network

PNAS | 11 of 11
https://doi.org/10.1073/pnas.2111821118
It may be instructive to note the dichotomy in the functions of these two different inputs: The cell uses Mod inputs to regulate its synaptic plasticity, but not to change its internal state, and it uses Syn inputs to change its internal state, but not to regulate synaptic plasticity.

Hence, our update rule suggests an additive term to compute the plasticity update at synapse $p/q$ at time $t$, $\Gamma_{p,q}$, which calculates multiplicative contributions of the modulatory signal $\alpha_{j,t}$ secreted by neuron $j$, the affinity of receptors of cell type $\beta$ to ligands of type $\alpha$, $w_{p,q,\beta}$, and the ET at the synapse $p/q$, $\eta_{p,q}$. The following two sections explain how two main components of $\Gamma$: cell-type–specific signals and ET, can be implemented.

**Cell-type–specific receptor affinities.** We explain cell-type–specific signaling implementation, notably, how type–specific receptor affinity $w_{p,q,\beta}$ is defined. As introduced in our learning rule derivation, $w_{p,q,\beta}$ is an approximation of gain $w_{p}$ (Eqs. 19 and 20), and we proposed to define $w_{p,q,\beta}$ as the weight average across its presynaptic and postsynaptic cell types:

$$w_{p,q,\beta} \approx \left\{ \begin{array}{ll}
\langle w_{\alpha_{t}}, \rho \rightarrow j \rangle & \text{otherwise},
\end{array} \right.$$  

where $\rho \rightarrow j$ denotes that there is a synaptic connection from neuron $\rho$ to $j$, motivated by the local diffusion assumption discussed in ref. 35, in which this type of signaling is registered only by local synaptic partners and therefore preserves the connectivity structure of $w_{p}$. One obvious variant of this receptor-affinity definition is one with a different spatial extent, for which we examine the opposite extreme in Fig. 4 D–F, where modulatory signals diffuse to all cells in the network. More specifically, the signaling gain $w_{p,q,\beta}$ is replaced by $\sum_{\rho} w_{p,q,\beta} \alpha_{j,t} \beta_{\alpha} \rho$, even for $w_{p} = 0$ so that modulatory signals diffuse to all cells with the same strength in the network.

For a proof of concept, we implemented MDGL with modulatory types mapped to the two main cell classes; i.e., cell-type–specific signaling gain $w_{p,q,\beta} \alpha_{j,t} \beta_{\alpha} \rho$, with $\alpha, \beta \in \{E, F\}$. We demonstrate the effectiveness of this cell-type discretization by comparing its learning performance to the case without cell-type discretization ($w_{p,q,\beta} = w_{p}$, i.e., each cell is its own type) and observed little difference in performance (SI Appendix, Fig. S7) exploring the sensitivity of ETs as derivatives can be found in ref. 22. Here, we briefly explain its implementation by expanding the factors in Eqs. 25 and 26 for both LIF and ALIF cells.

**LIF cells**, there is no adaptive threshold, so the hidden state consists only of the membrane potential. Thus, we have factors $\partial \eta_{p,q}/\partial w_{p}$, with pseudoderivative $h_{q}$ defined in Eq. 15, $\partial \eta_{p,q}/\partial w_{p,q,\beta}$ following Eq. 2.

For ALIF cells, there are two hidden variables, so the eligibility vector is now a two-dimensional vector $\frac{\partial \eta_{p,q}}{\partial h_{q}} = \partial \eta_{p,q}/\partial h_{q}$, $\partial \eta_{p,q}/\partial h_{q}$ pertaining to membrane potential $V_{p}$ and adaptive threshold state $b_{p}$. Following Eq. 3, one can obtain factors

$$\frac{\partial \eta_{p,q}}{\partial h_{q}} = \left[ \begin{array}{c}
\partial \eta_{p,q}/\partial h_{q}^1 \\
\partial \eta_{p,q}/\partial h_{q}^2
\end{array} \right] = \left[ \begin{array}{c}
\partial \eta_{p,q}/\partial h_{q}^1 \\
\partial \eta_{p,q}/\partial h_{q}^2
\end{array} \right] = \left[ \begin{array}{c}
\partial \eta_{p,q}/\partial h_{q}^1 \\
\partial \eta_{p,q}/\partial h_{q}^2
\end{array} \right]$$

Thus, the ET $\eta_{p,q}$ is scalar valued, regardless of the dimension of the eligibility vector.

**Data Availability.** Code for data generation and analysis is available in GitHub at https://github.com/Helena-Yuhan-Liu/MDGL-main.

**ACKNOWLEDGMENTS.** We thank Forrest Collins, Guillaume Lajoie, James Murray, Scott Owen, and Bosiljka Tasic for helpful feedback on the manuscript. We also thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support. Y.H.L. is supported by the National Science Foundation under grant no. 1811017.

12. B. A. Richards, T. P. Lillicrap, Dendritic solutions to the credit assignment problem. Curr. Opin. Neurobiol. 54, 28–36 (2019).

13. J. E. Rubin, C. Vich, M. Clapp, K. Noneman, T. Verstynen, The credit assignment problem in cortico-basal ganglia-thalamic networks: A review, and a possible solution. Eur. J. Neurosci. 53, 2234–2253 (2021).

14. T. P. Lillicrap, D. Cowden, D. B. Tweed, C. J. Akerman, Random synaptic feedback weights support error backpropagation for deep learning. Nat. Commun. 7, 13276 (2016).

15. A. Payeur, J. Guerguev, F. Zenke, B. A. Richards, R. Naud, Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. Nat. Neurosci. 24, 1010–1019 (2021).

16. I. Pozzi, S. Bohné, P. Roefselma, A biologically plausible learning rule for deep learning in the brain. arXiv (Preprint) 2018. https://arxiv.org/abs/1811.01768 (Accessed 2 August 2021).

17. J. Sacramento, R. P. Costa, Y. Bengio, W. Senn, Dendritic cortical microcircuits approximate the backpropagation algorithm. arXiv (Preprint) 2018. https://arxiv.org/abs/1810.11393 (Accessed 2 August 2021).

18. A. Laborieux et al., Scaling equilibrium propagation to deep ConvNets by drastically reducing its gradient estimator bias. Front. Neurosci. 15, 636374 (2021).

19. Y. Amit, Deep learning with asymmetric connections and Hebbian updates. Front. Comput. Neurosci. 13, 18 (2019).

20. B. Millidge, A. Tschantz, A. K. Seth, C. L. Buckley, Activation relaxation: A local dynamical approximation to backpropagation in the brain. arXiv (Preprint) 2019. https://arxiv.org/abs/2009.03599 (Accessed 2 August 2021).

21. J. M. Murray, Local online learning in recurrent networks with random feedback. eLife 8, e43299 (2019).

22. G. Bellec et al., A solution to the learning dilemma for recurrent networks of spiking neurons. Nat. Commun. 11, 3625 (2020).
Cell-type–specific neuromodulation guides synaptic credit assignment in a spiking neural network

Z. Brzosko, S. B. Mierau, O. Paulsen, Neuromodulation of spike-timing-dependent plasticity: Past, present, and future. *Neuron* 103, 563–581 (2019).

A. N. van den Pol, Neuropeptide transmission in brain circuits. *Cell* 184, 2733–2749.e16 (2021).

A. Suvrathan, Beyond STDP–towards diverse and functionally relevant plasticity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20190761 (2021).

W. Gerstner, M. Lehmann, V. Liakoni, D. Corneil, J. Brea, Eligibility traces and plasticity: Past, present, and future. *Neuron* 103, 563–581 (2019).

B. Engelhard et al., Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* 570, 509–513 (2019).

A. S. Morcos, C. D. Harvey, History-dependent variability in population dynamics of dendritic spines. *Science* 345, 1616–1620 (2014).

S. Yagishita et al., Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* 184, 2733–2749.e16 (2021).

B. A. Hamid, M. J. Frank, C. I. Moore, Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* 184, 2733–2749.e16 (2021).

T. Meyer, X. L. Qi, T. R. Stanford, C. Constantinides, Stimulus selectivity in dorsal and ventral prefrontal cortex after training in working memory tasks. *J. Neurosci.* 31, 6266–6276 (2011).

S. J. Smith, M. Hawrylycz, J. Rossier, U. Sümbl, New light on cortical neuropeptides and synaptic network plasticity. *Curr. Opin. Neurobiol.* 63, 176–188 (2020).

G. Jékely, The chemical brain hypothesis for the origin of nervous systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20190761 (2021).

B. A. Richards et al., A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761–1770 (2019).

D. Linsley, A. K. Ashok, L. N. Govindarajan, R. Liu, T. Serre, Stable and expressive recurrent vision models. arXiv [Preprint] (2020). https://arxiv.org/abs/2005.11362 (Accessed 25 September 2020).

D. Huh, T. J. Sejnowski, "Gradient descent for spiking neural networks" in 32nd Conference on Neural Information Processing Systems, S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, Eds. (Curran Associates Inc., Red Hook, NY, 2018), pp. 1433–1443.

R. Elliott, R. J. Dolan, Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. *J. Neurosci.* 19, 5066–5073 (1999).

N. W. Gouwens et al., Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat. Neurosci.* 22, 1182–1195 (2019).

H. Lu, H. Park, M. M. Poo, Spike-time-dependent BDNF secretion and synaptic plasticity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130132 (2013).

E. J. Donzis, N. C. Tronson, Modulation of learning and memory by cytokines: Signaling mechanisms and long term consequences. *Neurobiol. Learn. Mem.* 115, 68–77 (2014).

S. K. Esser et al., Convolutional networks for fast, energy-efficient neuromorphic computing. *Proc. Natl. Acad. Sci. U.S.A.* 113, 11441–11446 (2016).

M. Staykova, S. S. Jacobsen, C. A. Curie, L. Wang, D. C. Moore, G. M. Smith, A. M. Zito, J. P. Hetke, D. H. Sadik-T同时还存在一种特殊的神经元类型，使得能量和效能的平衡得以实现。