

Desiderata for Normative Models of Synaptic Plasticity

Colin Bredenberg

colin.bredenberg@mila.quebec

*Center for Neural Science, New York University, New York, NY 10003, U.S.A., and
Mila-Quebec AI Institute, Montréal, QC H2S 3H1, Canada*

Cristina Savin

cs5360@nyu.edu

*Center for Neural Science, New York University, New York, NY 10003, U.S.A., and
Center for Data Science, New York University, New York, NY 10011, U.S.A.*

Normative models of synaptic plasticity use computational rationales to arrive at predictions of behavioral and network-level adaptive phenomena. In recent years, there has been an explosion of theoretical work in this realm, but experimental confirmation remains limited. In this review, we organize work on normative plasticity models in terms of a set of desiderata that, when satisfied, are designed to ensure that a given model demonstrates a clear link between plasticity and adaptive behavior, is consistent with known biological evidence about neural plasticity and yields specific testable predictions. As a prototype, we include a detailed analysis of the REINFORCE algorithm. We also discuss how new models have begun to improve on the identified criteria and suggest avenues for further development. Overall, we provide a conceptual guide to help develop neural learning theories that are precise, powerful, and experimentally testable.

1 Introduction

Our identities change with time, gradually reshaping our experiences. We remember, we associate, we learn. However, we are only beginning to understand how changes in our minds arise from underlying changes in our brains. Of the many features of neural architecture that are altered over time, from the biophysical properties of individual neurons to the creating or pruning of synapses between neurons, changes in the strength of existing synapses remain the most prominent candidate for the neural substrate of longitudinal perceptual and behavioral change (Magee & Grienberger, 2020). Synaptic connections are easily modified, and these

Colin Bredenberg is the corresponding author.

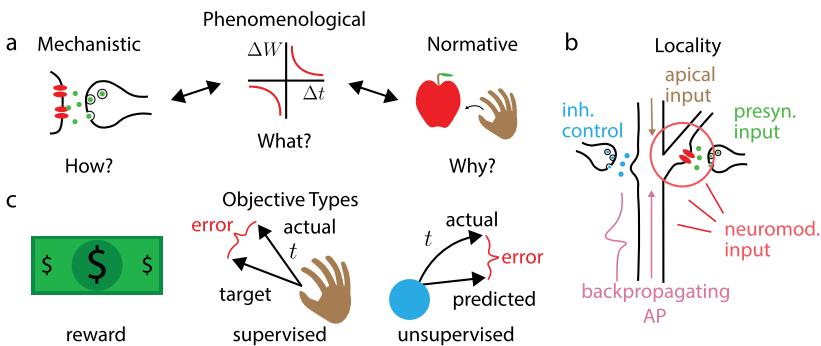


Figure 1: Spectrum of synaptic plasticity models. (a) By level of abstraction: Mechanistic models show how detailed biophysical interactions produce observed plasticity, phenomenological models concisely describe what changes in experimental variables (e.g., before and after relative spike timing Δt) affect plasticity (ΔW), and normative models explain why the observed plasticity implements capabilities that are useful to the organism. (b) Schematic illustrating the range of local variables that may be available for synaptic plasticity. (c) Classes of objective functions: Reward-based learning involves general feedback about how well the organism or network performed; supervised objectives specify explicit desired outcomes; unsupervised objectives do not require any form of explicit feedback.

modifications can persist for extended periods of time (Bliss & Collingridge, 1993). Further, synaptic modification has been associated with many of the brain's critical adaptive functions, including memory (Martin et al., 2000), experience-based sensory development (Levelt & Hübener, 2012), operant conditioning (Fritz et al., 2003; Ohl & Scheich, 2005), and compensation for stroke (Murphy & Corbett, 2009) or neurodegeneration (Zigmond et al., 1990). However, beyond these associations, it is often hard to establish a precise link between plasticity and a certain adaptive behavior. In this review, we distinguish "normative" modeling approaches from alternatives, demonstrate why they show promise for establishing precise links between mechanism and behavioral outcomes, and outline a set of desiderata that articulate how recent progress on normative plasticity models strengthens the link between plasticity and system-wide adaptive phenomena.

Plasticity models come in several flavors: phenomenological, mechanistic, and normative models (see Figure 1a) (Levenstein et al., 2020)—with the demarcation lines between them not always completely precise. Broadly, phenomenological models focus on concisely describing what happens in plasticity experiments with mathematical modeling; mechanistic modeling adds to this project by explaining how plasticity dynamics emerge from causal interactions between biophysical quantities. While

phenomenological and mechanistic models articulate how synaptic plasticity works, they do not explain why it exists in the brain, that is, what its importance is for neural circuits, behavior, or perception. Answering this question with any precision requires an appeal to normative modeling.

Normative models aim to answer this *why* question by connecting plasticity to observed network-level or behavioral-level phenomena, including memory formation (Hopfield, 1982; Lengyel et al., 2005; Savin et al., 2014) and consolidation (Fusi et al., 2005; Clopath et al., 2008; Benna & Fusi, 2016), reinforcement learning (Frémaux & Gerstner, 2016), and representation learning (Oja, 1982; Hinton et al., 1995; Rao & Ballard, 1999; Toyozumi et al., 2005; Savin et al., 2010). Guided by the intuition that plasticity processes have developed on an evolutionary timescale to near-optimally perform adaptive functions, normative plasticity theories are typically top-down in that they begin with a set of prescriptions about how synapses *should* modify in order to optimally perform a given learning-based function. Subsequently, with varying degrees of success, these theories attempt to show that real biology matches or approximates this optimal solution. Here, we review classical normative plasticity approaches and discuss efforts to improve them.¹ To provide concrete examples of these principles in action, in appendix C we describe the REINFORCE algorithm (Williams, 1992), explain how it can function as a normative plasticity model, and note its successes and failures to match our desiderata.

2 Desiderata for Normative Models

One of the biggest challenges for normative models of synaptic plasticity is their connection to biology: their predictions often tie biophysical phenomenology with function in ways that are hard to access experimentally. Therefore, it is a major challenge to identify how to improve normative models with relatively limited access to experimental data confirming or rejecting their predictions. In what follows, we articulate a set of desiderata that can serve as both an organizing tool for understanding the contributions of recent normative plasticity modeling efforts and as a set of intermediate objectives for the development of new models in the absence of explicit experimental rejection or confirmation of older work. Normative plasticity models do not need to satisfy all desiderata to be useful. For example, several seminal normative plasticity models fail to accommodate known facts about biology (e.g., Hopfield networks (Hopfield, 1982) and Boltzmann machines (Ackley et al., 1985)). We argue that any normative plasticity model can be improved by making it conform more closely to our desiderata. Each principle is desirable for some combination of the

¹In the interest of conciseness, we discuss only long-term plasticity, not including short-term plasticity.

following reasons: first, it may help ensure that the plasticity model actually qualifies as normative; second, it may require a model to accommodate known facts about biology; and third, it may ensure that models can be compared to existing experimental literature and generate genuinely testable experimental predictions. Most of these desiderata are relatively intuitive and simple. However, it has proven incredibly difficult for existing models of any adaptive cognitive phenomenon—from sensory representation learning, to associative memory formation, to reinforcement learning—to satisfy all of them in tandem.

2.1 Improving Performance on a Specified Objective. Many popular normative frameworks view neural plasticity as an approximate optimization process (Lillicrap et al., 2020; Richards et al., 2019), wherein synaptic modifications progressively reduce a scalar loss function. Within this perspective, the function of synaptic plasticity is to improve performance on this objective.² Thus, the modeling process can be divided into two steps: articulating an appropriate objective and subsequently demonstrating that a synaptic plasticity mechanism improves performance.

Normative theories of synaptic plasticity developed to date usually involve some combination of supervised, unsupervised, or reinforcement learning objectives (see Figure 1c). The choice of objective function for a neural system influences the resultant form and scope of applicability of the model. For instance, supervised learning implies the existence of either an internal (e.g., motor error signals (Gao et al., 2012; Bouvier et al., 2018) or saccade information indicating that a visual scene has changed (Illing et al., 2021)) or external teacher (e.g., zebra finch song learning (Fiete et al., 2007)). Unsupervised teaching signals can be provided by prediction, as in generative modeling frameworks (Fiser et al., 2010). This account of sensory coding is popular for both its ability to accommodate normative plasticity theories (Rao & Ballard, 1999; Dayan et al., 1995; Kappel et al., 2014; Isomura & Toyoizumi, 2016; Bredenberg et al., 2021) and its philosophical vision of sensory processing as a form of advanced model building, beyond simple sensory transformations. Bayesian inference frameworks can also be useful for systematically quantifying uncertainty about optimal synaptic parameter estimates and for adjusting learning rates accordingly (Kappel et al., 2015; Aitchison et al., 2021; Jegminat et al., 2022). However, alternative perspectives on sensory processing exist, including those based on maximizing the information about a sensory stimulus contained in a neural population (Attnave, 1954; Atick & Redlich, 1990) subject to metabolic efficiency constraints (Tishby et al., 2000; Simoncelli & Olshausen, 2001), and those based

² It should be noted that this is the simplest way to characterize improved performance, but not all formulations of learning easily fit into a simple optimization framework for example, associative learning in Hopfield networks (Hopfield, 1982) or multi-agent reinforcement learning (Zhang et al., 2021).

on contrastive methods (Oord et al., 2018; Illing et al., 2021), where a self-supervising internal teacher encourages the neural representation of some stimuli to grow closer together, while encouraging others to grow more discriminable.

Evaluating which objective function (or functions) best explains the properties of a neural system is hard: while some forms of objective function may have discriminable effects on plasticity (e.g. supervised versus unsupervised learning; Nayebi et al., 2020), others are even provably impossible to distinguish (see appendix A). This motivates the idea that for a given data set, it is plausible that one objective ($\tilde{\mathcal{L}}$) can masquerade as another (\mathcal{L}). In some cases, complex objective functions can masquerade as simple objectives, which may only be epiphenomenal. Take balancing excitatory and inhibitory inputs as an example objective for a neuron: this could be a goal on its own (Vogels et al., 2011) or a consequence of predictive coding (Brendel et al., 2020). In other cases, philosophically distinct frameworks, such as generative modeling, information maximization, or denoising may simply produce similar synaptic plasticity modifications because the frameworks often overlap heavily (Vincent et al., 2010) and may not be distinguishable on simple data sets without targeted experimental attempts to disambiguate between them.

Having addressed many difficulties associated with choosing a good objective function, we now move to difficulties involved in demonstrating that a particular synaptic plasticity rule decreases a chosen objective.³ How could such a property be proven? For a particular plasticity rule to reduce an objective $\mathcal{L}(\mathbf{W})$ that depends on synaptic weights \mathbf{W} , we need to show that the following principle holds:

$$\begin{aligned} \mathcal{L}(\mathbf{W} + \Delta\mathbf{W}) &< \mathcal{L}(\mathbf{W}) \\ \Rightarrow \mathcal{L}(\mathbf{W} + \Delta\mathbf{W}) - \mathcal{L}(\mathbf{W}) &< 0, \end{aligned} \tag{2.1}$$

for some update $\Delta\mathbf{W}$ determined by the plasticity rule. If we accept the additional supposition that $\Delta\mathbf{W}$ is very small, we can employ the first-order Taylor approximation (treating \mathbf{W} as a flattened vector: $\mathcal{L}(\mathbf{W} + \Delta\mathbf{W}) \approx \mathcal{L}(\mathbf{W}) + \frac{d\mathcal{L}}{d\mathbf{W}}(\mathbf{W})^T \Delta\mathbf{W}$). Substituting this approximation into our reduction criterion, we have after cancellation

$$\frac{d\mathcal{L}}{d\mathbf{W}}(\mathbf{W})^T \Delta\mathbf{W} < 0. \tag{2.2}$$

This shows that for small weight updates (slow learning rates), the inner product between a synaptic learning rule $\Delta\mathbf{W}$ and the gradient of the

³Some objectives (like reward functions) are best thought of as being maximized rather than minimized. Without loss of generality, in such cases we can minimize the negative reward function.

selected loss function $\mathcal{L}(\mathbf{W})$ with respect to the weight change must be negative. The simplest way to ensure that this is true is for $\Delta\mathbf{W}$ to equal a small scalar λ times the negative gradient of the loss ($-\lambda \frac{d\mathcal{L}}{d\mathbf{W}}(\mathbf{W})^T \frac{d\mathcal{L}}{d\mathbf{W}}(\mathbf{W}) = -\lambda \|\frac{d\mathcal{L}}{d\mathbf{W}}(\mathbf{W})\|_2^2 < 0$).⁴ If this were true, plasticity would be guaranteed to improve performance on the objective \mathcal{L} .

Some studies show empirically that this inner product is negative (Lillicrap et al., 2016; Marschall et al., 2020). A pure empirical demonstration that a learning algorithm aligns with the loss gradient on a particular task, network architecture, and data set does not necessarily generalize to the full range of relevant tasks. Moreover, trained networks are sensitive to hyperparameter choices, where small changes in simulated network parameters can effect large qualitative differences in network behavior (Xiao et al., 2021). Further, a battery of *in silico* simulations under a variety of different parameter settings and circumstances rapidly begins to suffer the curse of dimensionality, becoming almost as extensive as the collection of *in vivo* or *in vitro* experiments that it is attempting to explain.

For this reason, many studies construct mathematical arguments as to why equation 2.2 should hold for a given local synaptic plasticity rule by demonstrating that it either is a stochastic approximation to the true gradient (Williams, 1992; Spall, 1992) or maintains a negative inner product under reasonable assumptions (Bredenberg et al., 2021; Dayan et al., 1995; Ikeda et al., 1998; Meulemans et al., 2020). Mathematical analysis allows one to know quite clearly when a particular plasticity rule will decrease a loss function and identifies how plasticity mechanisms should change with changes in the network architecture or environment. However, analysis is often possible only under restrictive circumstances, and it is often necessary to supplement mathematical results with empirical simulations in order to demonstrate that the results extend to more general, more realistic circumstances.

2.2 Locality. Biological synapses can only change strengths using local chemical and electrical signals. “Locality” refers to the idea that a postulated synaptic plasticity mechanism should only refer to variables that could be conceivably available at a given synapse (see Figure 1b). Though locality may seem like an obvious biological requirement, it presents a great mystery: How does a system whose success or failure is determined by the joint action of many neurons distributed across the entire brain, improve performance through local changes? This is particularly puzzling, given that successful machine learning algorithms—including backpropagation (Werbos,

⁴ A negative inner product can also be achieved by taking $\Delta\mathbf{W}$ to be the negative loss gradient premultiplied by any positive-definite matrix, which could be dependent on the weights themselves. Updates of this form correspond to gradient descent with respect to different metrics (Surace et al., 2020); special cases include altering the learning rates for different parameters and natural gradient descent.

1974; Rumelhart et al., 1985) (see appendix B), backpropagation through time (Werbos, 1990), and real-time recurrent learning (Williams & Zipser, 1989)—need nonlocal propagation of learning signals.

Despite its importance as a guiding principle for normative theories of synaptic plasticity, locality is a slippery concept, primarily because of our limited understanding of the precise battery of biochemical signals available to a synapse and how those signals could be used to approximate quantities required by theories. As such, locality has resisted mathematical formalization until very recently (Bredenberg et al., 2023). Because of the complexities associated with assessing locality, normative theories typically declare success when some standard of plausibility is reached, where derived plasticity rules roughly match the experimental literature (Payeur et al., 2021) or only require reasonably simple functions of postsynaptic and presynaptic activity that a synapse could hypothetically approximate (Oja, 1982; Gerstner & Kistler, 2002; Scellier & Bengio, 2017; Williams, 1992).

In normative models of synaptic plasticity, the need for locality is in perpetual tension with the general need for some form of credit assignment (Lillicrap et al., 2020; Richards et al., 2019), a mechanism capable of signaling to a neuron that it is “responsible” for a network-wide error and should modify its synapses to reduce errors. Depending on a network’s objective, a system’s credit assignment mechanism could take a wide variety of forms, some small number of which may only require information about the pre- and postsynaptic activity of a cell (Oja, 1982; Pehlevan et al., 2015, 2017; Obeid et al., 2019; Brendel et al., 2020), but many of which appear to require the existence of some form of error (Scellier & Bengio, 2017; Lillicrap et al., 2016; Akroud et al., 2019) or reward-based (Williams, 1992; Fiete et al., 2007; Legenstein et al., 2010) signal.

The extent to which a credit assignment signal postulated by a normative theory meets the standards of locality depends heavily on the nature of the signal. For instance, there is growing support for the idea that neuromodulatory systems, distributing dopamine (Otani et al., 2003; Calabresi et al., 2007; Reynolds & Wickens, 2002), norepinephrine (Martins & Froemke, 2015), oxytocin (Marlin et al., 2015), and acetylcholine (Froemke et al., 2013; Guo et al., 2019; Hangya et al., 2015; Rasmusson, 2000; Shinoe et al., 2005) signals, can propagate information about reward (Guo et al., 2019), expectation of reward (Schultz et al., 1997), and salience (Hangya et al., 2015) diffusely throughout the brain to induce or modify synaptic plasticity in their targeted circuits. Therefore, it may be reasonable for normative theories to postulate that synapses have access to global reward or reward-like signals, without violating the requirement that plasticity be affected only by locally available information (Frémaux & Gerstner, 2016).

Locality as a desideratum serves as a heuristic stand-in for the requirement that a normative model must be eventually held to the standard of experimental evidence. This is not to say that normative models cannot postulate neural mechanisms that have not yet been observed experimentally.

However, for such an exercise to be constructive, the theory should clearly articulate how it deviates from the current state of the experimental field and how these deviations can be tested (see section 2.7; see appendix C for a concrete example of this process).

2.3 Architectural Plausibility. The learning algorithm implemented by a plasticity model can require specific architectural motifs to exist in a neural circuit in order to deliver reward, error, or prediction signals. These might include diffuse neuromodulatory projections (see Figure 4b) or neuron-specific top-down synapses onto apical dendrites (Richards & Lillicrap, 2019). Such architectural features are required for the learning algorithm in question and are known to exist in a wide range of cortical areas. However, normative plasticity models should not depend on circuit features that have been demonstrated not to exist in the modeled system, because spurious architectural features can be used to “cheat” at achieving locality by postulating unrealistic credit assignment mechanisms (see appendix B). In what follows, we highlight several particularly important architectural motifs that have been the focus of recent work.

Unlike the deterministic rate-based models typically used in machine learning, neurons communicate through discrete action potentials, with variability due to, for example, synaptic failures or task-irrelevant inputs (see Figure 2a; Faisal et al., 2008). Normative theories that employ rate-based activations (Bredenberg et al., 2020; Scellier & Bengio, 2017) or assume that the input-output function of neurons is approximately linear (Oja, 1982) may not extend to this more realistic discrete, stochastic, and highly nonlinear setting. Further, such theories inherently produce plasticity rules that ignore the precise relationship between pre- and postsynaptic spike times and will consequently be unable to capture spike-timing-dependent plasticity (STDP) phenomenology. Fortunately, learning rules that were originally formulated using rate-based models have subsequently been extended to spiking network models to great effect by leveraging methods that use stochasticity or explicit approximations to enable credit assignment through nondifferentiable spikes (Bohte et al., 2002; Pfister et al., 2006; Huh & Sejnowski, 2018; Shrestha & Orchard, 2018; Bellec et al., 2018; Neftci et al., 2019). Reward-based Hebbian plasticity based on the REINFORCE algorithm (see appendix C) (Williams, 1992) has been generalized to stochastic spiking networks (Pfister et al., 2006), while real-time recurrent learning approximations (Murray, 2019) and predictive coding methods (Rao & Ballard, 1999) have subsequently been extended to deterministic spiking networks (Bellec et al., 2020; Brendel et al., 2020). Therefore, a lack of a generalization to spiking networks is not necessarily a death knell for a normative theory, but many theories lack either an explicit generalization to spiking or a clear relationship to STDP, and the mathematical formalism that defines these methods may require significant modification to accommodate the change.

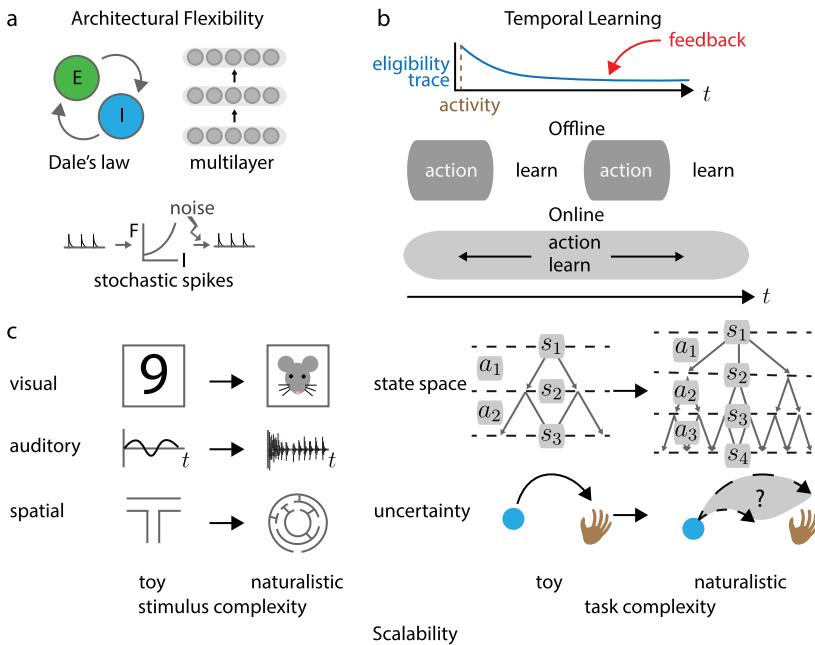


Figure 2: Architecture and scalability considerations for normative models. (a) Features of biological networks: separation of excitatory and inhibitory neuron populations, stochastic and spiking input-output functions for individual neurons, and multilayer, recurrent connectivity. (b) Mechanics of temporal credit assignment: eligibility traces store information about coactivity throughout time locally to a synapse and subsequently modify synaptic connections when paired with feedback information, either online or offline. (c) Increasing complexity in stimuli (left) and task structure (right). Different sensory features (e.g., visual, auditory, or spatial information) can all be made more naturalistic by matching real-world statistics. Task complexity can be modulated by increasing the number of action options (a) and sequential state (s) transitions required to achieve its goals or by introducing different forms of uncertainty.

Real biological networks have a diversity of cell types with different neurotransmitters and connectivity motifs. At the bare minimum, a normative model should be able to accommodate Dale's law (see Figure 2a), which stipulates that the neurotransmitters released by a neuron are either excitatory or inhibitory but not both (O'Donohue et al., 1985). Though this might seem like a simple principle, enforcing Dale's principle can seriously damage the performance of artificial neural networks without careful architectural considerations (Cornford et al., 2021). Furthermore, the mathematical results of many canonical models of synaptic modification rely on

symmetric connectivity between neurons, including Hopfield networks (Hopfield, 1982), Boltzmann machines (Ackley et al., 1985), contrastive Hebbian learning (Xie & Seung, 2003), and predictive coding (Rao & Ballard, 1999). This symmetry is partially related to the symmetric connectivity required by the backpropagation algorithm (see appendix B). Symmetric connectivity means that the connection from neuron A to neuron B must be the same as the reciprocal connection from neuron B to neuron A. It inherently violates Dale's law, because it means that entirely excitatory and entirely inhibitory neurons can never be connected to one another: the positive sign for one synapse and the negative sign for the reciprocal connection violates symmetry. Some models, such as Hopfield networks (Sompolinsky & Kanter, 1986) and equilibrium propagation (Ernoult et al., 2020) have been extended to demonstrate that moderate deviations from symmetry can exist and still preserve function. Further, a recent mathematical reformulation of predictive coding has demonstrated that interlayer symmetric connectivity is not necessary (Golkar et al., 2022). Therefore, recent results indicate that many canonical models believed to depend on symmetric connectivity can be improved on.

Many early plasticity models, including Oja's rule (Oja, 1982) and perceptron learning (Rosenblatt, 1958), as well as more modern model recurrent network models focused on learning temporal tasks (Murray, 2019), are designed to greedily optimize layer-wise objectives, and their mathematical justifications do not generalize to multilayer architectures. Though greedy layer-wise optimization may be sufficient for some forms of unsupervised learning (Illing et al., 2021), it is not clear how such greedy methods would be able to support many complex supervised or reinforcement learning tasks humans are known to learn (Lillicrap et al., 2020) that involve coordinating sensorimotor transformations across cortical areas (but see Zador, 2019). Generalizing layer-local learning to multilayer objective functions has been the focus of much recent work: many multilayer models can be seen as generalizations of perceptron learning (Bengio, 2014; Hinton et al., 1995; Rao & Ballard, 1999), with other models such as those derived from similarity matching (Pehlevan et al., 2017) or Bienenstock-Cooper-Munro theory (Bienenstock et al., 1982; Cooper, 2004; Intrator & Cooper, 1992) receiving similar treatment (Obeid et al., 2019; Halvagal & Zenke, 2023). We will refer to this form of multilayer signal propagation as "spatial" credit assignment, and will refer to relaying information across time as "temporal" credit assignment (see Figure 2b and section 2.4). As we will discuss in the next section, models that do not support temporal credit assignment are not able to account for learning in inherently sequential tasks.

2.4 Temporal Credit Assignment. Because so many learned biologically relevant tasks involving temporal decision making (Gold & Shadlen, 2007) or working memory (Compte et al., 2000; Wong & Wang, 2006; Ganguli et al., 2008) inherently leverage information from the past to inform

future behavior and because neural signatures associated with these tasks exhibit rich recurrent dynamics (Brody et al., 2003; Shadlen & Newsome, 2001; Mante et al., 2013; Sohn et al., 2019), many aspects of learning in the brain require a normative theory of synaptic plasticity that works in recurrent neural architectures and provides an account of temporal credit assignment.

As it currently stands, the majority of normative synaptic plasticity models focus on spatial credit assignment, which presents distinct challenges when compared to temporal credit assignment (Marschall et al., 2020). In fact, many theories that provide a potential solution to spatial credit assignment do so by requiring networks to relax to a steady state on a timescale much faster than inputs (Hopfield, 1982; Scellier & Bengio, 2017; Bredenberg et al., 2020; Xie & Seung, 2003; Ackley et al., 1985), which effectively prevents networks from having the rich, slow, internal dynamics required for many temporal motor (Hennequin et al., 2012) and working memory (Wong & Wang, 2006) tasks. Other methods appear to be agnostic to the temporal properties of their inputs but have not yet been combined with existing plasticity rules that perform approximate temporal credit assignment within local microcircuits (Murray, 2019; Bellec et al., 2020).

New algorithms do provide potential solutions to temporal credit assignment through either explicit approximation of real-time recurrent learning (Marschall et al., 2020; Bellec et al., 2020; Murray, 2019), by leveraging principles from control theory (Gilra & Gerstner, 2017; Alemi et al., 2018; Meulemans et al., 2022), or by leveraging principles of stochastic circuits that are fundamentally different from traditional explicit gradient-based calculation methods (Bredenberg et al., 2020; Miconi, 2017). Many use what is called “eligibility traces” (Izhikevich & Desai, 2003; Gerstner et al., 2018; see Figure 2b)—a local synaptic record of coactivity—to identify associations between rewards and neural activity that may have occurred much further in the past. We suggest that these models capture something fundamental about learning across time and that much work remains to combine these with spatial learning rules to construct normative models of full spatiotemporal learning.

2.5 Learning during Behavior. The relationship between learning and behavior can vary widely depending on the experimental context (see Figure 2b): learning-related changes can occur concomitantly with action (Bittner et al., 2015; Sheffield et al., 2017; Grienberger & Magee, 2022) (“online” learning), during brief periods of quiescence between trials (Pavlides & Winson, 1989; Bönstrup et al., 2019; Liu et al., 2021), or over periods of extended sleep (Gulati et al., 2017; Eschenko et al., 2008; Girardeau et al., 2009) (“offline” learning). Therefore, whether a normative plasticity model uses offline or online learning should be determined by the experimental context.

However, many classical algorithms—especially those that support multilayer spatial credit assignment (Ackley et al., 1985; Xie & Seung, 2003; Dayan et al., 1995)—are constrained to modeling only offline learning, because they require distinct training phases, during at least one phase of which activity of neurons is driven for learning, rather than performative purposes; these algorithms have begun to be extended to online learning only recently. For instance, algorithms such as Wake-Sleep (Hinton et al., 1995; Dayan et al., 1995) have been adapted such that the second phase becomes indistinguishable from perception (Bredenberg et al., 2020; Ernoult et al., 2020). Other recent models allow for simultaneous multiplexing of top-down learning signals and bottom-up inputs (Greedy et al., 2022), which enables online learning. These results suggest that future work may fruitfully adapt existing offline algorithms to provide good models of explicitly online learning in the brain.

2.6 Scalability in Dimensionality and Complexity. Models of brain learning need to be able to scale to handle the full complexity of the problems a given model organism has to solve. However, this is a point that can be difficult to verify: How can we guarantee that adding more neurons and more complexity will not make a particular collection of plasticity rules more effective? As a case study, consider REINFORCE (Williams, 1992), an algorithm that for the most part satisfies our other desiderata for normative plasticity for the limited selection of tasks in naturalistic environments that are explicitly rewarded (see appendix C). However, though REINFORCE demonstrably performs better than its precursor weight perturbation (Jabri & Flower, 1992), as the dimensionality of its stimuli, the number of neurons in the network and the delay time between neural activity and reward increase, the performance of the algorithm decays rapidly both analytically and in simulations (Werfel et al., 2003). This is primarily caused by the high variance of gradient estimates provided by the REINFORCE algorithm and is only partially ameliorated by methods that reduce its variance (Bredenberg et al., 2021; Ranganath et al., 2014; Mnih & Gregor, 2014; Miconi, 2017). Thus, adding more complexity to the network architecture actually impairs learning.

We do not mean to imply that all normative plasticity algorithms should be demonstrated to meet human-level performance or even that they should match state-of-the-art machine learning methods. Machine learning methods profit in many ways from their biological implausibility, and the human brain itself has orders of magnitude more neural units and synapses than have ever been simulated on a computer, all of them capable of processing totally in parallel. Therefore, direct comparison to the human—or any other—brain is also not fair. We propose the far softer condition that as the complexity of input stimuli and tasks increases, within the range supported by current computational power, plasticity rules derived from

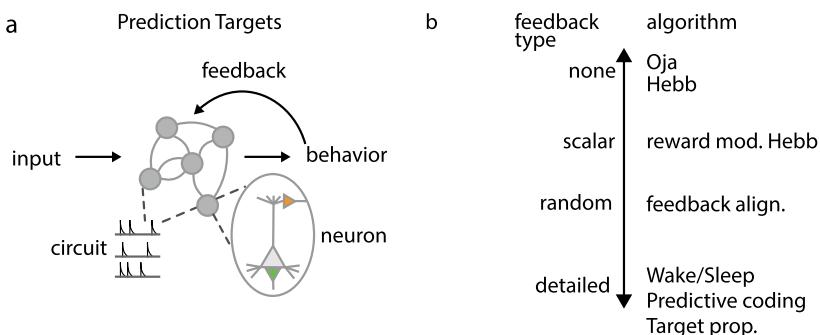


Figure 3: Testing normative theories. (a) Normative plasticity theories can be assessed through four different experimental lenses centered on individual neurons, circuits of collectively recorded neurons, the training signals delivered to a circuit, and the organism's overall behavior over the course of learning. (b) Different normative plasticity theories postulate different levels of detail for the feedback signals received by individual neurons.

normative theory should continue to perform well in both simulation and, preferably, analytically.

Complexity is multifaceted, and involves features of both stimulus and task (see Figure 2c). Even stimuli with very high-dimensional structure can fail to capture critical features of naturalistic stimuli, which can be much more difficult to learn from; for instance, existing plasticity models have great difficulty scaling to naturalistic image data sets (Bartunov et al., 2018). Further, in natural environments, rewards are often provided after long sequences of complex actions; supervised feedback is sparse, if present at all; and an organism's self-preservation often requires navigating both uncertainty and complex multi-agent interactions. Modern reinforcement learning algorithms are only just beginning to make progress with some of these difficulties (Kaelbling et al., 1998; Arjona-Medina et al., 2019; Raposo et al., 2021; Hung et al., 2019; Zhang et al., 2021), but as yet there are no normative plasticity models that describe how any of the human capabilities used to solve these problems could be learned through cellular adaptation. This suggests that scaling the ability of normative models to handle both complex stimuli and task structures is a major avenue of improvement for future methods.

2.7 Generating Testable Predictions. Testable predictions can be defined via several different experimental lenses, at the level of (1) individual neurons or synapses, (2) populations of neurons, (3) the feedback mechanisms that shape learning in neural circuits, and (4) learning at a behavioral level (see Figure 3a). Accurately distinguishing one mechanism from

another will likely require a synthesis of experiments spanning all four lenses.

2.7.1 Individual Neurons. Experiments that focus on individual neurons, including paired-pulse stimulation (Markram et al., 1997), mechanistic characterizations of plasticity (Graupner & Brunel, 2010), pharmacological explorations of neuromodulators that induce or modify plasticity (Bear & Singer, 1986; Reynolds & Wickens, 2002; Froemke et al., 2007; Gu & Singer, 1995), and characterization of local dendritic or microcircuit properties mediating plasticity (Froemke et al., 2005; Letzkus et al., 2006; Sjöström & Häusser, 2006) form the bulk of the classical literature underlying phenomenological and mechanistic modeling. These studies characterize what information is locally available at synapses and what can be done with that information, as well as which properties of cells can be altered in an experience-dependent fashion.

Existing normative theories differ in the nature of their predictions for plasticity at individual neurons. Reward-modulated Hebbian theories require that feedback information be delivered by a neuromodulator like dopamine, serotonin, or acetylcholine (Frémaux & Gerstner, 2016) and that this feedback modulates plasticity at the local synapse by changing the magnitude or sign of plasticity depending on the strength of feedback. In contrast, some unsupervised normative theories require no feedback modulation of plasticity (Pehlevan et al., 2015, 2017), and others argue that detailed feedback information arrives at the apical dendritic arbors of pyramidal neurons to modulate plasticity, which is also partially supported in the hippocampus (Bittner et al., 2015, 2017) and cortex (Larkum et al., 1999; Letzkus et al., 2006; Froemke et al., 2005; Sjöström & Häusser, 2006).

Independent of the exact feedback mechanism, models differ in how temporal associations are formed. Algorithms related to REINFORCE assume that local synaptic eligibility traces integrate over time fluctuations in coactivity of the post- and presynaptic neuron local to a synapse. These postulated eligibility traces are stochastic, summing gaussian fluctuations in activity (Miconi, 2017) that consequently produce temporal profiles similar to Brownian motion. In contrast, methods based on approximations to real-time recurrent learning propose eligibility traces that are deterministic records of coactivity whose time constants are directly connected to the dynamics of the neuron itself (Bellec et al., 2020), while other hybrid approaches predict eligibility traces that are deterministic but are related more to predicted task timescale than the dynamics of the cell (Roth et al., 2018). Though there do exist known cellular processes that naturally track coactivity, like NMDA receptors (Bi & Poo, 1998), and that store traces of this coactivity longitudinally, like CaMKII (Graupner & Brunel, 2010), how the properties of these known biophysical quantities relate to the predictions of various normative theories, and whether there are other biological alternatives remains unclear.

2.7.2 Neural Circuits. The functional effects of plasticity and their relationship to behavior manifest most directly at the level of neural populations (Marschall & Savin, 2023). Determining how circuits encode task-relevant information and affect motor actions requires methods that record large groups of neurons, such as 2-photon calcium imaging, multi-electrode recordings, fMRI, EEG, and MEG, as well as methods that manipulate large populations, like optogenetic (Rajasethupathy et al., 2016) stimulation.

First, a population-level lens is useful for evaluating hypotheses about the nature of the objective function, where one starts by training neural networks on a battery of objectives and tests which objective produces the closest correspondence between neural activity in the model and that recorded in the brain. This approach has been used in the ventral (Yamins et al., 2014; Yamins & DiCarlo, 2016) and dorsal (Mineault et al., 2021) visual streams, as well as in auditory cortex (Kell et al., 2018) and medial entorhinal cortex (Nayebi et al., 2021). Often changes in artificial neural network activity throughout time are sufficient to determine the objective optimized by the network as well as its learning algorithm (Nayebi et al., 2020), an approach that could also potentially be applied to recorded neural activity over learning.

Second, circuit recordings could test predictions about the existence of different phases of the dynamics, as required by some normative models. For instance, the Wake-Sleep algorithm (Dayan et al., 1995) proposes that neural circuits should spend extended periods of time (e.g., during dreaming) generating similar activity patterns to those evoked by natural stimulus sequences. There is plenty of room for experiments to more clearly map predictions and components of similar normative models onto well-documented neural phenomena, such as sleep or potentially replay phenomena (Girardeau et al., 2009; Eschenko et al., 2008).

Finally, some algorithms make specific predictions about inhibitory microcircuitry. Impression learning, for instance, suggests that a population of inhibitory interneurons could gate the influence of apical and basal dendritic inputs to the activity of pyramidal neurons (Bredenberg et al., 2021), and some learning algorithms propose that top-down error signals are partially computed by local inhibitory interneurons (Sacramento et al., 2017; Greedy et al., 2022). Therefore, to completely distinguish different theories, it may be necessary to analyze the connectivity and plasticity between small groups of different cell types. Because circuit recording and manipulation methods often sacrifice temporal resolution (Hong & Lieber, 2019) and have difficulty inferring biophysical properties of individual synapses and cells, these methods are best used in concert with single neuron studies to jointly tease apart the multilevel predictions of various normative models.

2.7.3 Feedback Mechanisms. The most direct way to distinguish normative plasticity algorithms is on the basis of the nature of their feedback

mechanisms (see Figure 3b). Though no feedback is necessary for some unsupervised algorithms, like Oja's rule, any form of supervised or reinforcement learning will require some form of top-down feedback. However, across models, the level of precision of feedback varies considerably. The simplest feedback is scalar, conveying reward (Williams, 1992), state fluctuation (Payeur et al., 2021), or context (e.g., saccade; Illing et al., 2021) or attention (Roelfsema & Ooyen, 2005; Pozzi et al., 2020) information. Beyond this, the space of proposed mechanisms expands considerably: back-propagation approximations like feedback alignment (Lillicrap et al., 2016) and random-feedback online learning (RFLO) (Murray, 2019) propose that random error feedback between layers of neurons can provide a sufficient learning signal, whereas algorithms based on control theory propose that low-rank or partially random projections carrying supervised error signals are sufficient (Gilra & Gerstner, 2017; Alemi et al., 2018). Other algorithms propose even more detailed feedback, with individual neurons receiving precise, carefully adapted projections carrying learning-related information. These algorithms propose that top-down projections to apical dendrites (Urbanczik & Senn, 2014) or local interneuron neurons (Bastos et al., 2012) perform spatial credit assignment, but the nature of this signal can differ considerably across different algorithms. It could be a supervised target, carrying information about what the neuron state "should" be to achieve a goal (Guerguiev et al., 2017; Payeur et al., 2021), or it could be a prediction of the future state of the neuron (Bredenberg et al., 2021).

So far, different feedback mechanisms have received only partial support. For example, acetylcholine projections to auditory cortex could subserve a form of reward-based learning: they modulate perceptual learning (Froemke et al., 2013) and display a diversity of responses related to both reward and attention (Hangya et al., 2015), but contrary to simple reward-based learning algorithms, these response properties adapt over the course of learning in concert with auditory cortex (Guo et al., 2019). This suggests that while traditional models of reward-modulated Hebbian plasticity may be correct to a first approximation, a more detailed study of the adaptive capabilities of neuromodulatory centers may be necessary to update the theories.

While a growing number of studies indicate that projections to apical synapses of pyramidal neurons do play a role in inducing plasticity and that these projections themselves are also plastic (i.e., nonrandom; Bittner et al., 2015, 2017), very little is known about the nature of the signal—a critical component for distinguishing several different theories. In the visual system, presentation of unfamiliar images without any form of reward or supervision can modify both apical and basal dendrites throughout time (Gillon et al., 2021), and in the hippocampus, apical input to CA1 pyramidal neurons while animals acclimatize to new spatial environments is sufficient to induce synaptic plasticity (Bittner et al., 2015, 2017). There is further evidence for explicit motor error signals carried by climbing fiber pathways

in the cerebellar system being used for plasticity (Gao et al., 2012; Bouvier et al., 2018).

In biofeedback training settings, animals can selectively control the firing rates of individual neurons to satisfy arbitrary experimental conditions for reward (Fetz, 2007), suggesting the existence of highly flexible credit assignment systems, which are not constrained by evolutionary predetermination.⁵ Other brain-computer interface (BCI) experiments more directly quantify the limits of this flexibility. In particular, animals have been shown to adapt more easily to BCI decoder perturbations that occur within the manifold of neural activity, relative to outside of manifold perturbations (Sadler et al., 2014), which may be reflective of constraints on the credit assignment system (Feulner & Clopath, 2021) (but see Humphreys et al., 2022; Payeur et al., 2023). Moreover, recent evidence suggests that apical dendrites may receive precise learning signals in the retrosplenial cortex during BCI tasks (Francioni et al., 2023), which could underlie these remarkable capabilities.

2.7.4 Behavior. In much the same way that psychophysical studies of human or animal responses define constraints on what the brain's perceptual systems are capable of, behavioral studies of learning can do quite a lot to describe the range of phenomena that a model of learning must be able to capture, from operant conditioning (Niv, 2009), to model-based learning (Doll et al., 2012), rapid language learning (Heibeck & Markman, 1987), unsupervised sensory development (Wiesel & Hubel, 1963), or consolidation effects (Stickgold, 2005). Behavioral studies can also outline key limitations in learning, which are perhaps reflective of the brain's learning algorithms—for example, the brain's failure to perform certain types of adaptation after critical periods of plasticity (Wiesel & Hubel, 1963).

These existing experimental results stand as (often unmet) targets for normative theories of plasticity, but in addition, normative theories themselves suggest further studies that may test their predictions. In particular, manipulation of learning mechanisms may have predictable effects on animals' behavior, as seen when acetylcholine receptor blockage in mouse auditory cortex prevented reward-based learning in animals (Guo et al., 2019) and nucleus basalis stimulation during tone perception longitudinally improved animals' discrimination of that tone (Froemke et al., 2013). Other algorithms have as-yet-untested predictions for behavior; for instance, experimentally increasing the influence of top-down projections should bias behavior toward commonly occurring sensory stimuli according to both predictive coding (Rao & Ballard, 1999; Friston, 2010) and impression learning (Bredenberg et al., 2021). For other detailed feedback algorithms (see Figure 3b), manipulating top-down projections may disrupt learning but

⁵This is a challenge for normative plasticity models that predefine the outputs of the circuit and approximately backpropagate errors from these outputs.

would have a much more unstructured deleterious effect on perceptual behavior.

Overall, each experimental lens has its own advantages and disadvantages. Single-neuron studies are excellent for identifying the locally available variables that affect plasticity, circuit-level studies can help narrow down the objectives that shape neural responses and identify traces of offline learning, studies of feedback mechanisms can distinguish among different algorithms that postulate different degrees of precision in their feedback and in complexity of the teaching signal, and studies of behavior can place boundaries on what can be learned, as well as serve as a readout for manipulations of the mechanisms underlying learning. Each focus alone is insufficient to distinguish among all existing normative models, but in concert they show promise for identifying the neural substrates of adaptation.

3 Conclusion

Normative models of plasticity are compelling because of their potential to connect our brains' capacity for adaptation to their constituent synaptic modifications. Generating good theories is a critical part of the scientific process, but finding ways to close the loop by testing key predictions of new normative models has proved extraordinarily difficult. In this perspective, we have illustrated some of the sources of this difficulty, have shown how recent work has progressed on these fronts, and have identified ways forward for future models.

The core of a normative plasticity model is its plasticity rule, which dictates how a model synapse modifies its strength. To be a normative model—to explain why the plasticity mechanism is important for the organism—there must be a concrete demonstration that this plasticity rule supports adaptation critical for system-wide goals like processing sensory signals or obtaining rewards (see section 2.1). However, this system-wide goal must be achieved using only local information (see section 2.2). These two needs of a normative plasticity model are the fundamental source of tension: it is very difficult to demonstrate that a proposed plasticity rule is both local *and* optimizes a system-wide objective (see appendix B). Insufficient or partial resolution of this fundamental tension produces normative models that satisfy the other desiderata to a lesser degree; namely, they struggle to map accurately onto neural hardware (see section 2.3) or handle complex temporal stimuli and tasks online (see sections 2.4 to 2.6). To provide a case study of how our desiderata come to be satisfied (or not) in practice, we have included a tutorial for the REINFORCE algorithm in appendix C.

In this review, we have organized emerging theories according to how they satisfy and improve on our desiderata, as well as by how they can be tested. Theoreticians can use our desiderata (see section 2.1 to 2.6) and Table 1 as guides for where theoretical development is needed in order

Table 1: Summarizing Progress on the Desiderata.

Algorithm	Improv. Perf.	Local.	Arch. Plaus.	Temp. Credit	Learning During Behavior	Scalability in Dim. & Complexity
Backprop. (BP) (Werbos, 1974)	U/S/R (Williams, 1992)	✗	✓ (Lee et al., 2016)	✓ (Werbos, 1990)	✗	✓
REINFORCE (Williams, 1992)	U/S/R	✓	✓	✓ (Miconi, 2017)	✓	✗ (Werfel et al., 2003)
Oja (Oja, 1982)	U	✓	✗	✗	✓	✓
Pred. Coding (Rao & Ballard, 1999)	U/S (Whittington & Bogacz, 2017)	✓	✗	✓ (Friston & Kiebel, 2009)	✓	✓
Wake-Sleep (Dayan et al., 1995)	U	✓	✓ (Dayan & Hinton, 1996)	✓ (Dayan & Hinton, 1996)	✓	✓
Approx. Gradient (Lillicrap et al., 2016) (Akrout et al., 2019)	U/S* (Bellic et al., 2020)	✓	✓ (Bellic et al., 2020)	✓ (Murray, 2019) (Bellic et al., 2020)	✓ (Murray, 2019) (Bellic et al., 2020)	✓
Equil. Prop. (Scellier & Bengio, 2017)	U/S	✓	✗	✗	✓ (Emoult et al., 2020)	✓ (Laborie et al., 2021)
Target Prop. (Bengio, 2014)	U/S	✓	✓	✓ (Manchev & Spratling, 2020)	✗	✓ (Lee et al., 2015)

Notes: A ✓ indicates that an algorithm has been demonstrated to satisfy a particular desideratum in at least one study, whereas an ✗ indicates that it has not been demonstrated. If the demonstrating study is an improvement on the seminal work or is a new model, we provide a citation. Asterisks indicate that results have only been shown by simulation and lack mathematical support. U, S, and R indicate whether a given algorithm supports unsupervised, supervised, or reinforcement learning, respectively.

Table 2: Examples of Testable Predictions for Normative Plasticity Models.

Algorithm	Testable Predictions
REINFORCE (Williams, 1992)	Reward signals modulate plasticity
Oja (Oja, 1982)	Stochastic eligibility traces Exclusively Hebbian plasticity
Pred. Coding (Rao & Ballard, 1999)	Feedforward propagation of prediction errors Approx. symmetric feedback connectivity
Wake-Sleep (Dayan et al., 1995)	Offline generative replay driven by top-down inputs Top-down predictive inputs drive bottom-up plasticity
Approx. Gradient (Lillicrap et al., 2016; Akrout et al., 2019)	Neuron-specific top-down errors drive plasticity Smooth eligibility traces
Equil. Prop. (Scellier & Bengio, 2017)	The sign of plasticity changes while receiving instructive feedback
Target Prop. (Bengio, 2014)	Top-down target inputs drive bottom-up plasticity

to render normative models more biologically accurate and easier to test, while experimentalists can use the summary of their experimental predictions (see Table 2) to identify tests that distinguish different normative models from one another in specific neural systems. Even if existing algorithms prove not to be implemented exactly in the brain, they can provide key insights into how local synaptic modifications can produce valuable improvements in both behavior and perception for an organism. It seems sensible to use these algorithms as a springboard to produce more biologically realistic and powerful theories.

As the diversity of the experimental preparations suggests, there are increasingly strong arguments for several fundamentally different plasticity algorithms instantiated in different areas of the brain and across different organisms, subserving different functions. It is quite likely that many plasticity mechanisms work in concert to produce learning as it manifests in our perception and behavior. It is our belief that well-articulated normative theories can serve as the building blocks of a conceptual framework that tames this diversity and allows us to understand the brain's tremendous capacity for adaptation.

Appendix A: The Unidentifiability of an Objective

In this section we illustrate why the choice of objective function for a normative plasticity model is never uniquely determined by data. We consider two situations: the system has already settled to its optimal setting of its

weights, \mathbf{W}^* , and in the second we are able to observe the system's plasticity update $\Delta\mathbf{W}$.

A.1 Unidentifiability Based on an Optimum. Suppose that some setting of synaptic weights \mathbf{W}^* minimizes an objective function \mathcal{L} : $\mathcal{L}(\mathbf{W}^*) \leq \mathcal{L}(\mathbf{W}) \forall \mathbf{W}$. We might be tempted to argue that because \mathbf{W}^* minimizes \mathcal{L} , \mathcal{L} must be *the* objective that the system is minimizing. However, an infinite variety of alternative objectives shares the same minimum. To see this, take a new objective $\tilde{\mathcal{L}} = \sigma(\mathcal{L}(\mathbf{W}))$ for any differentiable, monotonically increasing function $\sigma(\cdot)$. Then we have

$$\mathcal{L}(\mathbf{W}^*) \leq \mathcal{L}(\mathbf{W}) \forall \mathbf{W} \quad (\text{A.1})$$

$$\Rightarrow \sigma(\mathcal{L}(\mathbf{W}^*)) \leq \sigma(\mathcal{L}(\mathbf{W})) \forall \mathbf{W} \quad (\text{A.2})$$

$$\Rightarrow \tilde{\mathcal{L}}(\mathbf{W}^*) \leq \tilde{\mathcal{L}}(\mathbf{W}) \forall \mathbf{W}, \quad (\text{A.3})$$

where the second equality follows from the order preservation property of $\sigma(\cdot)$. This means that \mathbf{W}^* also minimizes $\tilde{\mathcal{L}}$, that is, we will be unable to arbitrate between whether the system is "attempting" to minimize $\tilde{\mathcal{L}}$ or \mathcal{L} on the basis of the optimized network state given by \mathbf{W}^* .

A.2 Unidentifiability Based on an Update Rule. Suppose instead that we were able to observe the adaptive plasticity mechanism of a system and were able to verify that it really does decrease an objective function \mathcal{L} , that is, by equation 2.2,

$$\frac{d\mathcal{L}}{d\mathbf{W}}(\mathbf{W})^T \Delta\mathbf{W} \leq 0 \forall \mathbf{W}. \quad (\text{A.4})$$

We might now be tempted to argue that by observing the plasticity rule itself, $\Delta\mathbf{W}$, we will be better able to assert that the system, by virtue of consistently decreasing \mathcal{L} , is "attempting" to minimize \mathcal{L} . However, the exact same family of alternative objectives will also be minimized ($\tilde{\mathcal{L}} = \sigma(\mathcal{L}(\mathbf{W}))$ for any differentiable, monotonically increasing function $\sigma(\cdot)$). To see this, we observe

$$\frac{d\mathcal{L}}{d\mathbf{W}}(\mathbf{W})^T \Delta\mathbf{W} \leq 0 \forall \mathbf{W} \quad (\text{A.5})$$

$$\Rightarrow \frac{d\sigma(\mathcal{L}(\mathbf{W}))}{d\mathcal{L}(\mathbf{W})} \frac{d\mathcal{L}}{d\mathbf{W}}(\mathbf{W})^T \Delta\mathbf{W} \leq 0 \forall \mathbf{W} \quad (\text{A.6})$$

$$\Rightarrow \frac{d\tilde{\mathcal{L}}}{d\mathbf{W}}(\mathbf{W})^T \Delta\mathbf{W} \leq 0 \forall \mathbf{W}, \quad (\text{A.7})$$

where the first implication follows from the fact that $\sigma(\cdot)$ is differentiable and increasing (it has strictly positive derivative), and the second implication follows from the chain rule. This implies that plasticity rules ($\Delta\mathbf{W}$) and

trained neural circuits (\mathbf{W}^*) can at most partially constrain the space of viable objective functions the system could be minimizing.

Appendix B: Why Can't the Brain Do Explicit Gradient Descent? _____

We have provided one surefire way to decrease an objective function by modifying the parameters of a neural network—simply take small steps in the direction of the gradient of the loss (see section 2.1). To appreciate the challenges faced by theories of normative plasticity, it's important to understand why a biological system could not do this. In this section we provide a simplified argument as to why gradient descent within multilayer neural networks produces parameter updates, thus failing our most critical desideratum for a normative plasticity theory (see section 2.2). More detailed arguments for multilayer neural networks can be found here (Lillicrap et al., 2020), and descriptions of why gradient descent becomes even more implausible for recurrent neural networks trained with either back-propagation through time (Werbos, 1990) or real-time recurrent learning (Williams & Zipser, 1989) can be found here (Marschall et al., 2020).

The weight transport problem is the most basic reason that gradient descent is implausible for neural networks. Suppose that we have a stimulus-dependent network response, $\mathbf{r}(\mathbf{W}^{\text{in}}) = f(\mathbf{W}^{\text{in}} \mathbf{s})$, where \mathbf{r} is an $N \times 1$ vector and \mathbf{W}^{in} is an $N \times N^s$ weight matrix mapping stimuli \mathbf{s} into responses after a pointwise nonlinearity $f(\cdot)$. This network response is decoded into a network output, $o(\mathbf{W}^{\text{in}}, \mathbf{s}) = \mathbf{W}^{\text{out}} \mathbf{r}(\mathbf{W}^{\text{in}})$, where \mathbf{W}^{out} is a $1 \times N$ vector mapping network responses into a scalar output. Now suppose for simplicity that our loss for a single stimulus example is given by

$$\mathcal{L} = \frac{1}{2} (\hat{o} - o(\mathbf{W}^{\text{in}}, \mathbf{s}))^2. \quad (\text{B.1})$$

This objective is trying to bring the stimulus-dependent network response $o(\mathbf{W}^{\text{in}}, \mathbf{s})$ close to the target output \hat{o} and is zero if and only if $o = \hat{o}$. A reasonable hypothesis would be that the gradient of this objective function with respect to a synaptic weight, $\mathbf{W}_{ij}^{\text{in}}$, will produce a parameter update that is local: we will see that this is not true. Taking the gradient, we have

$$\frac{\partial}{\partial \mathbf{W}_{ij}^{\text{in}}} \mathcal{L} = \frac{1}{2} \frac{\partial}{\partial \mathbf{W}_{ij}^{\text{in}}} (\hat{o} - o(\mathbf{W}^{\text{in}}, \mathbf{s}))^2 \quad (\text{B.2})$$

$$= (\hat{o} - o) \frac{\partial}{\partial \mathbf{W}_{ij}^{\text{in}}} o(\mathbf{W}^{\text{in}}, \mathbf{s}) \quad (\text{B.3})$$

$$= (\hat{o} - o) \mathbf{W}_i^{\text{out}} \frac{\partial}{\partial \mathbf{W}_{ij}^{\text{in}}} f_i(\mathbf{W}^{\text{in}} \mathbf{s}) \quad (\text{B.4})$$

$$= (\hat{o} - o) \mathbf{W}_i^{\text{out}} f'_i(\mathbf{W}^{\text{in}} \mathbf{s}) \mathbf{s}_j. \quad (\text{B.5})$$

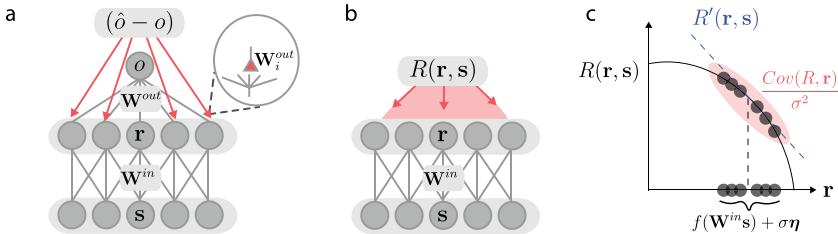


Figure 4: Weight transport and REINFORCE. (a) Traditional gradient descent propagates a credit assignment signal $(\hat{o} - o)W_i^out$ to each neuron r_i . How this pathway could have access to W_i^out is unclear: this is the weight transport problem. (b) REINFORCE resolves the weight transport problem by projecting a scalar reward signal $R(r, s)$ to all synapses. (c) By correlating this reward with fluctuations in neural activity, neurons can approximate the true gradient.

Breaking down this final update, we can see three terms: an error, $(\hat{o} - o)$, the neuron's output weight W_i^out , and an approximately Hebbian term $f'_i(W^in s) \mathbf{s}_j$, which requires only a combination of pre- and postsynaptic activity. One might be tempted to organize the plasticity rule into an error feedback signal received by the neuron, scaled by a neuron-specific synaptic weight W_i^out , and then combined with Hebbian coactivity to produce a synaptic update (see Figure 4a). This would have the form of a three-factor plasticity rule (Frémaux & Gerstner, 2016), combining weighted feedback with pre- and postsynaptic activity. However, the weight transport problem is as follows: W_i^out provides the strength of a synapse in the feedforward pathway. How could it possibly come to be that a feedback learning pathway would have access to the same synaptic weight? The answer is that there is no evidence for such a system of weight sharing across feed-forward and feedback pathways in the brain, though there are many hypotheses about how such a system could in theory be approximated by a normative plasticity algorithm. This problem becomes more pronounced in multilayer networks, where the error signal must be propagated through many interconnected connectivity layers.

It is also worth noting two key differentiability assumptions inherent in this approach. For one, we assume not only that the loss function \mathcal{L} is differentiable, but that some “error calculating” part of the brain does differentiate it. This requires knowledge of what the desired network output \hat{o} should be, which for many real-world tasks is not possible. Second, we assume that the network activation function $f(\cdot)$ is differentiable. Since neurons typically emit binary spikes, this differentiability assumption is not necessarily valid, though several modern methods have circumvented this problem by using either stochastic neuron models (Williams, 1992; Dayan & Hinton, 1996) or clever optimization tricks (Bellic et al., 2020). In

subsequent sections, we describe one canonical algorithm that employs clever tricks to circumvent the weight transport problem.

Appendix C: REINFORCE

In this section, we provide a mathematical tutorial on the REINFORCE learning algorithm (Williams, 1992), a mechanism for updating the parameters in a stochastic neural network for reinforcement learning objective functions. Its chief advantages are twofold. First, it only requires you to be able to evaluate an objective function (the reward received on any given trial), not the gradient of the objective function with respect to the parameters (see Figure 4b). This is very useful in situations in which the relationship between rewards and network outputs is not clear to an agent, as would be the case in many reinforcement learning scenarios. Second, under a broad range of biologically reasonable assumptions about a neural network architecture, the parameter updates produced by this algorithm are local, meaning the information required for a parameter update would reasonably be available to a synapse in the brain. This algorithm produces updates that are within the class of reward-modulated Hebbian plasticity rules. The chief disadvantage of this algorithm is its comparative data inefficiency relative to backpropagation. In practice, far more data samples (or, equivalently, much lower learning rates) will be required to produce the same improvements in performance compared to backpropagation (Werfel et al., 2003).

The REINFORCE algorithm and minor variations appear in different fields with different names. It is useful to keep track of these alternative names because they all use roughly the same derivation, with some improvements or field-specific modifications. In machine learning, the algorithm is often referred to as *node perturbation* (Richards et al., 2019; Lillicrap et al., 2020; Werfel et al., 2003), because it involves correlating fluctuations in neuron (node) activity with reward signals. In computational neuroscience, it is sometimes called *3-factor* or *reward-modulated Hebbian* plasticity (Frémaux & Gerstner, 2016), though REINFORCE is only one of several algorithms referred to by these blanket terms. In reinforcement learning, REINFORCE is often treated as a member of the more general class of *policy gradient* (Sutton & Barto, 2018) methods, which can be used to train any parameterized stochastic agent through reinforcement. Policy gradient methods need not commit to a neural network architecture and are consequently not always local. Finally, very similar methods are used for fitting variational Bayesian models and are in these contexts referred to as either *black box variational inference* (Ranganath et al., 2014) or *neural variational inference* (Mnih & Gregor, 2014).

In what follows, we provide a brief derivation of the REINFORCE learning algorithm for a one-layer feedforward neural network. We then discuss

the many extensions of the algorithm as well as its strengths and limitations as a normative plasticity model.

C.1 Network Model. Most neural networks used in machine learning are deterministic. However, neurons in biological systems fluctuate across trials and stimulus presentations, so modeling them as stochastic is often more appropriate. It will turn out that these fluctuations can be used to produce parameter updates in a way that a deterministic system could not.

First, we assume that there are stimuli drawn from some stimulus distribution, $p(\mathbf{s})$, and we will define the neural network response to a given stimulus drawn from this distribution as

$$\mathbf{r} = f(\mathbf{W}^{in} \mathbf{s}) + \sigma \boldsymbol{\eta}, \quad (C.1)$$

where the $\boldsymbol{\eta}$ is the source of random fluctuations, which, for simplicity, is drawn from a standard normal distribution ($\mathcal{N}(0, 1)$). In this equation, \mathbf{s} is an $N_s \times 1$ vector, \mathbf{W}^{in} is an $N_r \times N_s$ matrix, $f(\cdot)$ is the tanh nonlinearity, and $\boldsymbol{\eta}$ is an $N_r \times 1$ vector.

This equation defines a conditional probability distribution, $p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) \sim \mathcal{N}(f(\mathbf{W}^{in} \mathbf{s}), \sigma^2)$. There is an interesting point here: neuron activities are now samples from this conditional probability distribution, and so we can study how neurons behave on average by taking expectations over the probability distribution.

For simplicity and clarity, we restrict ourselves to this neural architecture for our derivation, but the basic principles apply more generally to a variety of noise sources and neural architectures (see section 3).

C.2 Defining the Objective. We assume that our goal is to maximize some instantaneous reward $R(\mathbf{r}, \mathbf{s})$ on average across many different samples of $R(\mathbf{r}, \mathbf{s})$ and \mathbf{s} . This allows us to write our objective function $\mathcal{O}(\mathbf{W}^{in})$ as

$$\mathcal{O}(\mathbf{W}^{in}) = \int R(\mathbf{r}, \mathbf{s}) p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) p(\mathbf{s}) d\mathbf{r} d\mathbf{s}. \quad (C.2)$$

In practice, this integral might be analytically impossible to integrate, but we can always approximate it (because it is an expectation) using samples from $p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in})$ and $p(\mathbf{s})$ as an empirical average over K samples \mathbf{r}_k and \mathbf{s}_k :

$$\mathcal{O}(\mathbf{W}^{in}) \approx \frac{1}{K} \sum_{k=0}^K R(\mathbf{r}^{(k)}, \mathbf{s}^{(k)}). \quad (C.3)$$

Procedurally, this would amount to sampling \mathbf{s} and \mathbf{r} each K times, calculating the reward for each trial, and taking an average.

C.3 Taking the Gradient. Now that we have our objective function, we can evaluate its derivative with respect to a particular synapse \mathbf{W}_{ij}^{in} in the network:

$$\frac{\partial \mathcal{O}(\mathbf{W}^{in})}{\partial \mathbf{W}_{ij}^{in}} = \frac{\partial}{\partial \mathbf{W}} \int R(\mathbf{r}, \mathbf{s}) p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) p(\mathbf{s}) d\mathbf{r} d\mathbf{s} \quad (\text{C.4})$$

$$= \int R(\mathbf{r}, \mathbf{s}) \left[\frac{\partial}{\partial \mathbf{W}_{ij}^{in}} p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) \right] p(\mathbf{s}) d\mathbf{r} d\mathbf{s}. \quad (\text{C.5})$$

We could theoretically stop here and evaluate $\frac{\partial}{\partial \mathbf{W}_{ij}^{in}} p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in})$ explicitly. However, in the same way that we can approximate $\mathcal{O}(\mathbf{W}^{in})$ as an empirical average over samples, we would like to be able to approximate our derivative as an average. To do this requires us to keep our loss in the form of an expectation over $p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in})p(\mathbf{s})$. We notice a convenient identity: $\frac{\partial}{\partial \mathbf{W}_{ij}^{in}} p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) = \frac{\partial}{\partial \mathbf{W}_{ij}^{in}} \exp(\log p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in})) = \left[\frac{\partial}{\partial \mathbf{W}_{ij}^{in}} \log p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) \right] p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in})$, which is a simple application of the chain rule. Inserting this identity into the above equation, we get

$$\frac{\partial \mathcal{O}(\mathbf{W}^{in})}{\partial \mathbf{W}_{ij}^{in}} = \int R(\mathbf{r}, \mathbf{s}) \left[\frac{\partial}{\partial \mathbf{W}_{ij}^{in}} \log p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) \right] p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) p(\mathbf{s}) d\mathbf{r} d\mathbf{s} \quad (\text{C.6})$$

$$\approx \frac{1}{K} \sum_{k=0}^K R(\mathbf{r}^{(k)}, \mathbf{s}^{(k)}) \left[\frac{\partial}{\partial \mathbf{W}_{ij}^{in}} \log p(\mathbf{r}^{(k)}|\mathbf{s}^{(k)}; \mathbf{W}^{in}) \right]. \quad (\text{C.7})$$

Though this is an approximation, we note that by the law of large numbers, we can improve its accuracy arbitrarily by increasing our number of samples K . In practice, however, taking $K = 1$ will prove to be the most straightforward way to get an update that is local in time; although such an update will still on average match the true gradient exactly, its high variance can lead to very inefficient learning.

We have left the derivation completely general up until this point. Different choices of $p(\mathbf{r}|\mathbf{s}; \mathbf{W})$ will produce different updates. Our particular choice gives

$$\frac{\partial}{\partial \mathbf{W}_{ij}^{in}} \log p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) = \frac{\partial}{\partial \mathbf{W}_{ij}^{in}} \sum_{i=0}^{N_r} \frac{1}{2\sigma^2} (\mathbf{r}_i - f_i(\mathbf{W}^{in} \mathbf{s}))^2 + C \quad (\text{C.8})$$

$$= \frac{1}{\sigma^2} \sum_{i=0}^{N_r} (\mathbf{r}_i - f_i(\mathbf{W}^{in} \mathbf{s})) \frac{\partial f_i(\mathbf{W} \mathbf{s})}{\partial \mathbf{W}_{ij}^{in}}. \quad (\text{C.9})$$

For a particular weight \mathbf{W}_{ij}^{in} , $\frac{\partial f_l(\mathbf{W}^{in} \mathbf{s})}{\partial \mathbf{W}_{ij}^{in}} = 0$ if $i \neq l$, so we have

$$\frac{\partial}{\partial \mathbf{W}_{ij}^{in}} \log p(\mathbf{r}|\mathbf{s}; \mathbf{W}) = \frac{1}{\sigma^2} (\mathbf{r}_i - f_i(\mathbf{W}\mathbf{s})) f'_i(\mathbf{W}\mathbf{s}) \mathbf{s}_j. \quad (\text{C.10})$$

Plugging this equation into equation C.4 gives the following parameter update:

$$\begin{aligned} \Delta \mathbf{W}_{ij}^{in} &\propto \frac{1}{K} \sum_{k=0}^K R(\mathbf{r}^{(k)}, \mathbf{s}^{(k)}) \left[\frac{1}{\sigma^2} (\mathbf{r}_i^{(k)} - f_i(\mathbf{W}^{in} \mathbf{s}^{(k)})) f'_i(\mathbf{W}^{in} \mathbf{s}^{(k)}) \mathbf{s}_j^{(k)} \right] \\ &\approx \frac{\partial \mathcal{O}(\mathbf{W}^{in})}{\partial \mathbf{W}_{ij}^{in}}. \end{aligned} \quad (\text{C.11})$$

If we want to update all of our parameters simultaneously using parallelized matrix operations, we can write this as an outer product:

$$\Delta \mathbf{W}^{in} \propto \frac{1}{K} \sum_{k=0}^K R(\mathbf{r}^{(k)}, \mathbf{s}^{(k)}) \left[\frac{1}{\sigma^2} (\mathbf{r}^{(k)} - f(\mathbf{W}^{in} \mathbf{s}^{(k)})) \odot f'(\mathbf{W}^{in} \mathbf{s}^{(k)}) \right] \mathbf{s}^{(k)T}, \quad (\text{C.12})$$

where \odot denotes a Hadamard (elementwise) vector product. Interestingly, the $\frac{1}{\sigma^2} (\mathbf{r} - f(\mathbf{W}^{in} \mathbf{s}))$ term here is exactly equal to η .

C.4 Why Don't We Need the Derivative of the Loss? One way of interpreting this parameter update is that neural units are correlating fluctuations in their neural activity with the rewards received to approximate $\frac{\partial R(\mathbf{r}, \mathbf{s})}{\partial \mathbf{r}}$ (see Figure 4c). To see this, first notice that

$$\mathbb{E} \left[b \left[\frac{1}{\sigma^2} (\mathbf{r} - f(\mathbf{W}^{in} \mathbf{s})) \odot f'(\mathbf{W}^{in} \mathbf{s}) \right] \mathbf{s}^T \right]_{p(\mathbf{r}|\mathbf{s})} = 0, \quad (\text{C.13})$$

for any constant b , because $\mathbb{E} [\mathbf{r} - f(\mathbf{W}^{in} \mathbf{s})]_{p(\mathbf{r}|\mathbf{s})} = 0$. If we take $b = \mathbb{E} [R(\mathbf{r}, \mathbf{s})]_{p(\mathbf{r}|\mathbf{s})}$, then we can rewrite the gradient without changing its expected value:

$$\begin{aligned} \frac{\partial \mathcal{O}(\mathbf{W}^{in})}{\partial \mathbf{W}_{ij}^{in}} &= \int (R(\mathbf{r}, \mathbf{s}) - \mathbb{E} [R(\mathbf{r}, \mathbf{s})]_{p(\mathbf{r}|\mathbf{s})}) \left[\frac{1}{\sigma^2} (\mathbf{r}_i - f_i(\mathbf{W}^{in} \mathbf{s})) f'_i(\mathbf{W}^{in} \mathbf{s}) \mathbf{s}_j \right] \\ &\quad \times p(\mathbf{r}|\mathbf{s}; \mathbf{W}^{in}) p(\mathbf{s}) d\mathbf{r} d\mathbf{s} \end{aligned} \quad (\text{C.14})$$

$$= \int \frac{1}{\sigma^2} \text{Cov}(R(\mathbf{r}, \mathbf{s}), \mathbf{r}_i) [f'_i(\mathbf{W}^{in} \mathbf{s}) \mathbf{s}_j] p(\mathbf{s}) d\mathbf{s}, \quad (\text{C.15})$$

where $\text{Cov}(R(\mathbf{r}, \mathbf{s}), \mathbf{r}_i) = \int (R - \mathbb{E} [R]_{p(\mathbf{r}|\mathbf{s})})(\mathbf{r}_i - \mathbb{E} [\mathbf{r}_i]_{p(\mathbf{r}|\mathbf{s})}) p(\mathbf{r}|\mathbf{s}) d\mathbf{r}$ is the stimulus-conditioned covariance between network firing rates and reward.

The sample-based parameter update is therefore using the fluctuations in neural activity to compute this covariance.

C.5 Biological Plausibility Assessment. Now that we have derived REINFORCE, we can examine its qualities as a normative plasticity theory. First, we ask: Is this algorithm “local” (see section 2.2)? The gradient for a particular synapse, $\frac{\partial \mathcal{O}(\mathbf{W}^{in})}{\partial \mathbf{W}_{ij}^{in}}$, can be approximated with samples in an environment with stimuli \mathbf{s} , firing rates \mathbf{r} , and rewards $R(\mathbf{r}, \mathbf{s})$ by $R(\mathbf{r}, \mathbf{s}) \left[\frac{1}{\sigma^2} (\mathbf{r}_i - f_i(\mathbf{W}^{in}\mathbf{s})) f'_i(\mathbf{W}^{in}\mathbf{s}) \mathbf{s}_j \right]$. To decide whether this could be a plasticity rule implemented (or, more realistically, approximated) by a biological system, we need to think about what pieces of information a synapse would have to have available.

First, the synapse needs \mathbf{s}_j , which amounts to just the presynaptic input, a common feature of any Hebbian synaptic plasticity rule. Second, the synapse needs $\frac{1}{\sigma^2} (\mathbf{r}_i - f_i(\mathbf{W}^{in}\mathbf{s})) f'_i(\mathbf{W}^{in}\mathbf{s})$. $\frac{1}{\sigma^2}$ is a constant, and so can be absorbed into the learning rate. \mathbf{r}_i is the postsynaptic firing rate, which is also a common feature of any Hebbian plasticity rule. $(\mathbf{W}^{in}\mathbf{s})_i$ is the current injected into the postsynaptic neuron, and $f_i(\cdot)$ and $f'_i(\cdot)$ are both monotonic functions of this current, so it is quite conceivable that these values could be approximated by a biochemical process. Third, every synapse needs access to the scalar reward value received on a given trial, $R(\mathbf{r}, \mathbf{s})$. This is the most “nonlocal” information involved in the parameter update; however, there exist many theories about how neuromodulatory systems in the brain can deliver information about reward diffusely to many synapses and induce plasticity (see section 2.2). To achieve this locality, we have implicitly assumed that we are performing gradient descent with respect to a Euclidean metric (Surace et al., 2020); using different metrics corresponds to premultiplying the full weight update vector $\Delta \mathbf{W}$ by a positive-definite matrix. The locality results discussed here hold if this positive definite matrix is diagonal, but otherwise nonlocal interactions may be introduced.

We have already demonstrated that REINFORCE is able to perform approximate gradient descent for reinforcement learning objective functions. This in itself makes the algorithm very promising as a normative plasticity model (see section 2.1). Its chief advantage is that it does not require detailed knowledge of the reward function $R(\mathbf{r}, \mathbf{s})$ (i.e., how to differentiate it), which means that an animal could simply receive a reward from its environment and relay that reward signal diffusely to its synapses. However, this also restricts the types of objectives that could plausibly be learned by a neural system. Unsupervised learning objectives like the ELBO require detailed knowledge of every neural activity of every neuron in the circuit in order to be calculable, and there is no evidence for downstream neural circuits that perform such calculations. Therefore, even though in principle REINFORCE can be used to train a neural network on any objective, explicit reinforcement is much more plausible than other alternatives.

We have only provided a derivation for a single-layer, rate-based neural network with additive gaussian noise, but REINFORCE extends quite readily to multilayer (Williams, 1992), spiking (Frémaux et al., 2013), and recurrent networks (Miconi, 2017) without any loss of locality. This indicates that the algorithm is both architecture-general (see section 2.3) and can handle temporal environmental structure (see section 2.4). Further, because a weight update can be calculated in a single trial, animals could use it to learn online (see section 2.5). The biggest point of failure for REINFORCE is that it scales poorly with high complexity in stimuli or task, large numbers of neurons, or prolonged delays in receipt of reward (Werfel et al., 2003; Fiete, 2004; Bredenberg et al., 2021). The greater the number of neurons that contribute to reward and the higher the complexity of the reward function, the harder it becomes to estimate the correlation between a single neuron and reward, which is a prerequisite for the algorithm's function. Thus, though the algorithm is an unbiased estimator of the gradient, it can still be so variable an estimate as to be effectively useless in complex contexts. This suggests that if animals exploit the principles of REINFORCE to update synapses, it is likely an approach paired with other algorithms or hybridized in a way that allows for better scalability.

The last way to assess REINFORCE is on the basis of how it can be tested (see section 2.7). The simplest way to test this algorithm is by examining whether scalar reward-like signals (i.e., $R(r, s)$) have a multiplicative effect on local plasticity in a circuit. At a single-neuron level, this corresponds to identifying neuromodulators that affect plasticity. At a feedback level, this corresponds to identifying neuromodulatory systems that project to the circuit in question and observing whether their stimulation or silencing improves or blocks circuit-level plasticity or behavioral learning performance, respectively. These steps do not identify REINFORCE as the only possibility, but they narrow down the field of possibilities considerably, removing all candidate algorithms that either do not require any feedback or require more detailed feedback signals (see Figure 3a).

Acknowledgments

We thank Blake Richards, Eero Simoncelli, Owen Marschall, Benjamin Lyo, Elliott Capek, Olivier Codol, and Yuhe Fan for their helpful feedback on this review. C.S. is supported by NIMH Award 1R01MH125571-01, NIH Award R01NS127122, by the National Science Foundation under NSF Award No. 1922658, and a Google faculty award.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(1), 147–169.

Aitchison, L., Jegminat, J., Menendez, J. A., Pfister, J.-P., Pouget, A., & Latham, P. E. (2021). Synaptic plasticity as Bayesian inference. *Nature Neuroscience*, 24(4), 565–571. 10.1038/s41593-021-00809-5

Akrout, M., Wilson, C., Humphreys, P. C., Lillicrap, T., & Tweed, D. (2019). *Using weight mirrors to improve feedback alignment*. arXiv:1904.05391.

Alemi, A., Machens, C., Deneve, S., & Slotine, J.-J. (2018). Learning nonlinear dynamics in efficient, balanced spiking networks using local plasticity rules. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Arjona-Medina, J. A., Gillhofer, M., Widrich, M., Unterthiner, T., Brandstetter, J., & Hochreiter, S. (2019). RUDDER: Return decomposition for delayed rewards. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems*, 32. Curran.

Atick, J. J., & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation*, 2(3), 308–320. 10.1162/neco.1990.2.3.308

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3), 183. 10.1037/h0054663

Bartunov, S., Santoro, A., Richards, B., Marrs, L., Hinton, G. E., & Lillicrap, T. (2018). Assessing the scalability of biologically-motivated deep learning algorithms and architectures. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*, 31. MIT Press.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695–711. 10.1016/j.neuron.2012.10.038

Bear, M. F., & Singer, W. (1986). Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature*, 320(6058), 172–176. 10.1038/320172a0

Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., & Maass, W. (2020). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 11(1), 1–15. 10.1038/s41467-020-17236-y

Bellec, G., Salaj, D., Subramoney, A., Legenstein, R., & Maass, W. (2018). Long short-term memory and learning-to-learn in networks of spiking neurons. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*, 31. Curran.

Bengio, Y. (2014). *How auto-encoders could provide credit assignment in deep networks via target propagation*. arXiv:1407.7906.

Benna, M. K., & Fusi, S. (2016). Computational principles of synaptic memory consolidation. *Nature Neuroscience*, 19(12), 1697–1706. 10.1038/nn.4401

Bi, G.-q., & Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24), 10464–10472. 10.1523/JNEUROSCI.18-24-10464.1998

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2(1), 32–48. 10.1523/JNEUROSCI.02-01-00032.1982

Bittner, K. C., Grienberger, C., Vaidya, S. P., Milstein, A. D., Macklin, J. J., Suh, J., Tonegawa, S., & Magee, J. C. (2015). Conjunctive input processing drives feature

selectivity in hippocampal CA1 neurons. *Nature Neuroscience*, 18(8), 1133–1142. 10.1038/nn.4062

Bittner, K. C., Milstein, A. D., Grienberger, C., Romani, S., & Magee, J. C. (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science*, 357(6355), 1033–1036. 10.1126/science.aan3846

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature*, 361(6407), 31–39. 10.1038/361031a0

Bohte, S. M., Kok, J. N., & La Poutre, H. (2002). Error-backpropagation in temporally encoded networks of spiking neurons. *Neurocomputing*, 48(1–4), 17–37. 10.1016/S0925-2312(01)00658-0

Bönstrup, M., Iturrate, I., Thompson, R., Cruciani, G., Censor, N., & Cohen, L. G. (2019). A rapid form of offline consolidation in skill learning. *Current Biology*, 29(8), 1346–1351.

Bouvier, G., Aljadeff, J., Clopath, C., Bimbard, C., Ranft, J., Blot, A., . . . Barbour, B. (2018). Cerebellar learning using perturbations. *eLife*, 7, e31599. 10.7554/eLife.31599

Bredenberg, C., Lyo, B. S. H., Simoncelli, E. P., & Savin, C. (2021). Impression learning: Online representation learning with synaptic plasticity. In M. A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. Wortman Vaughan (Eds.), *Advances in neural information processing systems*, 34.

Bredenberg, C., Simoncelli, E., & Savin, C. (2020). Learning efficient task-dependent representations with synaptic plasticity. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems*, 33. Curran.

Bredenberg, C., Williams, E., Savin, C., Richards, B. A., & Lajoie, G. (2023). Formalizing locality for normative synaptic plasticity models. In *Proceedings of the Thirty-Seventh Conference on Neural Information Processing Systems*.

Brendel, W., Bourdoukan, R., Vertechi, P., Machens, C. K., & Denéve, S. (2020). Learning to represent signals spike by spike. *PLOS Computational Biology*, 16(3), e1007692. 10.1371/journal.pcbi.1007692

Brody, C. D., Hernández, A., Zainos, A., & Romo, R. (2003). Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cerebral Cortex*, 13(11), 1196–1207. 10.1093/cercor/bhg100

Calabresi, P., Picconi, B., Tozzi, A., & Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends in Neurosciences*, 30(5), 211–219. 10.1016/j.tins.2007.03.001

Clopath, C., Ziegler, L., Vasilaki, E., Büsing, L., & Gerstner, W. (2008). Tag-trigger-consolidation: A model of early and late long-term-potentiation and depression. *PLOS Computational Biology*, 4(12), e1000248. 10.1371/journal.pcbi.1000248

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10(9), 910–923. 10.1093/cercor/10.9.910

Cooper, L. N. (2004). *Theory of cortical plasticity*. World Scientific.

Cornford, J., Kalajdzievski, D., Leite, M., Lamarquette, A., Kullmann, D. M., & Richards, B. (2021). *Learning to live with Dale's principle: ANNs with separate excitatory and inhibitory units*. bioRxiv.

Dayan, P., & Hinton, G. E. (1996). Varieties of Helmholtz machine. *Neural Networks*, 9(8), 1385–1403. 10.1016/S0893-6080(96)00009-3

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, 7(5), 889–904. 10.1162/neco.1995.7.5.889

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, 22(6), 1075–1081. 10.1016/j.conb.2012.08.003

Ernoult, M., Grollier, J., Querlioz, D., Bengio, Y., & Scellier, B. (2020). *Equilibrium propagation with continual weight updates*. arXiv:2005.04168.

Eschenko, O., Ramadan, W., Mölle, M., Born, J., & Sara, S. J. (2008). Sustained increase in hippocampal sharp-wave ripple activity during slow-wave sleep after learning. *Learning and Memory*, 15(4), 222–228. 10.1101/lm.726008

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4), 292–303. 10.1038/nrn2258

Fetz, E. E. (2007). Volitional control of neural activity: Implications for brain-computer interfaces. *Journal of Physiology*, 579(3), 571–579. 10.1113/jphysiol.2006.127142

Feulner, B., & Clopath, C. (2021). Neural manifold under plasticity in a goal driven learning behaviour. *PLOS Computational Biology*, 17(2), e1008621. 10.1371/journal.pcbi.1008621

Fiete, I. R. (2004). *Learning and coding in biological neural networks*. Harvard University Press.

Fiete, I. R., Fee, M. S., & Seung, H. S. (2007). Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *Journal of Neurophysiology*, 98(4), 2038–2057. 10.1152/jn.01311.2006

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences*, 14(3), 119–130. 10.1016/j.tics.2010.01.003

Francioni, V., Tang, V. D., Brown, N. J., Toloza, E. H., & Harnett, M. (2023). *Vectorized instructive signals in cortical dendrites during a brain-computer interface task*. bioRxiv.

Frémaux, N., & Gerstner, W. (2016). Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9, 85.

Frémaux, N., Sprekeler, H., & Gerstner, W. (2013). Reinforcement learning using a continuous time actor-critic framework with spiking neurons. *PLOS Computational Biology*, 9(4), e1003024.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. 10.1038/nrn2787

Friston, K., & Kiebel, S. (2009). Cortical circuits for perceptual inference. *Neural Networks*, 22(8), 1093–1104. 10.1016/j.neunet.2009.07.023

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6(11), 1216–1223. 10.1038/nn1141

Froemke, R. C., Carcea, I., Barker, A. J., Yuan, K., Seybold, B. A., Martins, A. R. O., ... Schreiner, C. E. (2013). Long-term modification of cortical synapses improves sensory perception. *Nature Neuroscience*, 16(1), 79–88. 10.1038/nn.3274

Froemke, R. C., Merzenich, M. M., & Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, 450(7168), 425–429. 10.1038/nature06289

Froemke, R. C., Poo, M.-m., & Dan, Y. (2005). Spike-timing-dependent synaptic plasticity depends on dendritic location. *Nature*, 434(7030), 221–225. 10.1038/nature03366

Fusi, S., Drew, P. J., & Abbott, L. F. (2005). Cascade models of synaptically stored memories. *Neuron*, 45(4), 599–611. 10.1016/j.neuron.2005.02.001

Ganguli, S., Huh, D., & Sompolinsky, H. (2008). Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, 105(48), 18970–18975. 10.1073/pnas.0804451105

Gao, Z., Van Beugen, B. J., & De Zeeuw, C. I. (2012). Distributed synergistic plasticity and cerebellar learning. *Nature Reviews Neuroscience*, 13(9), 619–635. 10.1038/nrn3312

Gerstner, W., & Kistler, W. M. (2002). Mathematical formulations of Hebbian learning. *Biological Cybernetics*, 87(5), 404–415. 10.1007/s00422-002-0353-y

Gerstner, W., Lehmann, M., Liakoni, V., Corneil, D., & Brea, J. (2018). Eligibility traces and plasticity on behavioral time scales: Experimental support of neo-Hebbian three-factor learning rules. *Frontiers in Neural Circuits*, 12, 53. 10.3389/fncir.2018.00053

Gillon, C. J., Pina, J. E., Lecoq, J. A., Ahmed, R., Billeh, Y., Caldejon, S., . . . Zylberberg (2021). *Learning from unexpected events in the neocortical microcircuit*. bioRxiv.

Gilra, A., & Gerstner, W. (2017). Predicting non-linear dynamics by stable local learning in a recurrent spiking neural network. *eLife*, 6, e28295. 10.7554/eLife.28295

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., & Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, 12(10), 1222–1223. 10.1038/nn.2384

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535–574. 10.1146/annurev.neuro.29.051605.113038

Golkar, S., Tesileanu, T., Bahroun, Y., Sengupta, A. M., & Chklovskii, D. B. (2022). *Constrained predictive coding as a biologically plausible model of the cortical hierarchy*. arXiv:2210.15752.

Graupner, M., & Brunel, N. (2010). Mechanisms of induction and maintenance of spike-timing dependent plasticity in biophysical synapse models. *Frontiers in Computational Neuroscience*, 4, 136. 10.3389/fncom.2010.00136

Greedy, W., Zhu, H. W., Pemberton, J., Mellor, J., & Ponte Costa, R. (2022). Single-phase deep learning in cortico-cortical networks. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems*, 35 (pp. 24213–24225). Curran.

Grienerger, C., & Magee, J. C. (2022). Entorhinal cortex directs learning-related changes in CA1 representations. *Nature*, 611(7936), 1–9. 10.1038/s41586-022-05378-6

Gu, Q., & Singer, W. (1995). Involvement of serotonin in developmental plasticity of kitten visual cortex. *European Journal of Neuroscience*, 7(6), 1146–1153. 10.1111/j.1460-9568.1995.tb01104.x

Guerguiev, J., Lillicrap, T. P., & Richards, B. A. (2017). Towards deep learning with segregated dendrites. *eLife*, 6, e22901. 10.7554/eLife.22901

Gulati, T., Guo, L., Ramanathan, D. S., Bodepudi, A., & Ganguly, K. (2017). Neural reactivations during sleep determine network credit assignment. *Nature Neuroscience*, 20(9), 1277–1284. 10.1038/nn.4601

Guo, W., Robert, B., & Polley, D. B. (2019). The cholinergic basal forebrain links auditory stimuli with delayed reinforcement to support learning. *Neuron*, 103(6), 1164–1177. 10.1016/j.neuron.2019.06.024

Halvagal, M. S., & Zenke, F. (2023). The combination of Hebbian and predictive plasticity learns invariant object representations in deep sensory networks. *Nature Neuroscience*, 26(11), 1–10. 10.1038/s41593-023-01460-y

Hangya, B., Ranade, S. P., Lorenc, M., & Kepcs, A. (2015). Central cholinergic neurons are rapidly recruited by reinforcement feedback. *Cell*, 162(5), 1155–1168.

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, 58(4), 1021–1034. 10.2307/1130543

Hennequin, G., Vogels, T. P., & Gerstner, W. (2012). Non-normal amplification in random balanced neuronal networks. *Physical Review E*, 86(1), 011909. 10.1103/PhysRevE.86.011909

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The “wake-sleep” algorithm for unsupervised neural networks. *Science*, 268(5214), 1158–1161. 10.1126/science.7761831

Hong, G., & Lieber, C. M. (2019). Novel electrode technologies for neural recordings. *Nature Reviews Neuroscience*, 20(6), 330–345. 10.1038/s41583-019-0140-6

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554–2558. 10.1073/pnas.79.8.2554

Huh, D., & Sejnowski, T. J. (2018). Gradient descent for spiking neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*, 31. Curran.

Humphreys, P. C., Daie, K., Svoboda, K., Botvinick, M., & Lillicrap, T. P. (2022). *BCI learning phenomena can be explained by gradient-based optimization*. bioRxiv.

Hung, C.-C., Lillicrap, T., Abramson, J., Wu, Y., Mirza, M., Carnevale, F., Ahuja, A., & Wayne, G. (2019). Optimizing agent behavior over long time scales by transporting value. *Nature Communications*, 10(1), 1–12. 10.1038/s41467-019-13073-w

Ikeda, S., Amari, S. I., & Nakahara, H. (1998). Convergence of the wake-sleep algorithm. In M. Kearns, S. Solla, & D. Cohn (Eds.), *Advances in neural information processing systems*, 11. MIT Press.

Illing, B., Ventura, J., Bellec, G., & Gerstner, W. (2021). Local plasticity rules can learn deep representations using self-supervised contrastive predictions. In M. A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. Wortman Vaughan (Eds.), *Advances in neural information processing systems*, 34. Curran.

Intrator, N., & Cooper, L. N. (1992). Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5(1), 3–17. 10.1016/S0893-6080(05)80003-6

Isomura, T., & Toyoizumi, T. (2016). A local learning rule for independent component analysis. *Scientific Reports*, 6(1), 28073. 10.1038/srep28073

Izhikevich, E. M., & Desai, N. S. (2003). Relating STDP to BCM. *Neural Computation*, 15(7), 1511–1523. 10.1162/089976603321891783

Jabri, M., & Flower, B. (1992). Weight perturbation: An optimal architecture and learning technique for analog VLSI feedforward and recurrent multi-layer networks. *IEEE Transactions on Neural Networks*, 3(1), 154–157. 10.1109/72.105429

Jegminat, J., Surace, S. C., & Pfister, J.-P. (2022). Learning as filtering: Implications for spike-based plasticity. *PLOS Computational Biology*, 18(2), e1009721.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2), 99–134. 10.1016/S0004-3702(98)00023-X

Kappel, D., Habenschuss, S., Legenstein, R., & Maass, W. (2015). Network plasticity as Bayesian inference. *PLOS Computational Biology*, 11(11), e1004485. 10.1371/journal.pcbi.1004485

Kappel, D., Nessler, B., & Maass, W. (2014). STDP installs in winner-take-all circuits an online approximation to hidden Markov model learning. *PLOS Computational Biology*, 10(3), e1003511. 10.1371/journal.pcbi.1003511

Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, 98(3), 630–644. 10.1016/j.neuron.2018.03.044

Laborieux, A., Ernoult, M., Scellier, B., Bengio, Y., Grollier, J., & Querlioz, D. (2021). Scaling equilibrium propagation to deep ConvNets by drastically reducing its gradient estimator bias. *Frontiers in Neuroscience*, 15, 129. 10.3389/fnins.2021.633674

Larkum, M. E., Zhu, J. J., & Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398(6725), 338–341. 10.1038/18686

Lee, D.-H., Zhang, S., Fischer, A., & Bengio, Y. (2015). Difference target propagation. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 498–515).

Lee, J. H., Delbrück, T., & Pfeiffer, M. (2016). Training deep spiking neural networks using backpropagation. *Frontiers in Neuroscience*, 10, 508.

Legenstein, R., Chase, S. M., Schwartz, A. B., & Maass, W. (2010). A reward-modulated Hebbian learning rule can explain experimentally observed network reorganization in a brain control task. *Journal of Neuroscience*, 30(25), 8400–8410. 10.1523/JNEUROSCI.4284-09.2010

Lengyel, M., Kwag, J., Paulsen, O., & Dayan, P. (2005). Matching storage and recall: Hippocampal spike timing-dependent plasticity and phase response curves. *Nature Neuroscience*, 8(12), 1677–1683. 10.1038/nn1561

Letzkus, J. J., Kampa, B. M., & Stuart, G. J. (2006). Learning rules for spike timing-dependent plasticity depend on dendritic synapse location. *Journal of Neuroscience*, 26(41), 10420–10429. 10.1523/JNEUROSCI.2650-06.2006

Levelt, C. N., & Hübener, M. (2012). Critical-period plasticity in the visual cortex. *Annual Review of Neuroscience*, 35, 309–330. 10.1146/annurev-neuro-061010-113813

Levenstein, D., Alvarez, V. A., Amarasingham, A., Azab, H., Gerkin, R. C., Hasenstaub, A., . . . Redish, A. D. (2020). *On the role of theory and modeling in neuroscience*. arXiv:2003.13825.

Lillicrap, T. P., Cownden, D., Tweed, D. B., & Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7(1), 1–10. 10.1038/ncomms13276

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6), 335–346. 10.1038/s41583-020-0277-3

Liu, Y., Mattar, M. G., Behrens, T. E., Daw, N. D., & Dolan, R. J. (2021). Experience replay is associated with efficient nonlocal learning. *Science*, 372(6544), eabf1357.

Magee, J. C., & Grienberger, C. (2020). Synaptic plasticity forms and functions. *Annual Review of Neuroscience*, 43, 95–117. 10.1146/annurev-neuro-090919-022842

Manchev, N., & Spratling, M. W. (2020). Target propagation in recurrent neural networks. *Journal of Machine Learning Research*, 21(7), 1–33.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474), 78–84. 10.1038/nature12742

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APS and EPSPs. *Science*, 275(5297), 213–215. 10.1126/science.275.5297.213

Marlin, B. J., Mitre, M., D'Amour, J. A., Chao, M. V., & Froemke, R. C. (2015). Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature*, 520(7548), 499–504. 10.1038/nature14402

Marschall, O., Cho, K., & Savin, C. (2020). A unified framework of online learning algorithms for training recurrent neural networks. *Journal of Machine Learning Research*, 21(1), 5320–5353.

Marschall, O., & Savin, C. (2023). *Probing learning through the lens of changes in circuit dynamics*. bioRxiv.

Martin, S. J., Grimwood, P. D., & Morris, R. G. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience*, 23(1), 649–711. 10.1146/annurev.neuro.23.1.649

Martins, A. R. O., & Froemke, R. C. (2015). Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. *Nature Neuroscience*, 18(10), 1483–1492. 10.1038/nn.4090

Meulemans, A., Carzaniga, F. S., Suykens, J. A., Sacramento, J., & Grewe, B. F. (2020). *A theoretical framework for target propagation*. arXiv:2006.14331.

Meulemans, A., Zucchetti, N., Kobayashi, S., Von Oswald, J., & Sacramento, J. (2022). The least-control principle for local learning at equilibrium. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems*, 35 (pp. 33603–33617). Curran.

Miconi, T. (2017). Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks. *eLife*, 6, e20899. 10.7554/eLife.20899

Mineault, P., Bakhtiari, S., Richards, B., & Pack, C. (2021). Your head is there to move you around: Goal-driven models of the primate dorsal pathway. In M. A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. Wortman Vaughan (Eds.), *Advances in neural information processing systems*, 34. Curran.

Mnih, A., & Gregor, K. (2014). Neural variational inference and learning in belief networks. In *Proceedings of the International Conference on Machine Learning* (pp. 1791–1799).

Murphy, T. H., & Corbett, D. (2009). Plasticity during stroke recovery: From synapse to behaviour. *Nature Reviews Neuroscience*, 10(12), 861–872. 10.1038/nrn2735

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

Nayebi, A., Attinger, A., Campbell, M., Hardcastle, K., Low, I., Mallory, C., . . . Yamins, D. (2021). Explaining heterogeneity in medial entorhinal cortex with task-driven neural networks. In M. A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. Wortman Vaughan (Eds.), *Advances in neural information processing systems*, 34. Curran.

Nayebi, A., Srivastava, S., Ganguli, S., & Yamins, D. L. (2020). *Identifying learning rules from neural network observables*. arXiv:2010.11765.

Neftci, E. O., Mostafa, H., & Zenke, F. (2019). Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6), 51–63. 10.1109/SP.2019.2931595

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139–154. 10.1016/j.jmp.2008.12.005

Obeid, D., Ramambason, H., & Pehlevan, C. (2019). *Structured and deep similarity matching via structured and deep Hebbian networks*. arXiv:1910.04958.

O'Donohue, T. L., Millington, W. R., Handelmann, G. E., Contreras, P. C., & Chronwall, B. M. (1985). On the 50th anniversary of Dale's law: Multiple neurotransmitter neurons. *Trends in Pharmacological Sciences*, 6, 305–308.

Ohl, F. W., & Scheich, H. (2005). Learning-induced plasticity in animal and human auditory cortex. *Current Opinion in Neurobiology*, 15(4), 470–477. 10.1016/j.conb.2005.07.002

Oja, E. (1982). Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3), 267–273. 10.1007/BF00275687

Oord, A. v. d., Li, Y., & Vinyals, O. (2018). *Representation learning with contrastive predictive coding*. arXiv:1807.03748.

Otani, S., Daniel, H., Roisin, M.-P., & Crepel, F. (2003). Dopaminergic modulation of long-term synaptic plasticity in rat prefrontal neurons. *Cerebral Cortex*, 13(11), 1251–1256. 10.1093/cercor/bhg092

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *Journal of Neuroscience*, 9(8), 2907–2918. 10.1523/JNEUROSCI.09-08-02907.1989

Payeur, A., Guerguiev, J., Zenke, F., Richards, B. A., & Naud, R. (2021). Burstdependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nature Neuroscience*, 24, 1010–1019. 10.1038/s41593-021-00857-x

Payeur, A., Orsborn, A. L., & Lajoie, G. (2023). *Neural manifolds and gradient-based adaptation in neural-interface tasks*. bioRxiv, 2023–03.

Pehlevan, C., Hu, T., & Chklovskii, D. B. (2015). A Hebbian/anti-Hebbian neural network for linear subspace learning: A derivation from multidimensional scaling of streaming data. *Neural Computation*, 27(7), 1461–1495. 10.1162/NECO_a_00745

Pehlevan, C., Sengupta, A. M., & Chklovskii, D. B. (2017). Why do similarity matching objectives lead to Hebbian/anti-Hebbian networks? *Neural Computation*, 30(1), 84–124. 10.1162/neco_a_01018

Pfister, J.-P., Toyoizumi, T., Barber, D., & Gerstner, W. (2006). Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Computation*, 18(6), 1318–1348. 10.1162/neco.2006.18.6.1318

Pozzi, I., Bohte, S., & Roelfsema, P. (2020). Attention-gated brain propagation: How the brain can implement reward-based error backpropagation. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems*, 33. Curran.

Rajasethupathy, P., Ferenczi, E., & Deisseroth, K. (2016). Targeting neural circuits. *Cell*, 165(3), 524–534.

Ranganath, R., Gerrish, S., & Blei, D. (2014). Black box variational inference. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics* (pp. 814–822).

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. 10.1038/4580

Raposo, D., Ritter, S., Santoro, A., Wayne, G., Weber, T., Botvinick, M., . . . Song, F. (2021). *Synthetic returns for long-term credit assignment*. arXiv:2102.12425.

Rasmusson, D. (2000). The role of acetylcholine in cortical synaptic plasticity. *Behavioural Brain Research*, 115(2), 205–218. 10.1016/S0166-4328(00)00259-X

Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, 15(4–6), 507–521. 10.1016/S0893-6080(02)00045-X

Richards, B. A., & Lillicrap, T. P. (2019). Dendritic solutions to the credit assignment problem. *Current Opinion in Neurobiology*, 54, 28–36. 10.1016/j.conb.2018.08.003

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., . . . Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, 22(11), 1761–1770. 10.1038/s41593-019-0520-2

Roelfsema, P. R., & Ooyen, A. v. (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural Computation*, 17(10), 2176–2214. 10.1162/0899766054615699

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386. 10.1037/h0042519

Roth, C., Kanitscheider, I., & Fiete, I. (2018). Kernel RNN learning (KeRNL). In *Proceedings of the International Conference on Learning Representations*.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation*. Technical report, La Jolla Institute for Cognitive Science, San Diego.

Sacramento, J., Costa, R. P., Bengio, Y., & Senn, W. (2017). *Dendritic error backpropagation in deep cortical microcircuits*. arXiv:1801.00062.

Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., . . . Batista, A. P. (2014). Neural constraints on learning. *Nature*, 512(7515), 423–426. 10.1038/nature13665

Savin, C., Dayan, P., & Lengyel, M. (2014). Optimal recall from bounded metaplastic synapses: Predicting functional adaptations in hippocampal area CA3. *PLOS Computational Biology*, 10(2), e1003489. 10.1371/journal.pcbi.1003489

Savin, C., Joshi, P., & Triesch, J. (2010). Independent component analysis in spiking neurons. *PLOS Computational Biology*, 6(4), e1000757. 10.1371/journal.pcbi.1000757

Scellier, B., & Bengio, Y. (2017). Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Frontiers in Computational Neuroscience*, 11, 24. 10.3389/fncom.2017.00024

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599. 10.1126/science.275.5306.1593

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4), 1916–1936. 10.1152/jn.2001.86.4.1916

Sheffield, M. E., Adoff, M. D., & Dombeck, D. A. (2017). Increased prevalence of calcium transients across the dendritic arbor during place field formation. *Neuron*, 96(2), 490–504. 10.1016/j.neuron.2017.09.029

Shinoe, T., Matsui, M., Taketo, M. M., & Manabe, T. (2005). Modulation of synaptic plasticity by physiological activation of M1 muscarinic acetylcholine receptors in the mouse hippocampus. *Journal of Neuroscience*, 25(48), 11194–11200. 10.1523/JNEUROSCI.2338-05.2005

Shrestha, S. B., & Orchard, G. (2018). SLAYER: Spike layer error reassignment in time. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*, 31. Curran.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1), 1193–1216. 10.1146/annurev.neuro.24.1.1193

Sjöström, P. J., & Häusser, M. (2006). A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons. *Neuron*, 51(2), 227–238.

Sohn, H., Narain, D., Meirhaeghe, N., & Jazayeri, M. (2019). Bayesian computation through cortical latent dynamics. *Neuron*, 103(5), 934–947. 10.1016/j.neuron.2019.06.012

Sompolinsky, H., & Kanter, I. (1986). Temporal association in asymmetric neural networks. *Physical Review Letters*, 57(22), 2861. 10.1103/PhysRevLett.57.2861

Spall, J. C. (1992). Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3), 332–341. 10.1109/9.119632

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437(7063), 1272–1278. 10.1038/nature04286

Surace, S. C., Pfister, J.-P., Gerstner, W., & Brea, J. (2020). On the choice of metric in gradient-based theories of brain function. *PLOS Computational Biology*, 16(4), e1007640. 10.1371/journal.pcbi.1007640

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Tishby, N., Pereira, F. C., & Bialek, W. (2000). *The information bottleneck method*. arXiv.physics.0004057.

Toyoizumi, T., Pfister, J.-P., Aihara, K., & Gerstner, W. (2005). Generalized Bienenstock–Cooper–Munro rule for spiking neurons that maximizes information transmission. *Proceedings of the National Academy of Sciences*, 102(14), 5239–5244. 10.1073/pnas.0500495102

Urbanczik, R., & Senn, W. (2014). Learning by the dendritic prediction of somatic spiking. *Neuron*, 81(3), 521–528. 10.1016/j.neuron.2013.11.030

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., & Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(12).

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., & Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, 334(6062), 1569–1573. 10.1126/science.1211095

Werbos, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. PhD diss., Harvard University.

Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, 78(10), 1550–1560. 10.1109/5.58337

Werfel, J., Xie, X., & Seung, H. S. (2003). Learning curves for stochastic gradient descent in linear feedforward networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems*, 29 (pp. 1197–1204). Curran.

Whittington, J. C., & Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a predictive coding network with local Hebbian synaptic plasticity. *Neural Computation*, 29(5), 1229–1262. 10.1162/NECO_a_00949

Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology*, 26(6), 1003–1017. 10.1152/jn.1963.26.6.1003

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3), 229–256.

Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1(2), 270–280. 10.1162/neco.1989.1.2.270

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4), 1314–1328. 10.1523/JNEUROSCI.3733-05.2006

Xiao, Z.-C., Lin, K. K., & Young, L.-S. (2021). A data-informed mean-field approach to mapping of cortical parameter landscapes. *PLOS Computational Biology*, 17(12), e1009718. 10.1371/journal.pcbi.1009718

Xie, X., & Seung, H. S. (2003). Equivalence of backpropagation and contrastive Hebbian learning in a layered network. *Neural Computation*, 15(2), 441–454. 10.1162/089976603762552988

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3), 356–365. 10.1038/nn.4244

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23), 8619–8624. 10.1073/pnas.1403112111

Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications*, 10(1), 3770. 10.1038/s41467-019-11786-6

Zhang, K., Yang, Z., & Başar, T. (2021). Multi-agent reinforcement learning: A selective overview of theories and algorithms. In K. G. Vamvoudakis, Y. Wan, F. L.

Lewis, & D. Cansever (Eds.), *Handbook of reinforcement learning and control* (pp. 321–384). Springer.

Zigmond, M. J., Abercrombie, E. D., Berger, T. W., Grace, A. A., & Stricker, E. M. (1990). Compensations after lesions of central dopaminergic neurons: Some clinical and basic implications. *Trends in Neurosciences*, 13(7), 290–296. 10.1016/0166-2236(90)90112-N

Received August 9, 2023; accepted February 6, 2024.