

Chapter 1

Introduction

1.1 Motivation

The outstanding learning capabilities of the human brain have been found to be elusive and as yet impossible to fully explain or replicate in silicio. While in recent years the power of classical machine learning solutions has improved even beyond human capabilities for some tasks, their underlying algorithms cannot serve as a model of human cognition. Some reasons why brains and machines appear irreconcilable relate to questions about network structure and neuron models. Yet more pressingly, almost all the most powerful artificial neural networks are trained with the Backpropagation of errors algorithm, which has long been considered to be impossible for neurons to implement. Hence, Neuroscience has dismissed this algorithm in an almost dogmatic way for many years after its development, stating that the brain must employ a different mechanism to learn.

Yet in recent years, there has been a resurgence of research by neuroscientists towards reconciling biological and artificial neural networks in spite of these concerns. This led to a number of experimental results indicating that brains might be capable of performing something very similar to Backpropagation after all. Furthermore, despite rigorous efforts, no unifying alternative to this learning principle was found which performs well enough to account for the brain's unmatched capabilities.

Hence, there now exists a vibrant community developing alternative ways to implement this algorithm - or some approximation of it. These novel approaches are capable of replicating an increasing number of properties of biological brains. Nevertheless, many issues remain unsolved, and a lot of neuronal features remain unaccounted for in brain models that are capable of any kind of learning. It is this open problem, to which I want to dedicate my efforts in this thesis. After reviewing the existing literature, I have selected a promising model of learning in cortical circuits. This model uses multi-compartment neuron models and local plasticity rules to implement a variant of Backpropagation. In this project, I will investigate and attempt to further improve its concordance with data on the human neocortex. I will use

the approach of computationally modelling the model while progressively adding biological features, attempting to retain learning performance in the process.

1.2 The Backpropagation of errors algorithm

The Backpropagation of errors algorithm (*Backprop*) (Schmidhuber, 2014) is the workhorse of modern machine learning and is able to outperform humans on a growing number of tasks (LeCun et al., 2015). Particularly for training deep neural networks it has remained popular and largely unchanged since its initial development. Its learning potential stems from its unique capability to attribute errors in the output of a network to activations of specific neurons and connections within its hidden layers. This property also forms the basis of the algorithm’s name; After an initial forward pass to form a prediction about the nature of a given input, a separate backward pass propagates the arising error through all layers in reverse order. During this second network traversal, local error gradients dictate to what extent a given weight needs to be altered so that the next presentation of the same sample would elicit a lower error in the output layer. It has been argued that through this mechanism, Backprop solves the *credit assignment problem* - i.e. the question to what degree a parameter contributes to an error signal - optimally (Lillicrap et al., 2020). With this critical information in hand, computing parameter changes that decrease error becomes almost trivial. As biological neural networks are likewise subject to the credit assignment problem, finding a general solution to it promises to be invaluable to neuroscience. For a long time Backprop was believed to be unsuitable for networks of biological neurons for several reasons.

1.3 Concerns over biological plausibility

While Backprop continues to prove exceptionally useful in conventional machine learning systems, it is viewed critically by many neuroscientists. For one, it relies on a slow adaptation of synaptic weights, and therefore requires a large amount of examples to learn rather simple input-output mappings. In this particular way, its performance is far inferior to the powerful one-shot learning exhibited by humans (Brea and Gerstner, 2016). Yet more importantly, no plausible mechanisms have yet been found by which biological neural networks could implement the algorithm. In fact, Backprop as a way by which brains may learn has been dismissed entirely by much of the neuroscience community for decades (Grossberg, 1987; Crick, 1989; Mazzoni et al., 1991; O’Reilly, 1996). This dismissal is often focussed on three mechanisms that are instrumental for the algorithm (Whittington and Bogacz, 2019; Bengio et al., 2015; Liao et al., 2016):

1.3.1 Local error representation

Neuron-specific errors in Backprop are computed and propagated by a mechanism that is completely detached from the network itself, which requires access to the entirety of the network state. In order to compute the weight changes for a given layer, the algorithm takes as an input the activation and synaptic weights of all downstream neurons. In contrast, plasticity in biological neurons is largely considered to be primarily dependent on factors that are local to the synapse (Abbott and Nelson, 2000; Magee and Grienberger, 2020; Urbanczik and Senn, 2014). While neuromodulators are known to influence synaptic plasticity, their dispersion is too wide to communicate neuron-specific errors. Thus, biologically plausible Backprop would require a method for encoding errors locally, i.e. close to the neurons to which they relate. This has been perhaps the strongest criticism of Backprop in the brain, as many questions regarding mechanisms for both computing and storing these errors remain unanswered as yet.

1.3.2 The weight transport problem

During the weight update stage of Backprop, errors are transmitted between layers with the same weights that are used in the forward pass. In other words, the magnitude of a neuron-specific error that is back-propagated through a given connection should be proportional to its impact on output loss during the forward pass. To replicate this, a neuronal network implementing Backprop would require feedback connections that mirror both the precise connectivity and synaptic weights of the forward connections. Bidirectional connections that could theoretically back-propagate errors are common in the cortex, yet it is unclear by which mechanism pairs of synapses would be able to align. This issue becomes particularly apparent when considering long-range pyramidal projections. In these, the feedforward and feedback synapses which need to be aligned would potentially be separated by a considerable distance.

1.3.3 Neuron models

Finally, the types of artificial neurons typically used in Backprop transmit a continuous scalar activation at all times, instead of discrete spikes. In theory, these activations correspond to the firing rate of a spiking neuron, giving this class of models the title *rate neurons*. Yet handling spike based communication requires more sophisticated neuron models than are typically employed in Backprop networks. Additionally, plasticity rules for rate neurons do not necessarily have an easily derived counterpart for spiking neurons. A notable example for this issue is Backprop itself; The local error gradient of a neuron is not trivial to compute for spiking neural networks (SNN), as a spiketrain has no natural derivative. Furthermore, a given neuron's activation in classical Backprop is computed from a simple weighted sum of all inputs. This fails to capture the complex nonlinearities of dendritic integration that are fundamental to cortical neurons (cf. Section 1.4.1). Finally, these abstract neurons - at least in classical Backprop - have no persistence through time. Thus, their activation is dictated strictly by

instantaneous presynaptic activity, in contrast to the leaky membrane dynamics exhibited by biological neurons.

1.4 Overcoming biological implausibility

Backprop has remained the gold standard against which most attempts at modelling learning in the brain eventually are compared. Also, despite its apparent biological implausibility, it does share some notable parallels to learning in the brain. Artificial neural networks (ANN) trained with Backprop have been shown to develop similar representations to those found in brain areas responsible for comparable tasks (Yamins and DiCarlo, 2016; Whittington et al., 2018; Khaligh-Razavi and Kriegeskorte, 2014; Kumbhani et al., 2016). Thus, numerous attempts have been made to define more biologically plausible learning rules which approximate Backprop to some degree. A full review of the available literature would be out of scope for this thesis, so only a few examples will be discussed in this section.

One approach to solve the issues around local error representations is, to drive synaptic plasticity through a global error signal (Potjans et al., 2011; Mozafari et al., 2018; Sutton and Barto, 2018). The appeal of this solution is that such signalling could be plausibly performed by neuromodulators like dopamine (Mazzoni et al., 1991; Seung, 2003; Izhikevich, 2007). These types solutions do not approximate Backprop, but instead lead to a kind of reinforcement learning. While some consider this the most plausible way for brains to learn (Sutton and Barto, 2018), performance of global error/reward signalling stays far behind that of the credit assignment performed by Backprop. Additionally, this class of algorithms requires even more examples of a training dataset, and was shown to scale poorly with network size (Werfel et al., 2003).

Two prominent classes of Backprop approximations have been developed, which are capable of locally representing errors. These algorithms encode errors in either activation changes over time or local membrane potentials. They will be discussed further in Section 1.6.

The weight transport problem was successfully addressed by a mechanism called *Feedback Alignment* (FA) (Lillicrap et al., 2014). This seminal paper shows that Backprop can still learn successfully when feedback weights are random. In addition to learning to represent an input-output mapping in forward weights, the network is trained to extract useful information from randomly weighted instructive pathways. The authors call this process *learning to learn*, and show that performance is even superior to classical Backprop for some tasks. This mechanism was further expanded to show that the principles of FA perform very well when biologically plausible plasticity rules are employed (Liao et al., 2016; Zenke and Ganguli, 2018). Another popular line of thought is - instead of computing local errors - to compute optimal activations for hidden layer neurons using autoencoders (Bengio, 2014; Lee et al., 2015; Ahmad et al., 2020). Approaches derived from this do not suffer from the weight transport problem,

and by design does not require local error representations. While these solutions (summed as *Target propagation* algorithms) solve the weight transport problem, they fall far behind traditional Backprop on more complex benchmark datasets like *CIFAR* and *ImageNet* (Bartunov et al., 2018).

Numerous approaches for implementing Backprop with more plausible neuron models exist, most of which employ variants of the *Leaky Integrate-and-fire* (LIF) neuron (Sporea and Grüning, 2013; Lee et al., 2016; Bengio et al., 2017; Lee et al., 2020). The aforementioned issue of computing the derivative over spiketrains has been solved in several ways, with the most prominent variant perhaps being *SuperSpike* (Zenke and Ganguli, 2018). One might therefore view this as the weakest criticism aimed at Backprop. Yet none of the employed neuron models come close to portraying the intricacies of biological neurons, and thus fail to provide explanations for their complexity. One aspect of this will be discussed in the upcoming section.

All of these studies successfully solve one or more concerns of biological plausibility, while still approximating Backprop to some degree. Yet none of them are able to solve all three simultaneously, and some of them introduce novel mechanisms that are themselves biologically questionable. It further appears that in all but a few cases, an increase in biological plausibility leads to a decrease in performance. Thus, whether Backprop could be implemented or approximate by biological neurons remains an open question.

1.4.1 Dendrites as computational elements

The issue of oversimplified neuron models is by far the most frequent to be omitted from explanations of the biological implausibility of Backprop (See for example (Meulemans et al., 2020; Lillicrap et al., 2014)). This disregard might stem from the fact that rate-based point neurons are employed in many of the most powerful artificial neural networks. This fact might be taken as an argument that the simple summation of synaptic inputs is sufficient for powerful and generalized learning. Modelling neurons more closely to biology would by this view only increase mathematical complexity and computational cost without practical benefit. Another hypothesis states that the dominance of point neurons stems from a "somato-centric perspective" within neuroscience (Larkum et al., 2018), which stems from the technical challenges inherent to studying dendrites *in vivo*. The vastly different amount of available data regarding these two neuronal components might have induced a bias in how neurons are modelled computationally. Some researchers have even questioned whether dendrites should be seen as more of a 'bug' than a 'feature' (Häusser and Mel, 2003), i.e. a biological necessity which needs to be overcome and compensated for.

Yet in recent years, with novel mechanisms of dendritic computation being discovered, interest in researching and explicitly modelling dendrites has increased. Particularly the vast dendritic branches of pyramidal neurons found in the cerebral cortex, hippocampus and amygd-

dala, were shown to perform complex integrations of their synaptic inputs (Spruston, 2008). These dendritic trees are capable of performing coincidence- (Larkum et al., 1999) and sequence detection (Branco et al., 2010). The size of dendritic trees is also known to discriminate regular spiking from burst firing pyramidal neurons (van Elburg and van Ooyen, 2010). Furthermore, pyramidal neuron dendrites are capable of performing computations, which were previously assumed to require multi-layer neural networks (Schiess et al., 2016; Gidon et al., 2020). See (Larkum, 2022) and (Poirazi and Papoutsi, 2020) for extensive reviews.

These neuroscientific insights have sparked hope that modelling dendritic compartments explicitly might aid machine learning in terms in both learning performance and energy efficiency (Chavlis and Poirazi, 2021; Guerguiev et al., 2017; Richards and Lillicrap, 2019; Eyal et al., 2018). It appears that, if not for computational gains, dendrites should be considered essential for any model attempting to explain the power of human learning. While the network discussed in this thesis includes very simple multi-compartment models, the choice of model was strongly influenced by the fact that segregated dendrites were considered at all.

1.5 Cortical microcircuits

Another feature of the brain which is often not considered in (biologically plausible) machine learning models is its intricate connectivity. This is quite understandable, as there is still some uncertainty about which brain areas would be involved in Backprop-like learning. It is also unclear, to what level of detail these areas would need to be modeled. It has been shown that the connectivity patterns of cortical circuits are superior to amorphous networks in some cases (Häusser and Maass, 2007), so there might be a computational gain from modeling network structure closer to biology. The question over network structure goes hand in hand with the choice of neuron models, as synaptic connections arrive at specific points of pyramidal neuron dendrites, depending on the origin of the connection (Felleman and Van Essen, 1991; Ishizuka et al., 1995; Larkum et al., 2018).

Several theories of cortical function focus more on reinforcement (Legenstein et al., 2008) or unsupervised learning (George and Hawkins, 2009; Häusser and Maass, 2017). Without dismissing these theories, this thesis will adopt the viewpoint that human brains require a form of gradient descent to successfully adapt to their ever-changing environments. Furthermore, we share the hypothesis that this kind of learning occurs predominantly in the neocortex (Marblestone et al., 2016).

The literature on the subject of learning historically appears to be somewhat split (although several important exceptions have been published recently). On the one hand, the "machine-learning" point of view largely considers the utility of network changes first, with considerations of biology appearing as an afterthought **TODO: cite**. On the other hand, intricate models of cortical circuits exist, which can so far not be trained to perform tasks Potjans and Diesmann (2014); Schmidt et al. (2018); van Albada et al. (2022). Within this thesis, I hope to contribute

to the body of literature between those extremes. For this, my approach will be to select a learning model that is already highly biologically plausible, and to attempt to improve its plausibility - without breaking the learning rule.

1.6 Model selection

The model selection progress was strongly influenced by a review article on biologically plausible approximations of Backprop (Whittington and Bogacz, 2019). The authors narrow the wide range of proposed solutions down to four algorithms that are both highly performant and largely biologically plausible. Due to impact of the paper on this thesis, their model comparison is depicted in Supplementary Table S1. The algorithms were in part selected for requiring minimal external control during training, as well as by the fact that they can all be described within a common framework of energy minimization (Scellier and Bengio, 2017). The first two models are Contrastive learning (O'Reilly, 1996), and its extension to time-continuous updates (Bengio et al., 2017). Both of these encode neuron-specific errors in the change of neural activity over time. One of their appeals is the fact that they rely on Hebbian (and Anti-Hebbian) plasticity, which are highly regarded in the neuroscience literature (Magee and Grienberger, 2020; Brea and Gerstner, 2016). Yet in the plasticity rule also lies their greatest weakness, as synapses need to switch between the two opposing mechanisms once the target for a given stimulus is provided. This switch requires a global signal that communicates the change in state to all neurons in the network simultaneously.

The second class of models was more appealing to me, as both variants are based on the predictive coding account in Neuroscience (Rao and Ballard, 1999), which deserves its own introduction.

1.6.1 Predictive coding

In this seminal model of processing in the visual cortex, each level of the visual hierarchy represents the outside world at some level of abstraction. Recurrent connections then serve communicate prediction errors and predictions up and down the hierarchy respectively, which the network attempts to reconcile. The authors show that through rather simple computations, these prediction errors can be minimized to obtain useful representations at each level of the hierarchy. They further show that a predictive coding network trained on natural images exhibits end-stopping properties previously found in mammalian visual cortex neurons. This work was instrumental to shaping the modern neuroscientific perspective of perception being largely driven by cortico-cortical feedback connections in addition to the feedforward processes. The extension of predictive coding principles from visual processing to the entire living system is promising to revolutionize neuroscience under the name of *Active inference* (Friston, 2008; Friston and Kiebel, 2009; Adams et al., 2015). By this view, the entire brain aims to minimize prediction errors with respect to an internal (generative) model of the world. A noteworthy

property of this hypothesis is that it implies an agents action in the world as 'just another' way in which it can decrease discrepancies between its beliefs and sensory information. In a seminal paper, a model of the cortical microcircuit (Haeusler and Maass, 2007) was shown to have a plausible way for performing the computations required by predictive coding (Bastos et al., 2012).

While predictive coding was originally described as a mechanism for unsupervised learning, through a slight modification it is also capable of performing Backprop-like supervised learning (Whittington and Bogacz, 2017). This is the third model considered in the review paper, in which values (i.e. predictions) and errors of a layer are encoded in separate, recurrently connected neuron populations. By employing only local Hebbian plasticity, this network is capable of approximating Backprop in multilayer perceptrons while conforming to the principles of predictive coding. The constraint on network topology was later relaxed by showing that the model is capable of approximating Backprop for arbitrary computation graphs (Millidge et al., 2022). The neuron-based predictive coding network was therefore an important contribution towards unifying the fields of Active inference and machine learning research. As noted in a recent review article:

“Since predictive coding is largely biologically plausible, and has many potentially plausible process theories, this close link between the theories provides a potential route to the development of a biologically plausible alternative to backprop, which may be implemented in the brain. Additionally, since predictive coding can be derived as a variational inference algorithm, it also provides a close and fascinating link between backpropagation of error and variational inference.” (Millidge et al., 2021)

With this perspective in mind, we turn to the final model discussed in the review paper.

1.6.2 The Dendritic error model

The predictive coding network stores local prediction errors in nodes (i.e. neurons) close to the nodes to which these errors relate. That errors may be represented within the activation of individual neurons is a promising hypothesis with some advantages, as well as results backing it up (Hertäg and Clopath, 2022). Yet there is a competing view, by which errors elicited by individual neurons may be represented by membrane potentials of their dendritic compartments (Guerguiev et al., 2017). The "Dendritic error model" (Sacramento et al., 2018) - as the name implies - follows this line of thought. It contains a highly recurrent network of both pyramidal- and interneurons, in which pyramidal neuron apical dendrites encode prediction errors. This view is supported by behavioral rodent experiments which show that stimulation to pyramidal neuron apical tufts in cortical layer 1 controls learning (Doron et al., 2020).

For the errors to be encoded successfully, the model requires a symmetry between feedforward and feedback sets of weights, which it has to learn prior to training. After that, apical

compartments behave like the error nodes in a predictive coding network. They are silent during a feedforward network pass, and encode local prediction errors in their membrane potential when a target is applied to the output layer. Since they are a part of the pyramidal neuron, only local information is required to minimize these prediction errors through a plasticity rule for multi-compartment neurons (Urbanczik and Senn, 2014). A critical observation made in (Whittington and Bogacz, 2019) is that the dendritic error model is mathematically equivalent to their predictive coding network **TODO: expand if I have time, otherwise this will be a ref.** All of these factors combined make the dendritic error model a promising model to help us further understand both predictive coding and deep learning in cortical circuits. While both the employed neuron and connectivity model are far behind some of the more rigorous cortical simulations, it is regarded in the literature as an important step towards integrating deep learning and neuroscience.

Nevertheless, the model still suffers from some constraints with regard to its biological plausibility; Both the predictive coding network and the dendritic error network require strongly constrained connectivity schemes, without which they cannot learn. This kind of specificity (in particular one-to-one relationships between pairs of neurons) are highly untypical for cortical connections (Thomson and Bannister, 2003). Hence, their exact network architectures are unlikely to be present in the cortex. The Dendritic error model additionally requires Pre-training to be capable of approximating Backprop. Both of these issues will be discussed in this thesis. Yet the most salient improvement to the network’s biological plausibility is likely, to change neuron models from rate-based to spiking neurons. It has been shown that the Plasticity rule employed by the network is capable of performing simple learning tasks when adapted to spiking neurons (Stapmanns et al., 2021). Yet, (to the best of my knowledge) there are no studies investigating if this variant is capable of learning more complex tasks on a network-level. A spiking implementation of the dendritic error network will therefore be the starting point for this thesis, upon which further analysis shall build.

- Gerstner, W. and Naud, R. (2009). How good are neuron models? *Science*, 326(5951):379–380.
- Gewaltig, M.-O. and Diesmann, M. (2007). Nest (neural simulation tool). *Scholarpedia*, 2(4):1430.
- Gidon, A., Zolnik, T. A., Fidzinski, P., Bolduan, F., Papoutsi, A., Poirazi, P., Holtkamp, M., Vida, I., and Larkum, M. E. (2020). Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science*, 367(6473):83–87.
- Ginzburg, I. and Sompolinsky, H. (1994). Theory of correlations in stochastic neural networks. *Physical review E*, 50(4):3171.
- Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive science*, 11(1):23–63.
- Guerguiev, J., Lillicrap, T. P., and Richards, B. A. (2017). Towards deep learning with segregated dendrites. *ELife*, 6:e22901.
- Haeusler, S. and Maass, W. (2007). A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cerebral cortex*, 17(1):149–162.
- Haider, P., Ellenberger, B., Kriener, L., Jordan, J., Senn, W., and Petrovici, M. A. (2021). Latent equilibrium: A unified learning theory for arbitrarily fast computation with arbitrarily slow neurons. *Advances in Neural Information Processing Systems*, 34:17839–17851.
- Häusler, S. and Maass, W. (2017). Inhibitory networks orchestrate the self-organization of computational function in cortical microcircuit motifs through stdp. *bioRxiv*, page 228759.
- Häusser, M. and Mel, B. (2003). Dendrites: bug or feature? *Current opinion in neurobiology*, 13(3):372–383.
- Hertäg, L. and Clopath, C. (2022). Prediction-error neurons in circuits with multiple neuron types: Formation, refinement, and functional implications. *Proceedings of the National Academy of Sciences*, 119(13):e2115699119.
- Hines, M. L. and Carnevale, N. T. (1997). The neuron simulation environment. *Neural computation*, 9(6):1179–1209.
- Ishizuka, N., Cowan, W. M., and Amaral, D. G. (1995). A quantitative analysis of the dendritic organization of pyramidal cells in the rat hippocampus. *Journal of Comparative Neurology*, 362(1):17–45.
- Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cerebral cortex*, 17(10):2443–2452.

- Kandel, E., Koester, J., Mack, S., and Siegelbaum, S. (2021). *Principles of Neural Science, Sixth Edition*. McGraw-Hill Education.
- Kandel, E. R. (1968). Dale’s principle and the functional specificity of neurons. In *Psychopharmacology; A Review of Progress, 1957–1967*, pages 385–398. US Government Printing Office Washington, DC.
- Kawaguchi, Y. (2001). Distinct firing patterns of neuronal subtypes in cortical synchronized activities. *Journal of Neuroscience*, 21(18):7261–7272.
- Khaligh-Razavi, S.-M. and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS computational biology*, 10(11):e1003915.
- Kubilius, J., Bracci, S., and Op de Beeck, H. P. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS computational biology*, 12(4):e1004896.
- Larkum, M. E. (2022). Are dendrites conceptually useful? *Neuroscience*, 489:4–14.
- Larkum, M. E., Petro, L. S., Sachdev, R. N., and Muckli, L. (2018). A perspective on cortical layering and layer-spanning neuronal elements. *Frontiers in neuroanatomy*, page 56.
- Larkum, M. E., Zhu, J. J., and Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398(6725):338–341.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.
- Lee, C., Sarwar, S. S., Panda, P., Srinivasan, G., and Roy, K. (2020). Enabling spike-based backpropagation for training deep neural network architectures. *Frontiers in neuroscience*, page 119.
- Lee, D.-H., Zhang, S., Fischer, A., and Bengio, Y. (2015). Difference target propagation. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7–11, 2015, Proceedings, Part I 15*, pages 498–515. Springer.
- Lee, J. H., Delbruck, T., and Pfeiffer, M. (2016). Training deep spiking neural networks using backpropagation. *Frontiers in neuroscience*, 10:508.
- Legenstein, R., Pecevski, D., and Maass, W. (2008). A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. *PLoS computational biology*, 4(10):e1000180.
- Leinweber, M., Ward, D. R., Sobczak, J. M., Attinger, A., and Keller, G. B. (2017). A sensorimotor circuit in mouse cortex for visual flow predictions. *Neuron*, 95(6):1420–1432.

- Liao, Q., Leibo, J., and Poggio, T. (2016). How important is weight symmetry in backpropagation? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.
- Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2014). Random feedback weights support learning in deep neural networks. *arXiv preprint arXiv:1411.0247*.
- Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., and Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6):335–346.
- Magee, J. C. and Grienberger, C. (2020). Synaptic plasticity forms and functions. *Annual review of neuroscience*, 43:95–117.
- Marblestone, A. H., Wayne, G., and Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. *Frontiers in computational neuroscience*, page 94.
- Mazzoni, P., Andersen, R. A., and Jordan, M. I. (1991). A more biologically plausible learning rule for neural networks. *Proceedings of the National Academy of Sciences*, 88(10):4433–4437.
- McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of neurophysiology*, 54(4):782–806.
- Meulemans, A., Carzaniga, F., Suykens, J., Sacramento, J., and Grewe, B. F. (2020). A theoretical framework for target propagation. *Advances in Neural Information Processing Systems*, 33:20024–20036.
- Millidge, B., Seth, A., and Buckley, C. L. (2021). Predictive coding: a theoretical and experimental review. *arXiv preprint arXiv:2107.12979*.
- Millidge, B., Tschantz, A., and Buckley, C. L. (2022). Predictive coding approximates backprop along arbitrary computation graphs. *Neural Computation*, 34(6):1329–1368.
- Millidge, B., Tschantz, A., Seth, A. K., and Buckley, C. L. (2020). Activation relaxation: A local dynamical approximation to backpropagation in the brain. *arXiv preprint arXiv:2009.05359*.
- Mozafari, M., Ganjtabesh, M., Nowzari-Dalini, A., Thorpe, S. J., and Masquelier, T. (2018). Combining stdp and reward-modulated stdp in deep convolutional spiking neural networks for digit recognition. *arXiv preprint arXiv:1804.00227*, 1.
- O’Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural computation*, 8(5):895–938.
- Poirazi, P. and Papoutsis, A. (2020). Illuminating dendritic function with computational models. *Nature Reviews Neuroscience*, 21(6):303–321.

- Potjans, T. C. and Diesmann, M. (2014). The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cerebral cortex*, 24(3):785–806.
- Potjans, W., Diesmann, M., and Morrison, A. (2011). An imperfect dopaminergic error signal can drive temporal-difference learning. *PLoS computational biology*, 7(5):e1001133.
- Rao, R. P. and 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.
- Richards, B. A. and Lillicrap, T. P. (2019). Dendritic solutions to the credit assignment problem. *Current opinion in neurobiology*, 54:28–36.
- Roelfsema, P. R. and Holtmaat, A. (2018). Control of synaptic plasticity in deep cortical networks. *Nature Reviews Neuroscience*, 19(3):166–180.
- Sacramento, J., Ponte Costa, R., Bengio, Y., and Senn, W. (2018). Dendritic cortical microcircuits approximate the backpropagation algorithm. *Advances in neural information processing systems*, 31.
- Scellier, B. and Bengio, Y. (2017). Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Frontiers in computational neuroscience*, 11:24.
- Schiess, M., Urbanczik, R., and Senn, W. (2016). Somato-dendritic synaptic plasticity and error-backpropagation in active dendrites. *PLoS computational biology*, 12(2):e1004638.
- Schmidhuber, J. (2014). Who invented backpropagation. *More.[DL2]*.
- Schmidt, M., Bakker, R., Hilgetag, C. C., Diesmann, M., and van Albada, S. J. (2018). Multi-scale account of the network structure of macaque visual cortex. *Brain Structure and Function*, 223(3):1409–1435.
- Seung, H. S. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron*, 40(6):1063–1073.
- Sporea, I. and Grüning, A. (2013). Supervised learning in multilayer spiking neural networks. *Neural computation*, 25(2):473–509.
- Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, 9(3):206–221.
- Stapmanns, J., Hahne, J., Helias, M., Bolten, M., Diesmann, M., and Dahmen, D. (2021). Event-based update of synapses in voltage-based learning rules. *Frontiers in neuroinformatics*, page 15.
- Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

- Svensson, E., Apergis-Schoute, J., Burnstock, G., Nusbaum, M. P., Parker, D., and Schiöth, H. B. (2019). General principles of neuronal co-transmission: insights from multiple model systems. *Frontiers in neural circuits*, 12:117.
- Thomson, A. M. and Bannister, A. P. (2003). Interlaminar connections in the neocortex. *Cerebral cortex*, 13(1):5–14.
- Thomson, A. M., West, D. C., Wang, Y., and Bannister, A. P. (2002). Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2–5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro. *Cerebral cortex*, 12(9):936–953.
- Urban-Ciecko, J. and Barth, A. L. (2016). Somatostatin-expressing neurons in cortical networks. *Nature Reviews Neuroscience*, 17(7):401–409.
- Urbanczik, R. and Senn, W. (2014). Learning by the dendritic prediction of somatic spiking. *Neuron*, 81(3):521–528.
- van Albada, S., Morales-Gregorio, A., Dickscheid, T., Goulas, A., Bakker, R., Bludau, S., Palm, G., Hilgetag, C.-C., and Diesmann, M. (2022). Bringing anatomical information into neuronal network models. In *Computational Modelling of the Brain*, pages 201–234. Springer.
- Van Albada, S. J., Rowley, A. G., Senk, J., Hopkins, M., Schmidt, M., Stokes, A. B., Lester, D. R., Diesmann, M., and Furber, S. B. (2018). Performance comparison of the digital neuromorphic hardware spinnaker and the neural network simulation software nest for a full-scale cortical microcircuit model. *Frontiers in neuroscience*, 12:291.
- van Elburg, R. A. and van Ooyen, A. (2010). Impact of dendritic size and dendritic topology on burst firing in pyramidal cells. *PLoS computational biology*, 6(5):e1000781.
- Vezoli, J., Falchier, A., Jouve, B., Knoblauch, K., Young, M., and Kennedy, H. (2004). Quantitative analysis of connectivity in the visual cortex: extracting function from structure. *The Neuroscientist*, 10(5):476–482.
- Werfel, J., Xie, X., and Seung, H. (2003). Learning curves for stochastic gradient descent in linear feedforward networks. *Advances in neural information processing systems*, 16.
- Whittington, J., Muller, T., Mark, S., Barry, C., and Behrens, T. (2018). Generalisation of structural knowledge in the hippocampal-entorhinal system. *Advances in neural information processing systems*, 31.
- Whittington, J. C. and 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.

- Whittington, J. C. and Bogacz, R. (2019). Theories of error back-propagation in the brain. *Trends in cognitive sciences*, 23(3):235–250.
- Yamins, D. L. and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356–365.
- Zenke, F. and Ganguli, S. (2018). Superspike: Supervised learning in multilayer spiking neural networks. *Neural computation*, 30(6):1514–1541.