

Opinion

# Contributions by metaplasticity to solving the Catastrophic Forgetting Problem

Peter Jedlicka,<sup>1,2,3,7,\*</sup> Matus Tomko,<sup>1,4,7</sup> Anthony Robins,<sup>5,7</sup> and  
Wickliffe C. Abraham<sup>6,7,\*</sup>

**Catastrophic forgetting (CF) refers to the sudden and severe loss of prior information in learning systems when acquiring new information. CF has been an Achilles heel of standard artificial neural networks (ANNs) when learning multiple tasks sequentially. The brain, by contrast, has solved this problem during evolution. Modellers now use a variety of strategies to overcome CF, many of which have parallels to cellular and circuit functions in the brain. One common strategy, based on metaplasticity phenomena, controls the future rate of change at key connections to help retain previously learned information. However, the metaplasticity properties so far used are only a subset of those existing in neurobiology. We propose that as models become more sophisticated, there could be value in drawing on a richer set of metaplasticity rules, especially when promoting continual learning in agents moving about the environment.**

## The catastrophic forgetting problem

The representations created by learning systems must be both stable enough to preserve information over time, yet plastic enough to accommodate new learning. These requirements are in direct opposition, creating a ‘stability–plasticity dilemma’ [1,2]. Compared with symbolic models (where representations are discrete), pattern-based systems such as the brain and **ANNs** (see [Glossary](#)), where the representing medium is shared, face an acute form of the dilemma. Because learning episodes make changes to the representing medium (synaptic connection strengths or artificial ‘weights’ between nodes), they typically interact with each other, allowing the plasticity required for new learning to disrupt the stability of older memories over time. This disruption is known as ‘**catastrophic forgetting**’ (CF) [3–5].

Evolution has largely resolved this dilemma in the brain. By contrast, ANNs, including **deep neural networks**, in general suffer from excessive plasticity that is associated with severe CF. Accordingly, an immense literature has evolved aiming to solve the CF problem while still permitting sequential or **continual learning**, as the brain does.

Here, we briefly review a few of the main approaches, some partially overlapping, that have been used in the ANN literature to solve CF (see [6–8] for reviews). Interestingly, most of these solutions have strong correlates, either intentionally or unintentionally, in biology [9]. We then note how common it is for correlates of biological **metaplasticity** [10] to also be included in algorithms, even if not explicitly stated as such. Given its usefulness so far, we describe other aspects of metaplasticity that are not commonly used but that might offer additional capability for network models to enhance adaptability while minimising CF.

## Solutions to catastrophic forgetting in ANNs

CF has been explored predominantly in feed-forward (recently deep) networks, where information is learned in ‘batches’ (e.g., separate episodes or tasks) with a supervised learning algorithm

## Highlights

Catastrophic forgetting (CF) is the sudden and severe loss of memory for previously stored information due to the learning of new information. Standard artificial neural networks have traditionally suffered from CF when undertaking sequential learning tasks.

A wide variety of strategies have been used to minimise forgetting, such as generative replay, constructive algorithms, complementary systems, sparse storage, and meta-learning.

Metaplasticity is the activity-dependent plasticity of future plasticity or learning. Various network models use metaplasticity-type strategies to reduce forgetting during continual learning.

Combining metaplasticity-type rules with other strategies often improves network performance even further.

Models may benefit even more by drawing on other properties of biological metaplasticity phenomena.

<sup>1</sup>ICAR3R – Interdisciplinary Centre for 3Rs in Animal Research, Faculty of Medicine, Justus Liebig University, Giessen, Germany

<sup>2</sup>Institute of Clinical Neuroanatomy, Neuroscience Center, Goethe University Frankfurt, Frankfurt/Main, Germany

<sup>3</sup>Frankfurt Institute for Advanced Studies, Frankfurt 60438, Germany

<sup>4</sup>Institute of Molecular Physiology and Genetics, Centre of Biosciences, Slovak Academy of Sciences, Bratislava, Slovakia

<sup>5</sup>Department of Computer Science, University of Otago, Dunedin 9016, New Zealand

<sup>6</sup>Department of Psychology, Brain Health Research Centre, University of Otago, Dunedin 9054, New Zealand

<sup>7</sup>All authors contributed equally.



[3–5,11,12]. It has also been studied in the context of self-supervised learning [13,14], reinforcement learning [12,15], and unsupervised learning [16,17]. Related problems occur in the context of Hopfield-type and other dynamical networks, where exceeding the capacity of the system leads to a complete collapse of recall [18]. While specific methods for addressing CF can occur in particular contexts, the main solution types are noted as follows.

### Generative replay

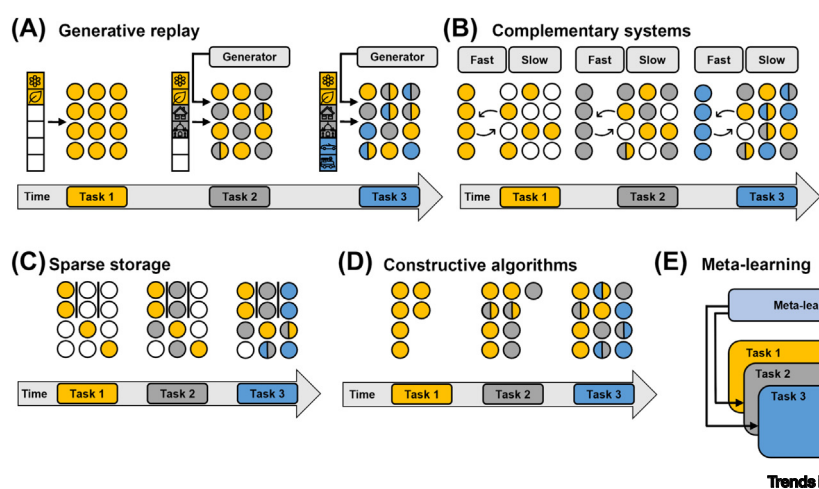
Of the early attempts to solve the CF problem in ANNs, the most effective was simply to rehearse previously learned information during new learning episodes [19]. However, the consequent need to access a separate store of old patterns is clumsy and unrealistic. Fortunately, the benefits of rehearsal can be attained by using approximations of old information, that is, patterns generated by sampling the behaviour of the trained ANN (Figure 1A). Originally termed ‘pseudorehearsal’ [4,20], the more popular recent term ‘generative replay’ [21] better captures the similarity with replay mechanisms involved in learning in the brain [22–25].

### Complementary systems

The complementary learning systems (CLS) approach is inspired by two memory systems in the brain, which as a simplification can be conceptualised as the rapidly learning hippocampus and the slowly learning neocortex [26–28]. The CLS (aka **dual-net/dual-memory**) strategy belongs to the so-called architectural approaches for avoiding CF. ANNs typically behave like the neocortical memory system because they learn slowly, while being repeatedly trained by a large number of input patterns [28]. To improve continual learning in these networks, models have implemented a second, rapidly learning hippocampus-like system that would mediate a replay of patterns [28,29], although in a different way to generative replay (Figure 1B). In a new version of the CLS, the hippocampal system can repeatedly train the neocortical system by replaying non-overlapping representations of multiple tasks [28,30]. Deep networks can also benefit from an additional memory system [23,30,31].

### Sparse storage

Sparsity of memory storage is another feature exploited in both ANNs and the brain to reduce representational overlap and CF (Figure 1C) [5,11,32–34]. Sparsity is closely linked to modularity. Both have been shown to improve learning new tasks while alleviating CF [35], providing robustness to noise, and enhancing memory capacity, storage efficiency, and speed of learning [36,37].



**Figure 1.** Diagrammatic representations of common types of solutions to catastrophic forgetting in artificial neural networks (ANNs). (A) Generative replay: approximations of old information are produced by a generator that are replayed and mixed with input to update the network. Coloured circles represent active neurons for a given task, colour-coded according to the respective task. (B) Complementary systems: the model consists of fast, hippocampus-like memory storage network and a slow neocortex-like memory storage network. The fast system repeatedly trains the slower system by replaying of stored patterns. (C) Sparse storage: only parts of a network are active for a given task. This reduces interference and produces modularity (task-specific subnetworks). (D) Constructive

algorithms: learning of a new task is ensured by growing or pruning nodes and connections. (E) Meta-learning: after learning each task, a meta-learning phase follows, during which the model learns how to update the learner's parameters for faster learning in the next task. Diagrams inspired by [7].

\*Correspondence:  
peter.jedlicka@informatik.med.uni-giessen.de (P. Jedlicka) and  
cliff.abraham@otago.ac.nz  
(W.C. Abraham).

Remarkably, biologically inspired mechanisms such as dendritic computation support the emergence of sparse task-specific subnetworks, reducing **task interference** and facilitating continual learning as compared with standard ANNs [38,39]. Sparse coding occurs widely across brain regions [28]. A classic example is the hippocampal dentate gyrus which uses very sparse representations (e.g., [40]) for pattern discrimination [41] and potentially for reducing CF [42].

### Constructive algorithms

In addition to adjusting the **parameters** of existing network architecture, constructive algorithms also learn by growing or pruning nodes and connections (Figure 1D). In early implementations of these algorithms, nodes were added so as to capture a particular component of error [43], or represent a new input category [44]. Whole latent networks can also be brought online for task-specific purposes [45]. Statistical and evolutionary methods have been used for optimisation [46]. Recent examples include applications to deep networks [47] and progressive networks [48,49]. Biological networks also use constructive algorithms in the form of structural plasticity, including growing and pruning of synapses [50] as well as neurogenesis [51]. Intriguingly, both synaptic structural plasticity and neurogenesis have been studied as solutions for CF in the brain [42] as well as in deep ANNs [31,52].

### Meta-learning

Meta-learning is a term from psychology and cognitive science covering various forms of 'learning to learn'. The topic received renewed interest within machine learning [53], with the goal of allowing models to train more quickly on new tasks by learning how to update the learner's parameters, for example, by determining the most effective parameters to update for each task (Figure 1E) [54]. Meta-learning can include the goal of minimising task interference/CF so as to facilitate continual learning [55–57]. Recent approaches have included synaptic metaplasticity as well [58].

### Palimpsest models

Related issues have been explored in the context of Hopfield-type/dynamical networks [18]. Such networks have a maximum storage capacity; when exceeded, it 'makes all memory states irretrievable' [18]. This breakdown could certainly be described as catastrophic, but it arises from different reasons than those for the CF discussed previously, namely, an absolute memory capacity being exceeded, rather than interference between tasks that are typically well within the storage capacity of the network. Techniques used to address this form of memory breakdown were typically based on 'unlearning' to increase capacity [59,60], or managed forgetting so as to preserve only recently learned information, giving rise to 'palimpsest' models [61,62].

### Metaplasticity and its expression in neurobiology

Despite the advantages offered by the above-mentioned approaches to solving CF, many researchers have found that using metaplasticity principles also contributes important functionality in their models, either alone or in combination with other tools. In the biological context, metaplasticity is defined as the ability of neural activity at one point in time to affect the direction, duration, or degree of synaptic plasticity in the future (Figure 2A) [63]. Before considering how metaplasticity concepts are being implemented in ANNs, we will briefly review some of the neurobiological properties of metaplasticity, without being concerned with mechanistic details (see [10,64] for detailed reviews).

Metaplasticity refers to a family of phenomena. In some cases, prior 'priming' activity in one set of synapses homosynaptically impairs (or enhances) the induction of long-term potentiation (LTP) while enhancing (impairing) long-term depression (LTD) at those same synapses. In other examples, prior activity can enhance heterosynaptically the duration of LTP or LTD at nearby synapses (e.g., synaptic tag and capture phenomena [65]). More dramatically, prior activity or indeed its

### Glossary

#### Artificial neural network (ANN):

mathematical model of artificial 'neurons' (as computational nodes/units) organised in layers and coupled via 'synapses' (as connections) with modifiable weights.

#### Bienenstock, Cooper, and Munro (BCM) computational model:

a computational model of visual cortex plasticity designed to explain experience-dependent modification due to altered visual experience. It incorporates a dynamic activity-plasticity function based on the history of cell firing in the network (i.e., metaplasticity).

**Catastrophic forgetting (CF):** severe loss of memory for previously learned information during the sequential learning of new information due to alterations of connection weights that are key for retaining the old information.

#### Continual (lifelong/cumulative)

**learning:** the ability to learn tasks in an ongoing and sequential way across time without undue forgetting of previous tasks. Modelling is driven by the understanding that the obvious solution of continuous replay of old tasks concurrently with new learning is unrealistic and too energy intensive.

#### Deep neural network:

an ANN consisting of many layers, where layers often have specialised functionality. Deep neural networks are typically more powerful than 'shallow' networks, and are currently the dominant paradigm in machine learning.

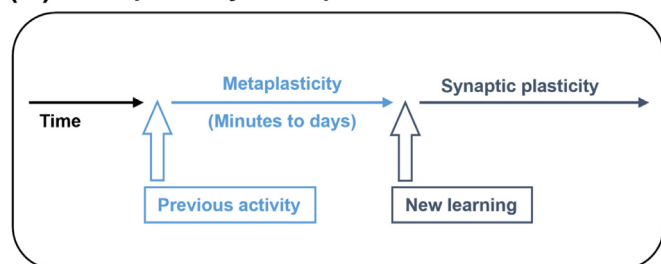
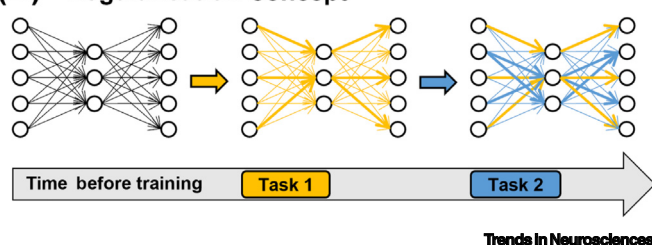
**Dual net:** a dual-memory architecture consisting of two structural components, each designed to use different learning strategies and which can have different functionalities.

**Hyperparameter:** values that control the learning process, for example, number of layers or nodes, learning rates, number of training epochs/iterations, and many more.

**Learning rate:** the rate at which synaptic or connection weights can be modified during learning. Metaplasticity-inspired models use rate parameters that are themselves modifiable based on experience or the needs of the network.

**Memory drift:** the experimentally observed phenomenon of slowly shifting combinations of cells representing information in a neural circuit.

**Metaplasticity:** the plasticity of synaptic plasticity involving activity-dependent and lasting changes

**(A) Metaplasticity concept****(B) Regularisation concept****Figure 2. Metaplasticity in biology and models.**

(A) Metaplasticity concept as typically studied in neurobiological experiments. Neural activity (or inactivity) affects the degree, direction, or duration of subsequently induced synaptic plasticity elicited either by experimental protocols or behaviour. (B) Metaplasticity-based regularisation for knowledge retention.

Diagrammatic representation of one type of the regularisation concept, whereby important connections (marked as thicker arrows) are estimated for each learned task, and changing them is penalised (connections remain as thicker arrows) when learning a next task.

of synaptic plasticity capability or learning rate, without necessarily changing the current synaptic strength.

**Parameter:** in the context of learning systems, parameters are values that are adjusted by the learning process; in ANNs, these are typically the weights and biases of nodes.

**Pareto optimality:** the state by which a system meets multiple competing objectives in an optimal way without necessarily optimising performance on any one of them.

**Regularisation:** modifies learning algorithms in ANNs to avoid overfitting and improve generalisation. Some regularisation techniques (e.g., elastic weight configuration and synaptic intelligence) have parallels to metaplasticity as they impose constraints on changes of synapses that are relevant for old knowledge. Thereby, they protect old knowledge against interference by new knowledge.

**Task interference:** memory of one task interferes with memory/learning of a different task. In proactive interference, old memory traces disturb learning of new memory traces. In retroactive interference, newly learned memory traces disturb old memory traces.

absence can change plasticity thresholds heterodendritically throughout a cell [66], or even across a whole network. The triggers of metaplasticity induction include primary synaptic neurotransmitters, neuromodulatory transmitters including gliotransmitters, and hormones.

Based on the breadth of properties noted previously, experimental and theoretical analyses suggest that metaplasticity phenomena may play several roles in learning by biological neural networks, including lowering the threshold for plasticity when the time is right, binding of information stored at neighbouring synapses and branchlets, or raising the threshold for plasticity to preserve recently learned information during its consolidation or to stabilise the overall level of network activity, avoiding hyperplasticity and neurotoxicity resulting from over-activation of the network [64,67].

**Metaplasticity-based solutions to catastrophic forgetting**

Metaplasticity-based approaches, for example, synaptic cascade models [68] (Box 1), have long been a powerful tool for alleviating CF in ANNs. More recent metaplasticity-related strategies were biologically inspired by synaptic consolidation [12,69], which slows down the weakening of previously strengthened synapses. This form of metaplasticity in the brain stabilises memories and has been hypothesised to make them resistant to interference by new memories. A different mechanism may be the addition of an inhibitory input to a dendritic spine which also receives an excitatory input, thereby reducing its plasticity [70]. Similarly, in ANNs, the metaplasticity-based approaches protect consolidated memories by decreasing synaptic plasticity rates. For instance, some **regularisation** methods penalise the changes of connections that are relevant for memories of previous tasks (Figure 2B) [12,71]. Thus, the regularisation approaches such as elastic weight consolidation (EWC) [12], synaptic intelligence (SI) [71], and memory aware synapses (MAS) [72] all estimate the importance of connections for previously learned tasks and penalise changes on them. The importance of a synapse (or uncertainty about its importance [72]), which is linked to its **learning rate**, can thus be viewed as a metaplastic state of the synapse [73,74]. Conceptually, synaptic weight changes can be considered as first-order adaptations in the network and learning rate changes (metaplasticity) as second-order adaptations [75].

### Box 1. Cascade models and synaptic meta-states

Application of metaplasticity principles such as regularisation implements a dynamically varying learning rate to connections depending on the history of activity and evaluated importance of the connection for a prior memory or memories. In a different formulation, synapses or connections have what have been termed meta-states, which refer to a linear set of discrete states for a depressed connection and a corresponding set for a potentiated one [68]. The greater the depth of the state, the harder it is to transition to the other connection weight [73]. Learning trials can move the weights up and down through the ‘cascade’ of meta-states, often without changing the connection strength until the threshold is reached for doing so, which varies depending on the meta-state the connection resides in at the time. The meta-states, therefore, can be conceptualised as metaplasticity variables.

Cascade-type models as described in the preceding text typically use binarised networks, which have limitations. But models have shown that meta-states can be used to enhance memory lifetimes and protect against over-writing, yet still provide for graceful degradation (forgetting) like real memories [68,69]. Moreover, such meta-state algorithms have been used successfully in models of real-world learning and choice under reward uncertainty [109]. In biology, hippocampal synapses have been shown to have an equivalent to meta-states, whereby the ability to obtain LTP or LTD is constrained to discrete states in some paradigms, with the transition between states dependent on the history activity at the synapses [110,111].

There is one important difference between earlier multistate cascade models of plasticity [68,69] and the regularisation methods such as EWC and SI. Cascade models with multiple hidden (metaplastic) states incorporating different degrees of plasticity [69] can simulate both remembering and forgetting memories. By contrast, EWC and SI focus only on the retention of learned memories by minimising the disruption by new memories. This is the reason why the regularisation approaches are sometimes also called ‘knowledge retention methods’ [55]. Synaptic plasticity rates in cascade models can increase or decrease with time. By contrast, synaptic plasticity rates in EWC/SI can only decrease with increasing simulation time and with newly stored more memory patterns [12]. Consequently, in EWC, exceeding the memory capacity eventually leads to a ‘blackout catastrophe’ [12] with no memory retention.

Given the advantages offered by EWC/SI/regularisation for continual learning, it is not unexpected that as ANNs evolved to incorporate more than one strategy to improve further their performance, metaplasticity-like algorithms became commonly incorporated. For example, SI has been recently combined with neurons equipped with active dendrites [38] to effect improved continual learning. These results suggest that biological realism in the form of active dendrites and metaplasticity jointly alleviates CF and improves knowledge retention for multiple tasks [38]. Even passive dendrites are capable of normalisation that improves learning in ANNs by slowing down the learning rate of highly connected neurons with presumably large dendrites [76]. Another recent study found that regularisation methods prevented CF for incremental learning of new tasks but not for incremental learning of new classes [24]. However, when combined with generative replay, regularisation contributed successfully to the continual learning of new classes. The authors concluded that regularisation (a type of metaplasticity) and generative replay are complementary approaches for suppressing memory interference, being analogous to cellular and systems consolidation in the brain.

Other recent brain-inspired models also combine metaplasticity with complementary methods to alleviate CF. This includes extending EWC/SI by adding a context-dependent gating signal (XdG) [77] or presynaptic plasticity and stochasticity [78]. In these models, synaptic stabilisation by EWC/SI complements the presynaptically or XdG-mediated sparsity of representations (but see also [79]). Likewise, it was shown that not only sparse storage but also meta-learning (‘learning to learn’) works better with metaplasticity, improving transfer learning [80]. Several other studies successfully combined regularisation with rehearsal/replay [81–83] or regularisation with architectural techniques [84–86] for continuous learning in deep networks.

Overall, in line with observations from neuroscience, hybrid approaches such as the conjunctive use of metaplasticity with other algorithms improve reduction of CF more than the individual approaches. However, currently used approaches rarely combine more than two strategies (for an overview, see Table 2 in [8]). Exceptions are ‘variational continual learning’, which integrates regularisation, rehearsal, and architecture strategies [87]; ‘growing dual-memory’, which incorporates regularisation, dual-memory architecture, generative replay, and constructive algorithms [88]; and ‘meta-experience replay’, which uses regularisation, rehearsal, and meta-learning [89].



Nevertheless, the potential synergies of merged strategies have not yet been exhaustively investigated. Therefore, future research should explore which combinations of approaches outperform other combinations on continuous learning benchmarks (cf. [8]), and how metaplasticity-based rules can contribute to performance enhancement.

### Other potential contributions of metaplasticity

As noted in the preceding text, the brain uses more than one strategy to minimise CF [7]. Thus, it is not surprising that ANNs are evolving towards a combination of principles that are also features of learning in neurobiology, such as sparsity of connectivity and representations, complementary systems, neurogenesis, and replay during sleep and other quiet states. These features may vary in exactly how they occur in the diverse neural circuits that underpin information storage, but we contend that metaplasticity is also fundamental to these processes, as it is for many network models as described previously.

Metaplasticity may have yet more to contribute to network models of continual learning, given that it operates at multiple levels of cell function, from the synaptic to network levels [64]. Moreover, these metaplasticity mechanisms can coexist, generating a complex neural state dependent on past experience. In ANNs, by contrast, the metaplasticity-style modifications of the learning rate are typically implemented only connection by connection, as in SI and EWC, or more widely based on reinforcement and top-down meta-learning, but not both. In either case, the learning rate is a single **hyperparameter** affecting both increases and decreases in weights equally, with some exceptions. Accordingly, we propose that deep learning models, especially those striving to implement lifelong learning by agents moving in changing environments, might increasingly benefit from including more sophisticated and diverse metaplasticity hyperparameters as seen in neurobiology. We give a few examples below of strategies that could merit consideration for incorporation into more models (see [Outstanding questions](#)).

- (i) Typically, for each connection, the learning rate applies equally to weight increases and decreases, and when the learning rate changes, it is applied equally to future weight increases and decreases. More flexibility, however, might be obtained by creating separate learning rates for increases and decreases, such that they could then be controlled proportionally, in opposite directions, or even completely independently, depending on the use and importance of that connection for learning particular tasks. This would be akin to homosynaptic metaplasticity, where, for example, LTP at a synapse can make it harder to get further LTP but increase the chance for LTD [90]. Of note, future LTD can also be inhibited after LTP induction, thus equivalent to lowering the learning rate in models, but this is typically only a temporary state in real synapses that would likely not be relevant to models without such temporal coding.
- (ii) On top of a learning rate per connection, there could be a learning rate for a node as well that changes depending on the node's usefulness in multi-trial (multi-shot) task learning. This would be equivalent to having modifiable excitability (i.e., intrinsic plasticity) for a neuron [91], which would make any of its synaptic connections more (or less) able to change in strength. A hyperparameter for this could assist in sparsifying cells and connections participating in any particular learning episode by focusing change to more active nodes and away from less active ones. A node learning rate (e.g., [34]) might also be applicable to generative algorithms, where newly formed nodes and connections to a network could have a transiently enhanced learning rate for the new information at hand, equivalent to what happens during neurogenesis where recently born neurons have a greater excitability and a lower threshold for LTP for a limited time during maturation [92].
- (iii) Further nuance could be achieved by implementing a competing interaction between connections terminating at the same node, that is, heterosynaptic metaplasticity. For example, LTP at

a subset of inputs could cause a reduction in the learning rate at adjacent connections [93]. This is a feature of the **Bienenstock, Cooper, and Munro (BCM) computational model** of synaptic plasticity, whereby over (under) activity of a cell leads to decreased (enhanced) future LTP but enhanced (decreased) LTD at all its afferent excitatory synapses. The model has been used to account for recovery of function in sensory cortex [67,94], and other types of plasticity [95]. It has the benefit also of helping maintain homeostasis of the net weights onto individual neurons, preventing under-utilisation or over-dominance of individual cells in the system. When applied in networks [96], this feature could contribute to sparsification, since activity-dependent use of a connection between nodes that strengthens that connection and increases the activity of the 'postsynaptic' node would then make it harder to generate such increases in weight at other incoming connections to that node when learning new tasks, directing positive weight changes to elsewhere in the network. In addition to sparsification, such heterosynaptic metaplasticity could be a way of enhancing signal-to-noise ratio of the various inputs that compete for control of the activity at that given node. Alternatively, in some circumstances, enhancing the learning rate heterosynaptically could prove beneficial [65,97]. Here, the lifetime of synaptic change is enhanced by the heterosynaptic sharing of plasticity-related proteins across time via the 'synaptic tag and capture' (STC) process [65]. Initial implementations of STC in models have revealed improved memory lifetime and strength [98,99], encouraging its extension to addressing the CF problem.

- (iv) Rather than learning always being available, there could be an overall threshold for learning in the network. For an agent moving or at least responding to a changing environment, it could be maladaptive to be trying to record all events into memory. Rather, through detection of novel or important stimuli or contexts, the learning rate could be ramped up for the whole network, or a segment of it in a complementary systems model, before relaxing back during more quiet periods of environmental interaction. This feature has already been adapted in part or in full by self-organising networks [100] and models of reinforcement [101]. It may also be useful for other purposes, such as models of psychotherapy, where they have been used to address situations where (from a conceptual viewpoint) the learning rate has been turned down by prior traumatic life experiences, leading to 'catastrophic remembering' [5,102]. This concept could also be useful if attempting to model reactivation of consolidated memories, allowing for their updating with new but related information, before reconsolidating through regularisation procedures.
- (v) Where networks are constrained by their size, this also limits the number of items that can be learned. But in lifelong learning models, it can be predicted that not all memories need to be retained, nor should spuriously stored information. This is essentially an inbuilt feature of palimpsest models, although when the rate of memory loss becomes too high then the system tips over into CF. There is also a use of unlearning algorithms particularly for recurrent neural networks such as Hopfield nets. There has been interest in applying such principles, including in models of sleep [59,101,103], although there may be costs to doing so [35], while generative replay may offer more advantages [60]. In the models of unlearning, the rate of change is typically fixed. However, it could be advantageous to have the ability to alter unlearning rates in ways that could optimise the decay of spurious memories or those no longer likely to be needed. Of course, a remaining trace could still be useful to underpin savings and updating during relearning.

The usefulness of these suggestions may depend in part on the type of model being implemented. For example, while most of the models described previously are ANNs, with layers of nodes and widespread connections between them, so-called spiking neural networks (SNNs) are also popular. Here, neurobiology principles are more evident, as nodes are more neuron-like with 'leaky integrate and fire' properties while the connections likewise become more synapse-like. When the neurons have dendritic structure, as opposed to being 'point' neurons with no structure, then the distribution of synapses on branching dendritic trees adds computational complexity and

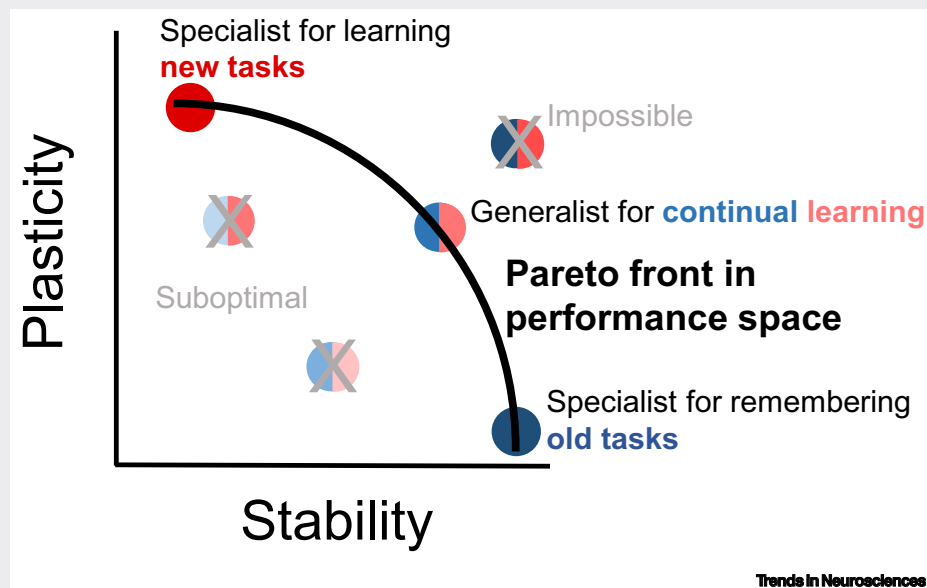
### Box 2. Plasticity–stability–economy trade-offs and Pareto optimality

Continual learning of new tasks requires a balance between plasticity and stability. In addition, memories cost energy, although biological neural networks process and memorise information at relatively low energy costs [112,113]. Nonetheless, the energy costs limit the amount of neural resources available, creating a trade-off for the performance of new and old tasks [6]. Importantly, no neural network can be optimal for all three objectives of plasticity, stability, and low energy cost (economy). Compellingly, the evolutionary framework of Pareto theory proposes that complex systems including neural networks evolved towards **Pareto optimality** (Figure 1) for the trade-offs between multiple competing objectives [114,115].

By definition, Pareto optimal networks cannot be improved in any objective without worsening other objectives (Figure 1). This suggests that biological networks avoid CF by finding solutions that optimise stability, plasticity, and energy efficiency. Accordingly, biologically realistic ANN models seem to achieve a better performance for the trade-off between learning accuracy and its energy cost (estimated as number of synapses) than their less realistic (e.g., binarised) counterparts [73,116]. Remarkably, the multi-objective Pareto optimisation approach improves CF more efficiently than a single-objective approach [117]. It remains to be addressed whether deep network models can reproduce these earlier findings.

CF is a consequence of too much plasticity with too little stability in ANNs. One metaplasticity approach for alleviating CF (e.g., EWC, SI) shifts the plasticity–stability trade-off towards stability for synapses that are relevant for prior tasks. In this way, metaplasticity facilitates continual learning by maintaining optimal plasticity–stability balance. Interestingly, a recent metaplasticity-based (cascade) model predicts that the energy cost of old memories is much higher than the cost of new memories [118]. It would be interesting in that context to explore the five CF-suppressing strategies in terms of their energy costs. Intuitively, one would expect generative replay, dual-network (CLS), and constructive algorithms (generating larger networks) to increase, but sparse storage to decrease the energy costs. Metaplasticity increases or decreases energy expenditure depending on conditions [118]. Joint contributions of all five strategies (and their distinct combinations) to the energy efficiency of CF suppression have not yet been systematically explored.

Among the goals for future research is to clarify whether changing brain networks move along low-dimensional manifolds in network parameter space [119], corresponding to optimal low-dimensional Pareto fronts [114] predicted by plasticity–stability–economy trade-offs.



**Figure 1. Pareto optimality of networks for the stability/plasticity trade-off.** Performance space for two competing objectives (plasticity and stability; coded by the two coordinates and the intensity of red and blue colour, respectively). Applying Pareto optimality theory [114] to biological neural networks suggests that networks might have evolved towards an optimal trade-off between plasticity and stability. Pareto optimal networks belong to the so-called Pareto front. They cannot increase their performance for one objective without decreasing it for other objectives. Pareto optimal networks with parameters that lead to the best performance for plasticity (or stability) are specialists (or archetypes [114]) for this objective and occupy one end of the Pareto front. Networks in between perform well for both objectives simultaneously and would be expected to support continual learning. They would avoid catastrophic forgetting (high plasticity/low stability) as well as entrenchment [120] or catastrophic remembering (low plasticity/high stability).



opportunity for architectural solutions for CF [38,104]. Moreover, heterosynaptic interactions including BCM-type rules could more evidently come into play during the unsupervised learning undertaken by spike-timing-dependent plasticity and contribute to CF mitigation [96,104]. Another type of approach of increasing interest is the use of neuromorphic hardware models. Here, the networks are composed of hardware components rather than software, typically involving models replicating ANNs or point neuron SNNs (see [105] for a review), but more recently also evolving to hardware models of neuron dendritic structure and synapses [106]. Hardware implementations of metaplasticity rules have been used to combat CF in such models, for example, cascade-like models of multistate synapses [107] and optimised learning rates via neuromodulation [80]. Mixed hardware and software models are also being implemented now, harnessing the advantages of each and including metaplasticity implementation [108].

### Concluding remarks

Since the earliest days of synaptic plasticity research, there has been extensive interest in how LTP and LTD are regulated by activity-dependent modulation, including by metaplasticity. Similarly, metaplasticity-type rules that regulate learning rates have long made contributions to ANNs needing to solve the stability–plasticity dilemma through balancing these two demands in a time- and space-dependent manner, while also being energy efficient (Box 2). We suggest that there are further potential gains to be achieved by harnessing the diverse properties of metaplasticity phenomena, even though this could complicate further the underlying algorithms. We encourage modellers and roboticists to keep drawing inspiration from the biological processes of plasticity regulation as they make ever-increasing improvements in the network adaptability that underpins continual learning.

### Acknowledgments

Funding support has been received through grants from the Health Research Council of New Zealand (16-597, 18-245) to W. C.A. and from the German Research Foundation (DFG) (JE 528/10-1) to P.J. The authors thank Terence Sejnowski, Igor Farkas, Martin Takac, and Alex Bird for discussions and comments on the manuscript, and Paulina Gronosova for assistance with the figures.

### Declaration of interests

The authors have no interests to declare.

### References

- Grossberg, S. (1987) Competitive learning: from interactive activation to adaptive resonance. *Cogn. Sci.* 11, 23–63
- Abraham, W.C. and Robins, A. (2005) Memory retention – the synaptic stability versus plasticity dilemma. *Trends Neurosci.* 28, 73–78
- McCloskey, M. and Cohen, N.J. (1989) Catastrophic interference in connectionist networks: the sequential learning problem. In *Psychology of Learning and Motivation* (Bower, G.H., ed.), pp. 109–165, Academic Press
- Robins, A. (1995) Catastrophic forgetting, rehearsal and pseudorehearsal. *Connect. Sci.* 7, 123–146
- French, R.M. (1999) Catastrophic forgetting in connectionist networks. *Tr. Cogn. Sci.* 3, 128–135
- Parisi, G.I. et al. (2019) Continual lifelong learning with neural networks: a review. *Neural Netw.* 113, 54–71
- Hadsell, R. et al. (2020) Embracing change: continual learning in deep neural networks. *Tr. Cogn. Sci.* 24, 1028–1040
- Lesort, T. et al. (2020) Continual learning for robotics: definition, framework, learning strategies, opportunities and challenges. *Inf. Fusion* 58, 52–68
- Kudithipudi, D. et al. (2022) Biological underpinnings for lifelong learning machines. *Nat. Mach. Intell.* 4, 196–210
- Abraham, W.C. (2008) Metaplasticity: tuning synapses and networks for plasticity. *Nat. Rev. Neurosci.* 9, 387–399
- Hurtado, J. et al. (2021) Overcoming catastrophic forgetting using sparse coding and meta learning. *IEEE Access* 9, 88279–88290
- Kirkpatrick, J. et al. (2017) Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. U. S. A.* 114, 3521–3526
- Gallardo, J. et al. (2021) Self-supervised training enhances online continual learning. *arXiv* Published online March 25, 2021. <http://dx.doi.org/10.48550/arXiv.2103.14010>
- Carpenter, G.A. (2019) Looking to the future: learning from experience, averting catastrophe. *Neural Netw.* 118, 204–207
- Atkinson, C. et al. (2021) Pseudo-rehearsal: achieving deep reinforcement learning without catastrophic forgetting. *Neurocomputing* 428, 291–307
- Karhunen, J. et al. (2015) Unsupervised deep learning: a short review in *Advances in Independent Component Analysis and Learning Machines* (Bingham, E. et al., eds), pp. 125–142, Academic Press
- Madaan, D. et al. (2022) Representational continuity for unsupervised continual learning. *arXiv* Published online October 13, 2021. <http://dx.doi.org/10.48550/arXiv.2110.06976>
- Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U. S. A.* 79, 2554–2558
- Ratcliff, R. (1990) Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychol. Rev.* 97, 285–308
- Robins, A. (1996) Consolidation in neural networks and in the sleeping brain. *Connect. Sci.* 8, 259–276

### Outstanding questions

The learning rate parameter tends to apply equally to weight increases and decreases, which is a rigid rule that constrains outcomes of any one pass through the network. Would having different but coupled modifiable learning rates for increasing and decreasing connection strengths add a fine-tuning capability to the models that could speed learning processes while also decreasing CF?

Metaplasticity phenomena co-occur in biological neural circuits, yet models using metaplasticity typically do so by implementing it either at individual connections (regularisation) or network-wide (neuromodulation). Would using multiple metaplasticity hyperparameters to enable learning rate changes homosynaptically, heterosynaptically, and network-wide assist with sparsification, modularisation, and associated CF mitigation without incurring too great an energy or processing time cost when learning sequential tasks?

Stimulus salience and novelty push neural circuits metaplastically into a state of readiness for plasticity and learning. For mobile agents traversing an environment across extended times, would implementing this capability, that is, the ability to bias the stability–plasticity trade-off in each direction as the situation demands, improve learning while decreasing forgetting, as well as enhancing energy efficiency?

Given the biological phenomenon of **memory drift**, where different cells come to represent memories over time, are regularisation-type approaches that preserve critical weights too limiting? Or does memory drift occur only in neural circuits that are the equivalent in models of hidden layers?

Neural circuits as well as models need to trade-off stability versus plasticity for continual learning. How do the CF suppression strategies compare with respect to the third consideration of energy efficiency? Will the answers come from neuromorphic hardware where energy can be explicitly measured?

21. Shin, H. *et al.* (2017) Continual learning with deep generative replay. *arXiv* Published online May 24, 2017. <http://dx.doi.org/10.48550/arXiv.1705.08690>
22. Ji, D. and Wilson, M.A. (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107
23. Kamra, N. *et al.* (2017) Deep generative dual memory network for continual learning. *arXiv* Published online October 28, 2017. <http://dx.doi.org/10.48550/arXiv.1710.10368>
24. van de Ven, G.M. *et al.* (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nat. Commun.* 11, 4069
25. Hayes, T.L. *et al.* (2021) Replay in deep learning: current approaches and missing biological elements. *Neural Comput.* 33, 2908–2950
26. McClelland, J.L. *et al.* (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457
27. O'Reilly, R.C. *et al.* (2014) Complementary learning systems. *Cogn. Sci.* 38, 1229–1248
28. Kumaran, D. *et al.* (2016) What learning systems do intelligent agents need? Complementary learning systems theory updated. *Tr. Cogn. Sci.* 20, 512–534
29. Hattori, M. (2014) A biologically inspired dual-network memory model for reduction of catastrophic forgetting. *Neurocomputing* 134, 262–268
30. McClelland, J.L. *et al.* (2020) Integration of new information in memory: new insights from a complementary learning systems perspective. *Philos. Trans. R. Soc. B Biol. Sci.* 375, 20190637
31. Parisi, G.I. *et al.* (2018) On the role of neurogenesis in overcoming catastrophic forgetting. *arXiv* Published online November 6, 2018. <http://dx.doi.org/10.48550/arXiv.1811.02113>
32. Rolls, E.T. and Treves, A. (1990) The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Netw. Comput. Neural Syst.* 1, 407–421
33. Ahmad, S. and Scheinkman, L. (2019) How can we be so dense? The benefits of using highly sparse representations. *arXiv* Published online March 27, 2019. <http://dx.doi.org/10.48550/arXiv.1903.11257>
34. Maneschi, L. *et al.* (2021) SpaRCe: improved learning of reservoir computing systems through sparse representations. *IEEE Trans. Neural Netw. Learn. Syst.* Published online August 16, 2021. <https://doi.org/10.1109/tnnls.2021.3102378>
35. Ellefsen, K.O. *et al.* (2015) Neural modularity helps organisms evolve to learn new skills without forgetting old skills. *PLoS Comput. Biol.* 11, e1004128
36. Spanne, A. and Jörntell, H. (2015) Questioning the role of sparse coding in the brain. *Trends Neurosci.* 38, 417–427
37. Feng, Y. and Brunel, N. (2021) Storage capacity of networks with discrete synapses and sparsely encoded memories. *arXiv* Published online December 13, 2021. <http://dx.doi.org/10.48550/arXiv.2112.06711>
38. Grewal, K. *et al.* (2021) Going beyond the point neuron: active dendrites and sparse representations for continual learning. *bioRxiv* Published online October 26, 2021. <https://doi.org/10.1101/2021.10.25.465651>
39. Iyer, A. *et al.* (2021) Avoiding catastrophe: active dendrites enable multi-task learning in dynamic environments. *arXiv* Published online December 31, 2021. <http://dx.doi.org/10.48550/arXiv.2201.00042>
40. Hainmueller, T. and Bartos, M. (2018) Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature* 558, 292–296
41. Leutgeb, J.K. *et al.* (2007) Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* 315, 961–966
42. Wiskott, L. *et al.* (2006) A functional hypothesis for adult hippocampal neurogenesis: avoidance of catastrophic interference in the dentate gyrus. *Hippocampus* 16, 329–343
43. Fahlman, S. and Lebiere, C. (1989) The cascade-correlation learning architecture. *Adv. Neural Inform. Process. Syst.* 2, 524–532
44. Carpenter, G.A. *et al.* (1989) Invariant recognition of cluttered scenes by a self-organizing ART architecture: CORT-X boundary segmentation. *Neural Netw.* 2, 169–181
45. Tsuda, B. *et al.* (2020) A modeling framework for adaptive lifelong learning with transfer and savings through gating in the prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 117, 29872–29882
46. Franco, L. *et al.*, eds (2009) *Constructive Neural Networks (Studies in Computational Intelligence v. 258)*, Springer
47. Zemouri, R.A. *et al.* (2019) A new growing pruning deep learning neural network algorithm (GP-DLNN). *Neural Comput. Appl.* 32, 18143–18159
48. Rusu, A.R. *et al.* (2016) Progressive neural networks. *arXiv* Published online June 15, 2016. <http://dx.doi.org/10.48550/arXiv.1606.04671>
49. Liu, C. *et al.* (2018) *Progressive neural architecture search. Proceedings of the European Conference on Computer Vision (ECCV)* pp. 19–34
50. Knoblauch, A. *et al.* (2014) Structural synaptic plasticity has high memory capacity and can explain graded amnesia, catastrophic forgetting, and the spacing effect. *PLoS ONE* 9, e96485
51. Ming, G. and Song, H. (2011) Adult neurogenesis in the mammalian brain: significant answers and significant questions. *Neuron* 70, 687–702
52. Draelos, T.J. *et al.* (2017) *Neurogenesis deep learning: extending deep networks to accommodate new classes. International Joint Conference on Neural Networks (IJCNN)* pp. 526–533
53. Thrun, S. and Pratt, L. (1998) Learning to learn: introduction and overview. In *Learning to Learn* (Thrun, S. and Pratt, L., eds), pp. 3–17, Springer
54. Finn, C. *et al.* (2017) *Model-agnostic meta-learning for fast adaptation of deep networks. Proceedings of the 34th International Conference on Machine Learning* pp. 1126–1135
55. Javed, K. and White, M. (2019) Meta-learning representations for continual learning. *Adv. Neural Inform. Process. Syst.* 32, 1–11
56. Hospedales, T. *et al.* (2020) Meta-learning in neural networks: a survey. *arXiv* Published online April 11, 2020. <http://dx.doi.org/10.48550/arXiv.2004.05439>
57. Beaulieu, S. *et al.* (2020) Learning to continually learn. *arXiv* Published online February 21, 2020. <http://dx.doi.org/10.48550/arXiv.2002.09571>
58. Tsyulmankov, D. *et al.* (2021) Meta-learning synaptic plasticity and memory addressing for continual familiarity detection. *Neuron* 110, 1–14
59. Hopfield, J.J. *et al.* (1983) 'Unlearning' has a stabilizing effect in collective memories. *Nature* 304, 158–159
60. Robins, A. and McCallum, S. (1999) The consolidation of learning during sleep: comparing the pseudorehearsal and unlearning accounts. *Neural Netw.* 12, 1191–1206
61. Sandberg, A. *et al.* (2000) A palimpsest memory based on an incremental Bayesian learning rule. *Neurocomputing* 32, 987–994
62. Amit, D.J. and Fusi, S. (1994) Learning in neural networks with material synapses. *Neural Comput.* 6, 957–982
63. Abraham, W.C. and Bear, M.F. (1996) Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci.* 19, 126–130
64. Hulme, S.R. *et al.* (2013) Emerging roles of metaplasticity in behaviour and disease. *Trends Neurosci.* 36, 353–362
65. Redondo, R.L. and Morris, R.G.M. (2011) Making memories last: the synaptic tagging and capture hypothesis. *Nat. Rev. Neurosci.* 12, 17–30
66. Hulme, S.R. *et al.* (2014) Mechanisms of heterosynaptic metaplasticity. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 1633
67. Bienenstock, E.L. *et al.* (1982) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* 2, 32–48
68. Fusi, S. *et al.* (2005) Cascade models of synaptically stored memories. *Neuron* 45, 599–611
69. Benna, M.K. and Fusi, S. (2016) Computational principles of synaptic memory consolidation. *Nat. Neurosci.* 19, 1697–1706
70. Villa, K.L. *et al.* (2016) Inhibitory synapses are repeatedly assembled and removed at persistent sites in vivo. *Neuron* 89, 756–769
71. Zenke, F. *et al.* (2017) Continual learning through synaptic intelligence. *Proc. Mach. Learn. Res.* 70, 3987–3995
72. Aljundi, R. *et al.* (2018) *Memory aware synapses: learning what (not) to forget. Proceedings of the European Conference on Computer Vision (ECCV)* pp. 139–154

73. Laborieux, A. *et al.* (2021) Synaptic metaplasticity in binarized neural networks. *Nat. Commun.* 12, 2549
74. Zohora, F.T. *et al.* (2021) *MetaplasticNet: architecture with probabilistic metaplastic synapses for continual learning*. 2021 IEEE International Symposium on Circuits and Systems (ISCAS) pp. 1–5
75. Bhalwankar, R. and Treur, J. (2020) *A second-order adaptive network model for learner-controlled mental model learning processes*. International Conference on Complex Networks and Their Applications pp. 245–259
76. Bird, A.D. *et al.* (2021) Dendritic normalisation improves learning in sparsely connected artificial neural networks. *PLoS Comput. Biol.* 17, e1009202
77. Masse, N.Y. *et al.* (2018) Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *Proc. Natl. Acad. Sci. U. S. A.* 115, E10467–E10475
78. Schug, S. *et al.* (2021) Presynaptic stochasticity improves energy efficiency and helps alleviate the stability-plasticity dilemma. *eLife* 10, e69884
79. Leibold, C. and Kempster, R. (2008) Sparseness constrains the prolongation of memory lifetime via synaptic metaplasticity. *Cereb. Cortex* 18, 67–77
80. Bohnstingl, T. *et al.* (2019) Neuromorphic hardware learns to learn. *Front. Neurosci.* 13, 483
81. Chaudhry, A. *et al.* (2018) Efficient lifelong learning with a-gem. *arXiv* Published online December 2, 2018. <http://dx.doi.org/10.48550/arXiv.1812.00420>
82. Wu, C. *et al.* (2018) Memory replay GANs: learning to generate new categories without forgetting. *Adv. Neural Inform. Process. Syst.* 31, 1–11
83. Ramapuram, J. *et al.* (2020) Lifelong generative modeling. *Neurocomputing* 404, 381–400
84. Furlanello, T. *et al.* (2016) Active long term memory networks. *arXiv* Published online June 7, 2016. <http://dx.doi.org/10.48550/arXiv.1606.02355>
85. Li, Z. and Hoiem, D. (2017) Learning without forgetting. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 2935–2947
86. Maltoni, D. and Lomonaco, V. (2019) Continuous learning in single-incremental-task scenarios. *Neural Netw.* 116, 56–73
87. Nguyen, C.V. *et al.* (2017) Variational continual learning. *arXiv* Published online October 29, 2017. <http://dx.doi.org/10.48550/arXiv.1710.10628>
88. Parisi, G.I. *et al.* (2018) Lifelong learning of spatiotemporal representations with dual-memory recurrent self-organization. *Front. Neurobot.* 12, 78
89. Flumer, M. *et al.* (2018) Learning to learn without forgetting by maximizing transfer and minimizing interference. *arXiv* Published online October 29, 2018. <http://dx.doi.org/10.48550/arXiv.1810.11910>
90. Milstein, A.D. *et al.* (2021) Bidirectional synaptic plasticity rapidly modifies hippocampal representations. *eLife* 10, e73046
91. Zhang, W. and Linden, D.J. (2003) The other side of the engram: experience-driven changes in neuronal intrinsic excitability. *Nat. Rev. Neurosci.* 4, 885–900
92. Ge, S. *et al.* (2007) A critical period for enhanced synaptic plasticity in newly generated neurons of the adult brain. *Neuron* 54, 559–566
93. Hulme, S.R. *et al.* (2012) Calcium-dependent but action potential-independent BCM-like metaplasticity in the hippocampus. *J. Neurosci.* 32, 6785–6794
94. Benuskova, L. *et al.* (1994) Dynamic synaptic modification threshold: computational model of experience-dependent plasticity in adult rat barrel cortex. *Proc. Natl. Acad. Sci. U. S. A.* 91, 4791–4795
95. Jedlicka, P. *et al.* (2015) A voltage-based STDP rule combined with fast BCM-like metaplasticity accounts for LTP and concurrent “heterosynaptic” LTD in the dentate gyrus in vivo. *PLoS Comput. Biol.* 11, e1004588
96. Halvagal, M.S. and Zenke, F. (2022) The combination of Hebbian and predictive plasticity learns invariant object representations in deep sensory networks. *bioRxiv* Published online March 19, 2022. <https://doi.org/10.1101/2022.03.17.484712>
97. Harvey, C.D. and Svoboda, K. (2007) Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* 450, 1195–1200
98. Pöpper, M. *et al.* (2011) Synaptic tagging, evaluation of memories, and the distal reward problem. *Learn. Mem.* 18, 58–70
99. Luboinski, J. and Tetzlaff, C. (2021) Memory consolidation and improvement by synaptic tagging and capture in recurrent neural networks. *Comm. Biol.* 4, 1–17
100. Grossberg, S. (1976) Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biol. Cybern.* 23, 121–134
101. Alired, J.M. and Roy, K. (2020) Controlled forgetting: targeted stimulation and dopaminergic plasticity modulation for unsupervised lifelong learning in spiking neural networks. *Front. Neurosci.* 14, 7
102. van Ments, L. and Treur, J. (2022) Disturbed by flashbacks: a controlled adaptive network model addressing mental models for flashbacks from PTSD. In *Mental Models and Their Dynamics, Adaptation, and Control: A Self-Modeling Network Modeling Approach* (Treur, J. and Van Ments, L., eds), pp. 99–116, Springer
103. Fachechi, A. *et al.* (2019) Dreaming neural networks: forgetting spurious memories and reinforcing pure ones. *Neural Netw.* 112, 24–40
104. Acharya, J. *et al.* (2022) Dendritic computing: branching deeper into machine learning. *Neuroscience* 489, 275–289
105. Tang, J. *et al.* (2019) Bridging biological and artificial neural networks with emerging neuromorphic devices: fundamentals, progress, and challenges. *Adv. Mater.* 31, 1902761
106. Kaiser, J. *et al.* (2022) Emulating dendritic computing paradigms on analog neuromorphic hardware. *Neuroscience* 489, 290–300
107. Zohora, F.T. *et al.* (2020) *Metaplasticity in multistate memristor synaptic networks*. 2020 IEEE International Symposium on Circuits and Systems (ISCAS) pp. 1–5
108. Li, Y. *et al.* (2022) Mixed-precision continual learning based on computational resistance random access memory. *Adv. Intell. Syst.* Published online May 17, 2022. <https://doi.org/10.1002/aisy.202200026>
109. Farashahi, S. *et al.* (2017) Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron* 94, 401–414.e406
110. Montgomery, J.M. and Madison, D.V. (2004) Discrete synaptic states define a major mechanism of synapse plasticity. *Trends Neurosci.* 27, 744–750
111. Ward, B. *et al.* (2006) State-dependent mechanisms of LTP expression revealed by optical quantal analysis. *Neuron* 52, 649–661
112. Laughlin, S.B. and Sejnowski, T.J. (2003) Communication in neuronal networks. *Science* 301, 1870–1874
113. Yu, L. and Yu, Y. (2017) Energy-efficient neural information processing in individual neurons and neuronal networks. *J. Neurosci. Res.* 95, 2253–2266
114. Shoval, O. *et al.* (2012) Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* 336, 1157–1160
115. Pallasdies, F. *et al.* (2021) Neural optimization: understanding trade-offs with Pareto theory. *Curr. Opin. Neurobiol.* 71, 84–91
116. Verbeke, P. and Verguts, T. (2019) Learning to synchronize: how biological agents can couple neural task modules for dealing with the stability-plasticity dilemma. *PLoS Comput. Biol.* 15, e1006604
117. Jin, Y. and Sendhoff, B. (2006) *Alleviating catastrophic forgetting via multi-objective learning*. The 2006 IEEE International Joint Conference on Neural Network Proceedings pp. 3335–3342
118. Karbowski, J. (2021) Energetics of stochastic BCM type synaptic plasticity and storing of accurate information. *J. Computat. Neurosci.* 49, 71–106
119. Mishra, P. and Narayanan, R. (2021) Stable continual learning through structured multiscale plasticity manifolds. *Curr. Opin. Neurobiol.* 70, 51–63
120. Mermillod, M. *et al.* (2013) The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects. *Front. Psychol.* 4, 504

## **Update**

### **Trends in Neurosciences**

Volume 46, Issue 10, October 2023, Page 893–894

DOI: <https://doi.org/10.1016/j.tins.2023.07.008>

## Correction

# Contributions by metaplasticity to solving the Catastrophic Forgetting Problem

Peter Jedlicka,\* Matus Tomko, Anthony Robins, and Wickliffe C. Abraham\*

\*Correspondence: [peter.jedlicka@informatik.med.uni-giessen.de](mailto:peter.jedlicka@informatik.med.uni-giessen.de) (P. Jedlicka) and [cliff.abraham@otago.ac.nz](mailto:cliff.abraham@otago.ac.nz) (W.C. Abraham).

DOI of original article: <https://doi.org/10.1016/j.tins.2022.06.002>

(Trends in Neurosciences, 45, 656–666, 2022)

In the print and online PDF versions of the article, references [15–49] were unfortunately missing due to a production error during late stages of article preparation. The references have been reinstated in the online version of the article, and are listed below. The Publisher apologizes to all for the inconvenience.

## References

- Atkinson, C. *et al.* (2021) Pseudo-rehearsal: achieving deep reinforcement learning without catastrophic forgetting. *Neurocomputing* 428, 291–307
- Karhunen, J. *et al.* (2015) Unsupervised deep learning: a short review. In *Advances in Independent Component Analysis and Learning Machines* (Bingham, E. *et al.*, eds), pp. 125–142, Academic Press
- Madaan, D. *et al.* (2022) Representational continuity for unsupervised continual learning. *arXiv* Published online October 13, 2021. <http://dx.doi.org/10.48550/arXiv.2110.06976>
- Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U. S. A.* 79, 2554–2558
- Ratcliff, R. (1990) Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychol. Rev.* 97, 285–308
- Robins, A. (1996) Consolidation in neural networks and in the sleeping brain. *Connect. Sci.* 8, 259–276
- Shin, H. *et al.* (2017) Continual learning with deep generative replay. *arXiv* Published online May 24, 2017. <http://dx.doi.org/10.48550/arXiv.1705.08690>
- Ji, D. and Wilson, M.A. (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107
- Kamra, N. *et al.* (2017) Deep generative dual memory network for continual learning. *arXiv* Published online October 28, 2017. <http://dx.doi.org/10.48550/arXiv.1710.10368>
- van de Ven, G.M. *et al.* (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nat. Commun.* 11, 4069
- Hayes, T.L. *et al.* (2021) Replay in deep learning: current approaches and missing biological elements. *Neural Comput.* 33, 2908–2950
- McClelland, J.L. *et al.* (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457
- O'Reilly, R.C. *et al.* (2014) Complementary learning systems. *Cogn. Sci.* 38, 1229–1248
- Kumaran, D. *et al.* (2016) What learning systems do intelligent agents need? Complementary learning systems theory updated. *Tr. Cogn. Sci.* 20, 512–534
- Hattori, M. (2014) A biologically inspired dual-network memory model for reduction of catastrophic forgetting. *Neurocomputing* 134, 262–268
- McClelland, J.L. *et al.* (2020) Integration of new information in memory: new insights from a complementary learning systems perspective. *Philos. Trans. R. Soc. B Biol. Sci.* 375, 20190637
- Parisi, G.I. *et al.* (2018) On the role of neurogenesis in overcoming catastrophic forgetting. *arXiv* Published online November 6, 2018. <http://dx.doi.org/10.48550/arXiv.1811.02113>
- Rolls, E.T. and Treves, A. (1990) The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Netw. Comput. Neural Syst.* 1, 407–421
- Ahmad, S. and Scheinkman, L. (2019) How can we be so dense? The benefits of using highly sparse representations. *arXiv* Published online March 27, 2019. <http://dx.doi.org/10.48550/arXiv.1903.11257>
- Manneschi, L. *et al.* (2021) SpaRCe: improved learning of reservoir computing systems through sparse representations. *IEEE Trans. Neural Netw. Learn. Syst.* Published online August 16, 2021. <https://doi.org/10.1109/tnls.2021.3102378>
- Ellefsen, K.O. *et al.* (2015) Neural modularity helps organisms evolve to learn new skills without forgetting old skills. *PLoS Comput. Biol.* 11, e1004128
- Spanne, A. and Jörentell, H. (2015) Questioning the role of sparse coding in the brain. *Trends Neurosci.* 38, 417–427
- Feng, Y. and Brunel, N. (2021) Storage capacity of networks with discrete synapses and sparsely encoded memories. *arXiv* Published online December 13, 2021. <http://dx.doi.org/10.48550/arXiv.2112.06711>
- Grewal, K. *et al.* (2021) Going beyond the point neuron: active dendrites and sparse representations for continual learning. *bioRxiv* Published online October 26, 2021. <https://doi.org/10.1101/2021.10.25.465651>
- Iyer, A. *et al.* (2021) Avoiding catastrophe: active dendrites enable multi-task learning in dynamic environments. *arXiv* Published online December 31, 2021. <http://dx.doi.org/10.48550/arXiv.2201.00042>
- Hainmueller, T. and Bartos, M. (2018) Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature* 558, 292–296
- Leutgeb, J.K. *et al.* (2007) Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* 315, 961–966
- Wiskott, L. *et al.* (2006) A functional hypothesis for adult hippocampal neurogenesis: avoidance of catastrophic interference in the dentate gyrus. *Hippocampus* 16, 329–343
- Fahlman, S. and Lebiere, C. (1989) The cascade-correlation learning architecture. *Adv. Neural Inform. Process. Syst.* 2, 524–532
- Carpenter, G.A. *et al.* (1989) Invariant recognition of cluttered scenes by a self-organizing ART architecture: CORT-X boundary segmentation. *Neural Netw.* 2, 169–181
- Tsuda, B. *et al.* (2020) A modeling framework for adaptive lifelong learning with transfer and savings through gating in the prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 117, 29872–29882





46. Franco, L. *et al.*, eds (2009) *Constructive Neural Networks (Studies in Computational Intelligence v. 258)*, Springer
47. Zemouri, R.A. *et al.* (2019) A new growing pruning deep learning neural network algorithm (GP-DLNN). *Neural. Comput. Appl.* 32, 18143–18159
48. Rusu, A.R. *et al.* (2016) Progressive neural networks. *arXiv* Published online June 15, 2016. <http://dx.doi.org/10.48550/arXiv.1606.04671>
49. Liu, C. *et al.* (2018) Progressive neural architecture search. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 19–34