UNIVERSITÄT
BERN

Graduate School for Cellular and Biomedical Sciences
University of Bern

A computational model for continual learning and synaptic consolidation

PhD Thesis submitted by

Pascal Leimer

Bettlach, SO
Switzerland

for the degree of

PhD in Neuroscience

Supervisor
Prof. Dr. Walter Senn
Department of Physiology
University of Bern, Switzerland

Co-advisor
Prof. Dr. Wulfram Gerstner
Brain Mind Institute
École Polytechnique Fédérale de Lausanne, Switzerland

Accepted by the Faculty of Medicine, the Faculty of Science and the Vetsuisse Faculty of the University of Bern at the request of the Graduate School for Cellular and Biomedical Sciences

Bern, Dean of the Faculty of Medicine

Bern, Dean of the Faculty of Science

Bern, Dean of the Vetsuisse Faculty Bern

Acknowledgment

My PhD studies were an exciting and enriching but sometimes also a frustrating project which would not have been possible for me to master alone. I would like acknowledge and express my gratitude towards all people who helped and supported me during this time.
First and foremost, I would like to thank my supervisor Walter Senn. He gave me the opportunity to dive into the fascinating field of neuroscience and was a precious help at all times throughout my PhD. His insightful advices strongly contributed to the success of my project. I would like to extend a thank you to my fellow students Elena Kreutzer, Dominik Spicher, Kristin Völk and Chang Zhao. The time we spent together at work and the many extracurricular activities made me enjoy every day. Without them something would be missing in my life. A thank you goes to João Sacramento who has always encouraged me with thoughtful advices, and Jakob Jordan who helped me to polish my writings. Furthermore, special thanks goes to all current and former members of the Senn lab who contributed to a scientific and social exchange during my PhD.
Most importantly, I would like to thank my fiancée Anna Holzbecher. She encourages me in all my endeavors and supports me with her adoring nature in all situations.

Abstract

How humans are able to learn and memorize is a long-standing question in science. Much progress has been achieved in recent decades to answer this question but the are still many open problems. One of these problems refers to the human ability to learn several tasks in sequence without forgetting.
In neuronal networks learning can interfere with pre-existing memories when the network is engaged in continual learning. The interference is particularly pronounced if, for instance, similar sensory stimuli require different responses depending on the context. Unlike in humans, this can lead to a memory loss termed catastrophic forgetting. To avoid interference and its fatal consequences, only a subset of synaptic weights should be consolidated. In this work we propose as computational model which performs selective consolidation by incorporating the synaptic tagging and capture hypothesis. This hypothesis, well grounded by experimental evidences, claims that synaptic consolidation requires both a synaptic-specific tag and diffusible plasticity-related proteins. We show that synaptic tagging and capture can be modeled by two classes of synaptic processes acting on different time scales. The two classes, characterized whether protein synthesis is required, are represented in our model by two synaptic components interacting with each other.
With our approach we demonstrate that synaptic consolidation can not only diminishes the problem of catastrophic forgetting during continual learning but also enables fast learning through strongly changing synaptic strengths during the early phase of long-term potentiation. The model reproduces various experimental observations on synaptic tagging and cross-tagging. It also explains why learning in psychophysical experiments is hampered when different types of stimuli are randomly intermixed.

Contents

Acknowledgment **V**

Abstract **VII**

Contents **VIII**

List of Figures **XI**

1 Introduction **1**

1.1 General introduction 1
1.1.1 Neuroscience 2
1.1.2 Computational neuroscience 3
1.1.3 Learning and memory 4
1.2 Neurobiological basis of synaptic plasticity 5
1.2.1 Historical notes on synaptic plasticity 5
1.2.2 Synaptic plasticity and time scales 8
1.2.3 Consolidation 9
1.3 Single neuron model 11
1.4 Learning rule 14
1.5 Reinforcement learning 17
1.6 Perceptual learning and roving 18
1.7 Catastrophic forgetting and continual learning 20
1.8 Computational models 21
1.8.1 Models of continual learning 21
1.8.2 Models of tagging and consolidation 23

1.8.3 Models explaining the roving phenomenon 24

2 Hypothesis and Aim **27**

3 Results **31**

4 Discussion **65**
4.1 Key findings 65
4.2 Further insights 66
4.2.1 Catastrophic forgetting 66
4.2.2 Small pattern distinctions 66
4.2.3 Fast learning 67
4.3 Biological relevance 68
4.3.1 Protein dependent and independent processes 68
4.3.2 Memory consolidation during breaks 69
4.3.3 Roving 69
4.4 Comparison with other models 69
4.5 Model predictions 71
4.6 Future work & model improvements 72
4.6.1 Towards more biological plausibility 72
4.6.2 New directions 73

Bibliography **75**

List of Figures

1 Neuron with cell body, dendrites and axon 3
2 Tagging and capture in dependence of stimulation strength 10
3 Perceptron neuron model 13
4 Neuron model with different dendritic and axonal ininital segment voltage 15
5 Bisection stimuli 19
6 Strategies to avoid catastrophic forgetting. 23
7 Classifying similar input patterns in different contexts fails with a 1-component plasticity model but succeeds with a 2-component model . . 38
8 Learning similar associations in sequence causes catastrophic forgetting with a classical 1-component rule 40
9 Competition-agnostic plasticity allows for fast learning without saturation 42
10 Selecting informative events for consolidation 44
11 Model accounts for synaptic tagging *in vivo* 46
12 The model also accounts for LTD tagging and cross-tagging 48
13 2-component rule accounts for roving 50
S1 Competition-agnostic plasticity create non-vanishing learning signals . . 60

CHAPTER 1

Introduction

Curiosity is always the first step in solving a problem.

— Galileo Galilei

1.1 General introduction

On a hot summer day in Bern crowds of people meet at the river Aare to stroll along the banks, to take a refreshing swim or lie down in the sun to reflect on the impressions of the day. While we do not consider any of these actions particularly challenging, all of them would be impossible to perform without the human ability to learn and memorize.

Without this ability, we would be restricted to innate behavior, incapable of active decision making and acquiring new knowledge. Some animal species without the physiological prerequisites to learn are able to swim, for example sponges, but they are strongly restricted in their diversity of actions.

This thesis addresses the question how humans and animals learn and memorize from a computational perspective. In particular, we focus on the problem of continual learning, i.e. learning in multiple subsequent contexts. In each context, a similar but not identical task has to be learned. Given the natural occurrence of continual learning

situations, the underlying learning mechanism is of particular interest to be studied.

The desire to understand how we learn is not only motivated by the quenching thirst for knowledge but can also help in the future to improve teaching methods, the treatment of learning disabilities and artificial intelligence. The study of learning and memory is therefore one of the core topics of neuroscience. In the next subsections we give a brief overview of how it fits in the research spectrum of neuroscience in general and computational neuroscience in particular. The subsequent parts of the introduction aim to provide the prior knowledge necessary to understand the scientific results of our work. A detailed description of the aim and hypothesis of our work is discussed in chapter 2, followed by the presentation of the main scientific results in chapter 3. Chapter 4 concludes this thesis with an outlook and proposals for future studies.

1.1.1 Neuroscience

Neuroscience is the scientific research area in which the structure and function of the nervous system is studied. It is traditionally classified as a subfield of biology but nowadays combines many more disciplines such as psychology, computer science, mathematics, philosophy and physics. Questions asked by neuroscientists can span a broad range of interests, for example:

- What is the biological basis of consciousness?
- Why do we dream?
- What is the neural basis of mental diseases?
- How are our memories stored and retrieved?

In vertebrates, the nervous system is composed of two main parts, the central nervous system and the peripheral nervous system. The most complex part is the brain which, together with the spinal cord, forms the central nervous system. The primary cell types found in the brain are neurons and supportive glial cells. The anatomical structure of neurons is diverse but typically consists of a cell body (soma), dendrites and an axon.

Neurons transmit information by conducting electrical and chemical signals. To process information, it is essential that signals are transferred from one neuron to another. This happens at synapses, a structure where the signal-emitting neuron (presynaptic neuron) is close to the membrane of the target neuron (postsynaptic neuron). Two

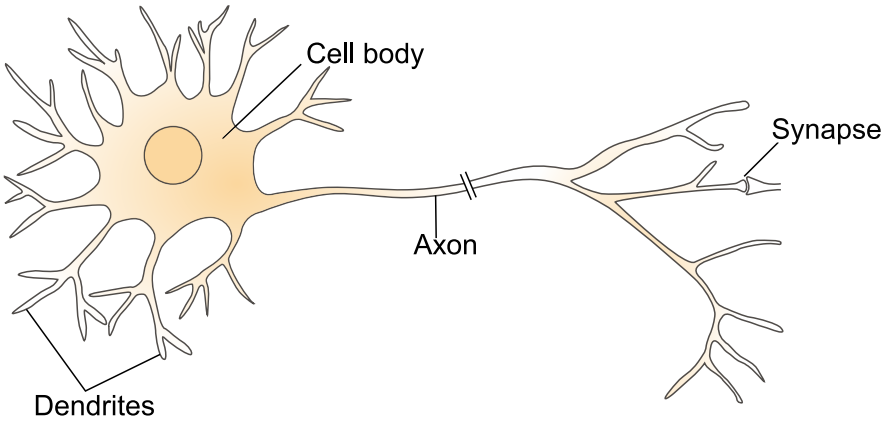

Figure 1: Various types of neurons exist. Most of them have a cell body (soma), dendrites and an axon.

fundamentally different types of synapses are known: chemical synapses, where signal transmission happens by releasing and binding of neurotransmitter, and gap junctions, which are capable of passing an electric current. Often synapses connect the axon of the presynaptic neuron with a dendritic spine of the postsynaptic neuron, though other points of contact also exist (Pereda, 2014). The human brain contains around 10^{11} neurons with a total of 10^{14} synaptic connections (Squire and Kandel, 2009). On average, these are on the order of 10^3 synapses per neuron, though this number changes strongly between different neuron types and varies with the individual's age.

The complexity of the brain forces scientist to study processes on a wide range of length scales. On the smallest scale, the molecular level, the interaction of ions, proteins and neurotransmitters is studied. On a larger scale, scientists investigate synapses and transmission of action potentials in neurons. Neurons form assemblies which play an important role for sensory processing, learning and memory. Here, knowledge about the connectivity of neurons and integration of neural signals can shed light on functioning of the brain. At the largest scale, larger parts of the brain and the interaction of brain areas are studied to understand neurological diseases, behavior or consciousness.

1.1.2 Computational neuroscience

Computational neuroscience is an interdisciplinary field which focuses on mathematical models and theories of neural structures and processes to address neuroscientific

questions. It incorporates and describes results of experiments and makes predictions for new experiments.

Models can describe a wide range of structures, e.g. synapses, neurons or neural networks. These models are often evaluated in computer simulations due to their complexity. Although models represent the nature in a simplified and incomplete way they are essential to detect relations which would not be visible by a looking either at the whole system or at isolated structures. They help us to formulate new hypotheses and to strengthen the scientific knowledge.

To avoid the risk of oversimplification many scientists consider so-called bottom-up models, i.e. models which describe a small system in a precise way and result in predictions for a larger system. In contrast, top-down models start from a higher level principle such as reward maximization from which one can derive predictions about the underlying mechanisms (Gerstner et al., 2012).

Both approaches have in common that they bridge the gap between multiple length and time scales while experiments, restricted by technical capabilities, often only look at processes at one particular scale. With models, we can combine findings and predictions from different levels of abstractions.

1.1.3 Learning and memory

Learning is the process of acquiring knowledge and skills while memory is the expression of what has been acquired (Kazdin, 2000).

Over the last decades, scientist have discovered many different memory systems in the brain. They differ from each other by having different time spans and by encoding different kinds of information. Roughly, one distinguishes between short-term memory and long-term memory. Short-term memory, albeit imprecisely defined, normally refers to memories which are only maintained temporarily. This includes the working memory, which has a limited capacity of around four items and a life time on the order of a few seconds (Baddeley, 2003). Short-term memory can also refer to memories which last for several hours. Since memories can be perturbed during this time they need to be consolidated to become stable (Squire and Kandel, 2009). Memories which last hours to months belong to the category of long-term memory (McGaugh, 2000).

Long-term memory can be broken down into explicit memory and implicit memory. Explicit memory, or declarative memory, is available to our consciousness. It can be further divided into two types: semantic memory, which stores general knowledge like

facts, ideas or concepts, and episodic memory keeping track of specific personal experiences. In contrast to the explicit memory is the implicit or nondeclarative memory, acquired and used unconsciously. One form of the implicit memory is the procedural memory, i.e. the encoding of certain motor skills (McClelland et al., 1995).

An important structure for explicit memories is the hippocampus (Eichenbaum, 2001) while implicit memories are hippocampus independent (Squire et al., 1992). Invertebrate animals likely only have implicit memories due to the lack of a hippocampus (Squire and Kandel, 2009). The hippocampus is not only important for learning and memory but also for spatial navigation (O'Keefe and Nadel, 1978) and the reward system (O'Neil et al., 2015).

How memories are formed is one of the big questions in biology. There is strong evidence that learning happens by changing the synaptic strength (Takeuchi et al., 2014) but many questions remain to be answered. Do short-term and long-term memory act independently or is information transformed from short-term memory to long-term memory? What triggers the formation of stable memories? Why do we forget certain information?

These questions have fascinated me during my PhD time. On the following pages I will present some answers I found. The core result, a computational model for learning and memory, contributes to a better understanding of learning and memory. The model, which performs learning based on reward-signals, uses experimental data from animals and humans and makes new testable predictions about learning and its physiological basis in neural circuits.

1.2 Neurobiological basis of synaptic plasticity

1.2.1 Historical notes on synaptic plasticity

Compared to other field of science, neuroscience is a rather young field of research with little comprehension till the late 19th century. Exemplary is the reticular theory proposed by German anatomist Joseph von Gerlach in 1871. He claims that everything in the nervous system is a single continuous network (von Gerlach, 1871). The theory was well accepted by the scientific community and strongly supported by the later Nobel laureate Camillo Golgi (De Carlos and Borrell, 2007).

The reticular theory was disputed among others by the Spanish pathologist Santiago

Ramón y Cajal. In his lecture delivered to British Royal Society in 1894, he showed that the nervous system is made up of many separate neurons (Jones, 1994). Besides, he suggested that neurons communicate with each other over special junctions, which are nowadays called synapses (Jones, 1994). For this research Santiago Ramón y Cajal received the Nobel Prize in Physiology or Medicine. The existence of synapses could only be proven in the 1950s when individual synapses could be observed with electron microscopes (De Robertis and Bennett, 1955; Palay, 1956).

A major impact on neuroscience research had the publication of the book *The Organization of Behavior* by the Canadian psychologist Donald Olding Hebb in 1949 (Hebb, 1949). Hebb postulated that if two neurons are near enough and the activity of one neuron leads to repeated or persistent firing of the other neuron, that some changes take place in one or both neurons such that the efficiency between the two neurons is increased. The theory often gets summarized by the phrase 'What wires together, fires together', although the phrase is imprecise since it misses the temporal aspects of Hebb's theory. Hebb did not have any experimental evidence for his postulate; it was a pure theoretical conclusion from the consideration that such a mechanism would stabilize specific neuronal activity patterns in the brain. We know nowadays that many synapses show Hebbian-like plasticity but also synapses with different characteristics can be found, so-called anti-Hebbian (Tzounopoulos et al., 2004) and non-Hebbian synapses (Kossel et al., 1990).

Two decades after the influential publication of Hebb, Terje Lømo and Timothy Bliss discovered long-term potentiation (LTP) (Bliss and Lømo, 1973). They stimulated presynaptic neurons in the perforant pathway of rabbits and recorded the response in postsynaptic neurons of the dentate gyrus. As expected, a single electrical pulse applied to the presynaptic neuron caused an excitatory postsynaptic potential (EPSP) in the postsynaptic neuron. New was the observation that the EPSP could be enhanced if he first delivered a high-frequency train of stimuli to the presynaptic neurons. This is a proof that the strength of a synapse can be modified, i.e. synapses are plastic. Nowadays we know that among others the strength of the stimuli and the ability to synthesize proteins determine whether the EPSP is enhanced persistently, temporarily or not at all (Frey and Morris, 1997, 1998a,b). While an EPSP makes a neuron more likely to generate a spike, also inhibitory postsynaptic potentials (IPSPs) can be observed, which make a neuron less likely to generate an action potential (Curtis and Eccles, 1960).

The counterpart of LTP, long-term depression (LTD), was discovered a few years later. LTD is an activity-dependent reduction in the strength of synapses and can be induced by low-frequency stimulation (Dunwiddie and Lynch, 1978; Dudek and Bear, 1992).

Hebb's postulate was extended by the observed process of spike-timing-dependent plasticity (STDP). STDP does not only depend on the correlation between pre- and post-synaptic activity but also on the respective timing. It was first postulated in a modeling study (Gerstner et al., 1996) and experimentally confirmed in neocortical and hippocampal neurons during the 1990s (Markram et al., 1997; Bi and Poo, 1998). In STDP experiments, two connected neurons get stimulated with varying interstimuli times. As predicted by Hebb, LTP is observed if the post-synaptic neuron fires after the pre-synaptic neuron. Contrary, and not mentioned by Hebb, is the observation of LTD if the pre-synaptic neuron fires after the post-synaptic neuron. The shorter the time interval between pre- and post-synaptic activity is, the stronger the effect of LTP and LTD, respectively. However, not all cell types show the same response. In other experiments using the same stimulation protocol but done in cerebellum-like structures, an inverted STDP curve can be found (Bell et al., 1997).

Synaptic plasticity as the basis of learning and memory

Besides the histological findings, Santiago Ramón y Cajal proposed a theory for learning in 1894, which is now called the *synaptic plasticity hypothesis*. It claims that the strength of a synaptic connection can be modified by neural activity and further suggested that these plastic connections could be the basis of learning (Squire and Kandel, 2009). Cajal's theory was to a great extent confirmed by subsequent scientific findings. The first experimental evidence for the hypothesis emerged in 1970 by Nobel laureate Eric Kandel and his colleagues who were able to identify cells responsible for the gill-withdrawal reflex in the sea slug *Aplysia* (Castellucci et al., 1970). Performing intracellular recordings from these cells, while habituating the animal to the reflex, they measured profound depression in the EPSP.

This finding strongly contributed to the comprehension of neuroscience, however, it does not establish a causal link between learning and synaptic changes. To prove the synaptic plasticity hypothesis, it was necessary to show that the suppression of plasticity impairs the learning process. This was done in pharmacological studies by blocking certain receptors, in molecular-genetic studies by knocking-out certain receptors or by

optogenetic modifications of neurons. The ultimate proof for the hypothesis is to induce memories by artificially changing synaptic strengths (Takeuchi et al., 2014).

1.2.2 Synaptic plasticity and time scales

The discovery of LTP and LTD raised the question which molecular processes are underlying the increase and decrease of EPSPs. Nowadays we know that the synaptic strength, or more technically *synaptic weight*, can be modified by various mechanisms acting on a wide range of time scales, thereby change the amplitude of EPSPs and the probability of spike generation.

Synaptic plasticity on the time scale of hundreds of milliseconds to seconds is termed short-term plasticity (Stevens and Wang, 1995; Markram and Tsodyks, 1996; Abbott, 1997). On chemical synapses, neurotransmitters are released on the presynaptic site to forward a signal to the postsynaptic site. The strength of the forwarded signal depends on the amount of available neurotransmitters and on the probability that neurotransmitters are released. Both factors may change during the signaling process leading to opposing effects. On one hand, the pool of available neurotransmitters depletes which causes a depression of the weight (short-term depression). On the other hand, spike generation leads to an influx of calcium into the axon terminal increasing the release probability of neurotransmitters and thus increases the synaptic weight (short-term facilitation). Both modifications are only temporary; without continued presynaptic activity, the synaptic weight quickly returns back to its baseline value.

In contrast, so called long-term plasticity gives rise to synaptic changes lasting minutes or more. It can be separated into an early phase of long-term plasticity (early LTP and early LTD) which lasts less than three hours and a late phase of long-term plasticity (late LTP and late LTD) beginning after one to three hours and lasting for ten hours or longer (Frey et al., 1993). The two phases can be distinguished by blocking protein synthesis. In that case, only early LTP/LTD is expressed while the late phase cannot be seen. In many experiments it is additionally observed, that synaptic strengths change rapidly during the early phase while changes are much slower during the late phase.

Various processes can contribute to long-lasting changes. Early LTP, mainly associated with protein synthesis independent processes, can originate by synaptic incorporation of additional AMPA receptors (Hayashi et al., 2000; Redondo and Morris, 2011) and an increase of neurotransmitter release (Kullmann and Nicoll, 1992). Moreover,

changes in the numbers of synaptic vesicles and their distribution leads to early LTP (Lynch, 2004). Contrary, processes contributing to late LTP are protein synthesis dependent, often leading to morphological changes. At Schaffer collateral synapses in the hippocampus, for example, late LTP is associated with a long-term enlargement (Masanori et al., 2004) and late LTD with a shrinkage (Zhou et al., 2004) of dendritic spines. In addition, increase in spine numbers contribute to the late phase (Lynch, 2004).

1.2.3 Consolidation

Experiments showed that initially after formation, memories are fragile and need to be consolidated to become persistent (Brashers-Krug et al., 1996). Consolidation processes can happen both on the system and on the synaptic level (Dudai, 2004).

On the system level, the significance of consolidation could be observed from the behavior of Henry Gustav Molaison, formerly known as patient H.M. (Viola and Moncada, 2014). Part of Molaison's medial temporal lobe was removed in a surgery to alleviate epileptic symptoms. After the surgery, Molaison lost the ability to form any factual long-term memories. Moreover, memories formed relatively close before the surgery were impaired while childhood memories were not. This is evidence that older memories do not rely on the medial temporal lobe while newer memories do (Corkin, 2002). Generally, we can say that system consolidation is a reorganization process where memories dependent on the hippocampus and associated cortices become independent of the hippocampus. The transfer may happen during replay events while sleeping and can last from weeks to years (Dudai, 2004; Frankland and Bontempi, 2005).

Consolidation on the synaptic level, the process turning early LTP/LTD into late LTP/LTD, acts on a much faster timescale. LTP induction experiments show that induced weight changes decay back to baseline within a few hours if the induction stimulus is weak. On the other hand, for strong induction stimuli one can observe a persistent weight change over many hours (Frey and Morris, 1997). In addition, it is found that protein synthesis is a prerequisite for synaptic consolidation (Reymann and Frey, 2007; Murakoshi and Yasuda, 2012).

A widely accepted framework to describe synaptic consolidation is the tagging and capture hypothesis (Frey and Morris, 1997; Redondo and Morris, 2011). The theory describes synaptic consolidation in four steps (Figure 2).

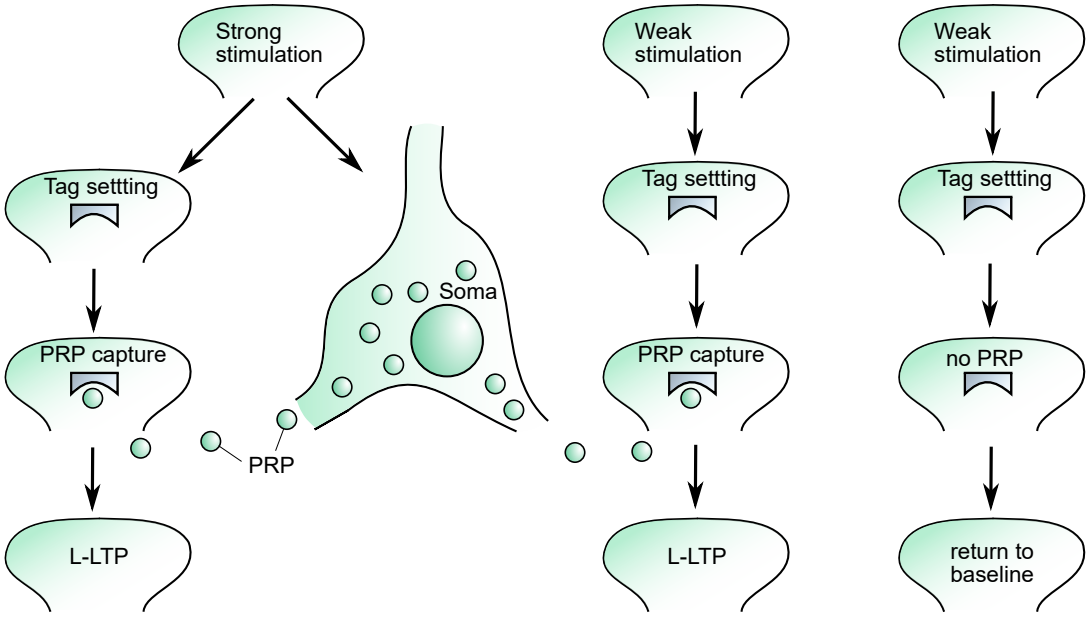

Figure 2: Sketch how the stimulation strength influences consolidation according to the tagging and capture hypothesis. A strong stimulation, e.g. strong tetanus, leads to setting of a tag and synthesis of plasticity related proteins (PRPs) in the soma or dendrites. Weak stimulation, e.g. weak tetanus, only leads to tag setting. If a tagged synapse captures PRPs, consolidation occurs. Otherwise, the induced weight change decays back to baseline. Figure adapted from Redondo and Morris (2011) with permission of Springer Nature.

1. LTP induction leads to the setting of a local, synaptic specific tag. These are molecular changes which mark that synaptic plasticity, i.e. weight changes, has occurred.

2. Strong LTP induction activates synthesis of diffusible plasticity-related proteins (PRPs). This can occur either in the soma or in dendrites.

3. PRPs get captured by tagged synapses.

4. In tagged synapses which captured PRPs, weights are stabilized and late LTP is formed.

This theory emerged from observations made in two-pathway studies, where two independent presynaptic populations projecting to the same postsynaptic neuron are

stimulated (Frey and Morris, 1997). Inhibiting protein synthesis in one pathway still leads to consolidation in both pathways, just like stimulating one pathway weakly and the other strongly. It does not matter whether the weak stimulation occurs up to 30 minutes before or after the strong stimulation. A weak stimulation consists for example of a single tetanus with 20 pulses at 100 Hz; a strong stimulation for example corresponds to a tetanus of three stimulus trains of 100 pulses at 100 Hz with 10 minutes intertrain interval. Stimulating both pathways weakly results only in early LTP, because due to the lack of PRPs, no late LTP can be established. Also, if one pathway received an extremely weak stimulus, weight changes could not be rescued by a strong stimulation on another pathway (Frey and Morris, 1997). These experiments were repeated *in vivo* (Shires et al., 2012) and for LTD induction (Sajikumar and Frey, 2004) with analog outcomes. In addition, cross-tagging experiments show that a strong induction of LTP leads to consolidation of weakly induced LTD on the same neuron and strong induction of LTD enables consolidation of weakly induced LTP (Sajikumar and Frey, 2004; Sajikumar et al., 2005).

1.3 Single neuron model

The function of individual neurons can be described at many levels of abstraction.

A famous example of a biophysical model which describes the complex interactions of ion channels is the Hodgkin-Huxley model (Hodgkin and Huxley, 1952). It was developed based on observations from experiments on the squid giant axon. The model provides a set of nonlinear differential equations that describe mechanistically how the all-or-none nature of a spike emerges and how action potentials are propagated within a neuron, linking neuronal structures to electrical elements. While the lipid bilayer is represented as a capacitance, three types of ion channels (sodium, potassium, leak) serve as electrical conductances. The sodium and potassium channels are voltage and time dependent while the voltage-independent leak channel represents other channel types.

Although widely used, the Hodgkin-Huxley model has been subject to criticism. On one hand, the model comprises only two voltage-dependent ion channels which is a severe simplification of the biophysical processes in the membrane. Researches therefore have developed many extended models (Li and Rinzel, 1994; Pospischil et al., 2008). On the other hand, the Hodgkin-Huxley model is too complex for mathematical analysis of neural networks. For this reason, simpler models are often used in theoretical

works (Meunier and Segev, 2002).

One of the simplest neuron models was proposed by Pitts and McCulloch (1943) and was inspired by the all-or-none character of action potentials. It suggests that neurons are working like a logical gate transforming binary inputs x_j via a step-like activation function into a single binary output y. The activation function acts on the sum of the inputs, returning 1 if the sum crosses a threshold θ and 0 otherwise,

$$y = \begin{cases} 1 & \text{for } \sum_j x_j \geq \theta \\ 0 & \text{for } \sum_i x_j < \theta \end{cases} \tag{1.1}$$

$$= \Theta\left(\sum_j x_j - \theta\right), \tag{1.2}$$

where $\Theta(\cdot)$ denotes the Heaviside step function. While the Pitts and McCulloch model captures the essence of a neuron's functionality, it is limited to binary variables and cannot deal with inputs of varying importance.

In this thesis we will use an adapted version of the neuron model known from the perceptron (Rosenblatt, 1958). A neuron receives multiple inputs x_j (Figure 3). The weighted sum of the inputs evokes a membrane voltage,

$$V_i = \sum_j w_{ij} x_j \tag{1.3}$$

where w_{ij} is the real-valued weight of the connection from input j to neuron i. The weight can in this context be defined as the transmission efficacy or strength of a synapse. The output of the neuron is an instantaneous firing rate r_i, calculated in a non-linear way from the voltage,

$$r_i = \varphi(V_i). \tag{1.4}$$

The function φ is called transfer or activation function. Common choices for the activation function are the step function, the identity function or the hyperbolic function tanh. For biological plausibility it is desired that the activation function is mapping the values to a bounded interval. Moreover, the activation function should also be dif-

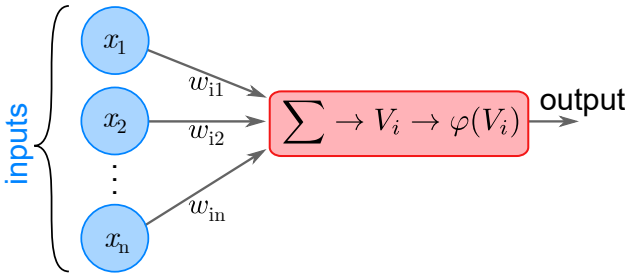

Figure 3: Simple neuron model. The neuron i (red) calculates a weighted sum of the inputs x_j (blue). This sum corresponds to the membrane voltage V_i. The output, a firing rate, is a function of the membrane voltage, $\varphi(V_i)$.

ferentiable for practical reasons (Dayan and Abbott, 2001). A widespread choice which fulfills these two criteria is the logistic function,

$$\varphi(x) = \frac{1}{1 + e^{-(x-a)/\beta}} \,. \tag{1.5}$$

It maps the voltages to firing rates such that they are bounded to the range $[0, 1]$. In addition, the activation function adds a nonlinearity which is in particular important for connected neural layers (Goodfellow et al., 2016).

The inputs x_j can be interpreted as normalized firing rates of a stochastic process, e.g. inhomogeneous Poisson process. In a homogeneous Poisson processes, the firing rate r is defined by the probability P to find a spike in a short time window of duration Δt,

$$r \equiv \lim_{t \to 0} \frac{P(t; t + \Delta t)}{\Delta t} \,. \tag{1.6}$$

For an inhomogeneous Poisson process the firing rate $r(t)$ is time-dependent. In a Poisson process spikes occur independently and stochastically. While it is a reasonable approximation to stochastic neuronal firing (Dayan and Abbott, 2001), it does not reproduce all neuronal properties like refractoriness or reliability of neuronal activity (Amarasingham, 2006).

As an extension to the perceptron neuron, we differentiate in our work between dendritic and axonal initial segment voltage (Figure 4). This becomes important when the neuron is implemented in a neural network to carry out a classification task. To get

an unambiguous classification, a winner-take-all mechanism is desirable, i.e. only one output neuron should fire in response to an input while all other output neurons should be silent. With lateral inhibition, according to which the first spike of a neuron in the output layer suppresses possible spikes in the other output neurons, the winner-take-all mechanism can be achieved. But in order to adapt the weights optimally, synapses need information about the unperturbed voltage. To solve this conflict, the winner-take-all mechanism with the lateral inhibition is deferred to the axonal initial segment. Since soma and dendrites represent a deep current sink viewed from the axon initial segment, the dendritic voltage is only barely influenced by the voltage in the axonal initial segment.

In a computational model, the above explained winner-take-all mechanism can be implemented easier without spikes. After onset of a new input pattern, the voltage resp. firing rate in the dendrites and the axonal inital segment are calculated according to Equations 1.3 and 1.4 . While the dendritic quantities stay constant until the onset of a new pattern, the firing rate in the axonal initial segment is only needed to calculate the probability that given the set of all output neurons, neuron k fires first. This probability is directly related to the normalized firing rates,

$$P_w(k|x) = \frac{\varphi(V_k)}{\sum_i \varphi(V_i)}, \tag{1.7}$$

where the sum runs over all neurons i in the output layer. After the winning neuron k is selected, the firing rate in the axonal initial segment is changed to a all-or-none function,

$$\varphi(V_k) \to \varphi_k^{\mathrm{WTA}} = \begin{cases} 1 & \text{neuron } k, \\ 0 & \text{all other neurons.} \end{cases} \tag{1.8}$$

1.4 Learning rule

During the learning process, which is usually modeled as an adjustment of synaptic weights, postsynaptic neurons change their firing activities by adapting synaptic weights in response to a given input. For example in the context of the perceptron model (Equation 1.3), a change w_{ij} results in a change of the postsynaptic voltage V_i and hence in a different firing statistics while keeping the input pattern fixed.

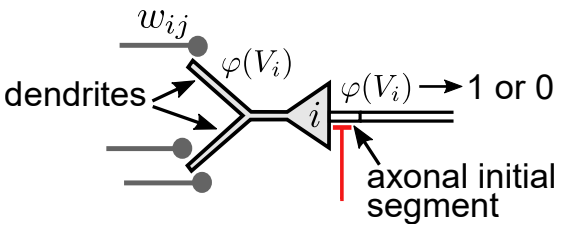

Figure 4: Neuron model with separate voltage in the dendrites and the axonal initial segment. The voltage in the dendrite stays constant till the input patterns change such that the synapse can change their weight w_{ij} as a function of the rate $\varphi(V_i)$. The voltage in the axonal initial segment is initially equal to that in the dendrite but may be inhibited by another neuron (red line).

A learning rule is a function which describes how weights are adapted over time. In general, it is a function of the neuron's input and output, measured either as spikes or rates, and the current weights, however, it can include any number of additional variables. Depending on the learning paradigm, an external signal like a teacher or reward signal is necessary (Stein et al., 2016).

Based on the existence and nature of a external signal, we divide the pool of learning rules into three groups.

In *unsupervised learning*, weights are strengthened or weakened only based on the stimulus and the response to the stimulus, i.e. no feedback signal is incorporated. Thus learning happens solely based on the statistics and correlations of neural activity.

A well-known example is given by Hebb's rule,

$$\tau \frac{dw}{dt} = y_i(t)x_j(t) \tag{1.9}$$

which is based on Hebb's postulates (Hebb, 1949) and suggests that simultaneous pre- and postsynaptic firing increases the weight. In discrete time Hebb's rule reads as follows

$$\Delta w_{ij}(t) = \eta\, y_i(t)x_j(t), \tag{1.10}$$

where η is the learning rate that determines how strongly weights are adjusted in each time step. In its original form, the rule accounts only for increases in synaptic strength, but it can be generalized to include decreases of weights arising from the failure of the

presynaptic neuron to be involved in the activation of the postsynaptic neuron. Hebb's rule is an unsupervised learning rule because weights are strengthened only based on the actual response to a stimulus. Another example of unsupervised learning is the implementation of principle component analysis in neuronal networks (Oja, 1982).

In *supervised learning*, plasticity is driven by a teacher signal ('supervisor') in order to learn a set of input-output pairs. Learning rules for this paradigm try to reduce the difference between the network's output and the desired output. The teacher signal encodes this error and feeds it back to the network. All weights are then individually adapted to improve performance. A well known example of this type of learning is the backpropagation-of-errors algorithm (Rumelhart et al., 1986). Although supervised learning is very successful in machine-learning, it is often not considered to be biologically plausible (Grossberg, 1987; Crick, 1989).

Reinforcement learning does not rely on such a explicit supervision signal, instead, the network receives a reward signal which feeds back whether the output was correct or incorrect. In mathematical terms, reward can be represented by a global binary variable. Unlike supervised learning, the network does not know which neuron contributed how strongly to the success or failure of a trial. Reinforcement learning is discussed in more detail in Section 1.5.

A common approach to find a new learning rule is to first define a loss or cost function that depends on the synaptic weights. While for a supervised learning rule the cost is typically the difference between the desired and actual network output, in reinforcement learning the negative expected reward is a possible choice. The goal of learning is then adapting the weights in such a way that the cost function is minimized. A strategy to efficiently minimize the cost function is following the direction of steepest descent. Since the gradient points into direction of the steepest descent, adapting the weights in direction of the gradient guarantees that learning ends up in a local minimum of the cost function.

In practice, however, gradient descent algorithms are not guaranteed to find the optimal solution. They can fail to find a good weight configuration if the cost function exhibits multiple minima or plateaus. The framework also presumes that the gradient of the cost function is exactly calculated or at least well approximated. In a biological setting weight changes happen continuously, thus the gradient is only sampled based on the current or past input-output relations (online-learning). In machine learning, which is not bound to biological constraints, one can repeatedly run through all examples of

the training dataset, and after each run calculate the gradient and perform weight updates (batch-learning).

Depending on the underlying structure, extensions of gradient descent that take the geometry of the output space into account can be more efficient (Amari, 1998). In this thesis we will use a learning rule which is derived from the gradient of a cost function, but incorporates additional modifications to prevent premature saturations.

1.5 Reinforcement learning

Reinforcement learning is learning how to best act in an environment using a reward signal as feedback. The learner (or often called 'agent') must find out by itself which actions are good by trying them out (Barto and Sutton, 2018). Thus, in its simplest version reinforcement learning corresponds to trial-and-error learning. The agent, interacting with an environment, gets exposed to inputs and undertakes varied actions based on a policy with the goal to maximize the numerical reward signal. The reward signal is traditionally positive, but can be negative if punishment is incorporated in the framework.

Positive reward signals reinforce the previous action and will hence increase the probability that the same action is taken again. Contrary, negative reward signals evoke the opposite effect. However, actions are not only based on the past experiences (exploitation), the agent should also make new choices to explore the environment. Finding the optimal balance between exploration and exploitation is a distinctive challenge when applying reinforcement learning (Barto and Sutton, 2018). The concrete implementations of the exploration-exploitation balance and the update of action probabilities depend on the specific algorithm. A classical example of a reinforcement learning task consists of finding a way through a maze where multiple sequences of actions can lead to reward but some of them might be shorter and thus preferred. If the agent found one way, the question is whether to exploit that solution or to discover the environment further to find even better solutions.

State-of-the-art reinforcement learning algorithms use concepts similar to those of biology. Most important to mention is the concept of neural networks. In machine-learning tasks with a limited number of possible states data can be stored in lookup tables. For systems with large number of states or actions, tables become unmanageable due to their memory consumption and computational time to learn the value of each

state individually. Thus, lookup tables get replaced by functional approximations. It turned out that combining neural networks and reinforcement learning works extremely well (Mnih et al., 2015). Alternative functional approximations are decision trees which are easier to understand than neural networks but may have weaker performance (Ze et al., 2013).

Another similarity between reinforcement algorithms from machine learning and brain function is the incorporation of reward-prediction error, i.e. the difference between the actual reward and the expected reward. In temporal-difference learning, a class of methods in machine learning, the difference between temporally successive estimates of the reward is used to improve learning performance (Sutton and Barto, 1981; Szepesvari, 2010). Animal experiments showed that neurons can carry information about the actual reward (Okada et al., 2009), about the expected reward (Oyama et al., 2015) and about the difference of the actual and expected reward (Hollerman and Schultz, 1998).

Given the current state of research, there is evidence that the reward prediction error is encoded by the activity of dopaminergic neurons (Montague et al., 1996; Schultz et al., 1997). These neurons are mainly located in the ventral part of the midbrain (Colombo, 2014) and have extensive axonal arbors projecting to many brain regions. An axon of a dopaminergic neuron releases dopamine at roughly 500,000 dendritic spines (Schultz, 1998). In that way it has modulatory effects on synaptic plasticity (Schultz, 1998).

1.6 Perceptual learning and roving

Perceptual learning is improving perception skills. It is a form of reinforcement learning (Herzog et al., 2012). In a typical perceptual learning experiment, participants are exposed to two or more stimuli which they need to distinguish. Perceptual learning experiments have been performed for many types of stimuli, e.g. auditory stimuli (Karmarkar and Buonomano, 2003) and visual stimuli (Karni and Sagi, 1991). Several aspects can influence the performance of perceptual learning, among others attention, task difficulty, and the order of stimuli presentation.

While perceptual performance generally improves if the number of trials increases, Parkosadze et al. (2008) showed that this is not always the case if two types of stimuli are presented randomly interleaved, a condition referred as *roving*. They used line bisection stimuli, i.e. a spatial interval bounded by two vertical parallel lines is bisected in two unequal broad components by a third vertical line. In each trial of the experiment,

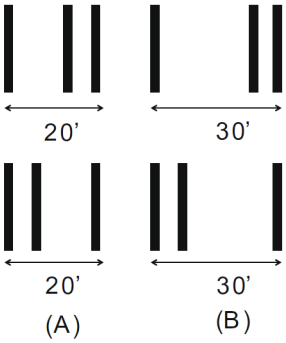

Figure 5: Bisection stimuli as used for studying the roving phenomenon. In each trial, subjects see for a short time one of the four stimuli alternatives and have to decide whether the middle line has an offset to the left or right. If in a session stimuli from set (A) or set (B) are used alone, performance increases with more trials carried out. However, if all four stimuli alternatives are used during one session, no learning is observed. Figure adapted from Tartaglia et al. (2009a) with permission of Elsevier.

subjects saw one of the two stimulus alternatives and they had to report whether the middle line is closer to the left or right outer line. Parkosadze et al. (2008) observed that perceptual performance increased during the experiment. In a second experiment, two types of bisection stimuli were used, one with an outer line distance of 20' and one with an outer line distance of 30' (Figure 5). If the stimuli were randomly interleaved performance did not increase anymore.

Learning impairment during roving conditions is however not always observed, for example, when the bisection stimuli had all the same length, but in one case the lines where vertical and in the second case the lines were rotated by 45°.

Similar results for other stimuli types were observed by Otto et al. (2006), Tartaglia et al. (2009a), Yu et al. (2004) and Zhang et al. (2008). Outside of perceptual learning, Flesch et al. (2018) observed learning impairments in a setting similar to roving. In their categorization experiment, the classification criterion was alterable. If longer blocks with the same classification criterion were trained, participants significantly performed better than when the classification criteria were randomly interleaved.

1.7 Catastrophic forgetting and continual learning

The brain is exposed to a continuous stream of external stimuli as a consequence of ongoing interactions with the environment. While different stimuli may activate different brain areas, a reasonable assumption is that at least for similar stimuli a common set of neurons is excited. This leads us to the question, how synaptic weights are modified if their pre- or postsynaptic neurons are excited in different situations? This question is especially relevant if we look at continual or life-long learning tasks, i.e. multiple tasks are learned each after the other.

Most computational models of learning focus on one task. Thus, they may not explain the correct behavior in a continual learning task setting. Actually, many models would exhibit catastrophic forgetting (or catastrophic interference), i.e. the previous learned task is forgotten on a catastrophically fast time scale when exposed to a new task (French, 1999). This happens because weights are continuously adapted to the current task. Therefore, also a weight configuration optimal for the previous task will be altered when learning a new task. If the new weight configuration is inappropriate for the previous task, that task is forgotten.

Humans are forgetful and to some extent this is beneficial. By forgetting the particulars, humans gain the possibility of abstracting and retaining the important essence of their memories. This helps humans to generalize, abstract and assemble general knowledge which is often more important than retaining a literal record of particular events (Squire and Kandel, 2009). Forgetting can happen passively by synaptic weight decay or actively due to interference (Wixted, 2004). However, in contrast to computational models, forgetting in humans normally happens gradually and not catastrophically (McCloskey and Cohen, 1989).

Catastrophic forgetting is one expression of the so called stability-plasticity dilemma (Carpenter and Grossberg, 1988; McClelland et al., 1995; Fusi et al., 2005; Kumaran et al., 2016). How to design models of synapses which are plastic enough to learn new tasks and stable enough to keep old memories? Too much plasticity results in ongoing forgetting of old memories but too much stability prevents learning of new memories. Many models exist which tackle the problem of catastrophic forgetting and the stability-plasticity dilemma (see Section 1.8.1) but so far it is still largely unclear how the nervous system handles the problem.

Catastrophic forgetting can also occur outside of continual learning. A neural network with a fixed architecture has an upper bound on the number of associations it

can store. If more associations are loaded into the network, memory retrieval becomes abruptly impossible (Robins and McCallum, 1998).

This PhD work introduces a model that makes use of synaptic consolidation and thereby prevents catastrophic forgetting naturally.

1.8 Computational models

In this section we present a selection of computational models which tackle the same problems or use similar methods as we do in our work.

1.8.1 Models of continual learning

Continual learning refers to learning multiple tasks each after the other. Sometimes the terms lifelong learning or incremental learning are used which we use interchangeably.

To our knowledge the earliest work tackling the problem of context interference in continual learning were Hinton and Plaut (1987). They used a fast and slow weight evolving according to the same learning rule but with different learning rates. In addition, the fast weights decay towards zero. Thus, long-term memories are stored in the slow weights. In simulations they could show that disrupted memories of associations can be retained by only retraining a fraction of the associations thanks to the fast weights. However, the regained memories fade away again when the fast weights decay.

Fusi et al. (2005) analyzed how the quality of memories degrades over time due to ongoing plasticity. They looked on one hand at the signal-to-noise ratio immediately after memory storage. This is a measure of the flexibility of the network to store new memories. On the other hand, they looked at memory lifetime. Preferably, both quantities should be large but it turns out that in a general model maximizing one quantity comes with the cost of decreasing the other quantity. This is an intrinsic problem of models where memories decay exponentially due to deleterious effects of plasticity. The authors show that memories decaying according to a power law suffer less from the trade off between flexibility and memory life time. Consequently, they propose a model where memories decay approximately according to a power law.

The model is based on on the finding that interactions of multiple exponential processes acting on a wide range of time scales generate a power-law dynamic (Anderson,

2001). Thus, the model has two levels of synaptic strength and each level has a cascade of metaplastic states. Synapses can switch from one synaptic strength level to the other as well as move down the cascade of metaplastic states. The transition probabilities between the states vary. The upper states are the most plastic ones and the further down the cascade a synapse moves the more stable it gets.

The cascade model by Fusi et al. (2005) uses binary synapse and linear chains of dynamical variables. In Benna and Fusi (2016) a new cascade model is proposed which deals with continuous weights and arbitrarily complex network interactions (Figure 6A). Kaplanis et al. (2018) could show that the cascade model can be used for learning.

To prevent catastrophic forgetting in deep neural networks, Kirkpatrick et al. (2017) introduced a new approach using so called elastic terms. In deep neural networks many weight configurations may exist that result in the same performance for a given learning task. Hence, it is likely that close to a found weight configuration for a previously learned task one can find a weight configuration which is optimal for a new yet to be learned task. The challenge is to find this second weight configuration close to the first one. The proposed mechanism adds a quadratic penalty term to the cost function. The penalty term is proportional to the diagonal of the Fisher information matrix which constrains the weights to stay in a region of low error for the first task while learning the second task (Figure 6B). As a consequence, the plasticity gets selectively decreased for specific weights. For each additional task another penalty term gets added, which leads to further reduction of plasticity.

Zenke et al. (2017) used a similar approach as Kirkpatrick et al. (2017) but the importance of weights are computed online during training by estimating the sensitivity of the cost function to a weight change.

A third work based on the same approach was described by Aljundi et al. (2018). They also estimate an importance for each weight and reduce plasticity for important weights accordingly. Importance is estimated based on the sensitivity of the learned function to a weight change. The estimation is done in an online manner, but the method can also be added after the network is trained. Importance can be computed on any set of data without the need for labels.

Yet another method to counteract catastrophic forgetting is explored by Isele and Cosgun (2018) and Rolnick et al. (2018). They use a buffer to store previous learned tasks. During learning a new task, replays of prior tasks are incorporated. This also

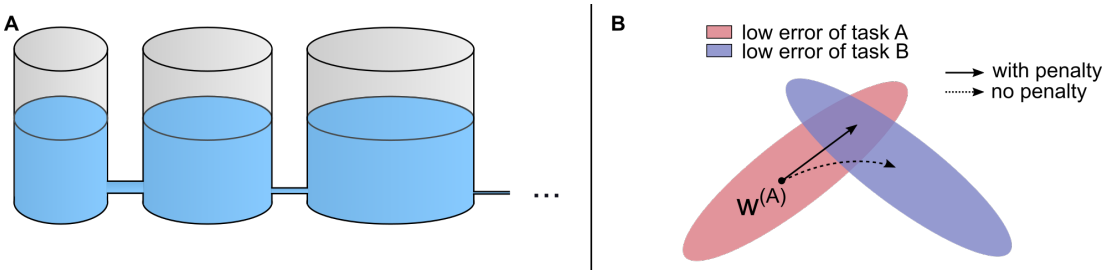

Figure 6: (A) A series of beakers with increasing size and decreasing tube width resembles the models by Fusi et al. (2005) and Benna and Fusi (2016). A chain of exponential decays with different time constants produces approximately a power-law decay. (B) With a penalty term, Kirkpatrick et al. (2017) Zenke et al. (2017) and Aljundi et al. (2018) ensure that task A is remembered while learning task B. If the cost function of task B is minimized alone (dotted arrow), task A is forgotten. With the penalty term (solid arrow) learning is restricted to stay in a region with low error for task A. Figures adapted from Kaplanis et al. (2018) and Kirkpatrick et al. (2017).

works, if the buffer has a limited capacity and not all previous tasks can be stored completely in it.

1.8.2 Models of tagging and consolidation

Barrett et al. (2009) use a stochastic Markov process to explain tagging and capturing. It contains six states: weak basal, strong basal, early LTD, early LTP, late LTD and late LTP. Between these states ten transitions are possible (weak $\leftrightarrow$ strong basal state, weak/strong basal state $\leftrightarrow$ early LTP/LTD, early LTP/eLTD $\rightarrow$ late LTP/LTD, late LTP/LTD $\rightarrow$ weak/strong basal state) described by a total of seven transition rate parameters. Synapses are each in one of these states and depending on the state either have a weight w or $2w$. Stimulations change the transition rate parameters, thus allowing the synapse to change their state. Synapses in the states early LTD or early LTP are assumed to be tagged. Strong stimulations allows the transition from early LTD to late LTD and early LTP to late LTP. Both transitions are described by the same parameters thus describing both phenomena of capturing and cross-capturing.

Clopath et al. (2008) models synapses as discrete quantities which can switch from a non-tagged state to a 'high state' (during LTP induction) or to a 'low state' (during LTD induction). If a synapse is in a high or low state, it is tagged. The tag resets

stochastically to zero with a certain rate. If the total number of tags reaches a threshold, protein synthesis is triggered. As a consequence, a consolidation variable which has initially two stable fixed points loses one fixed point. The synaptic weight is a linear combination of its high state value, low state value and consolidation variable value. A shortcoming of that model is that there is no separation between tags and initial expression of synaptic plasticity.

Ziegler et al. (2015) developed a model using a similar concept as Clopath et al. (2008). It can however account for several more phenomena; it includes depotentiation effects and it separates tags from initial expression of synaptic plasticity. A synaptic state consists of three layers, a weight, a tag and a scaffold. The weight is the only layer which influences the post-synaptic voltage. Each layer is described by a noisy variable evolving in a double-well potential, thus each layer has two stable fixed points. In addition, each layer gets influenced by its neighbor layer(s) modeled by two coupling constants. With this setup, many phenomena like tagging and consolidation can be well described. For example, the induction of LTP changes the value of the weight from its low state to its high state. Tagging corresponds to opening the gating variable between the weight and the tag, and thus drags the tag from its low state to its high state. If this does not occur, the weight value decays back from its high state to its low state. Consolidation is modeled similar. The gating variable between the tag and the scaffold opens by he presence of a reward or surprise signal. This allows the scaffold to get dragged by the tag from its low state to the high state. The model also accounts in a similar way for the phenomenon of tag resetting and LTD induction.

1.8.3 Models explaining the roving phenomenon

The bisection stimulus task under roving conditions (see Section 1.6) can be learned by the simplest supervised models, e.g. single-layer perceptron, because the task is linearly separable. In contrast, humans are not able to learn it as experiments showed. It is therefore worth investigating why humans are inferior to these simple models.

As Herzog et al. (2012) conclude, perceptual learning is neither supervised learning nor unsupervised learning but reinforcement learning. Purely supervised learning can be ruled out because perceptual learning also occurs without feedback. But feedback speeds up perceptual learning which rules out unsupervised learning.

Herzog et al. (2012) continue to point out that two classes of reinforcement learning

exist. One class are the so-called R-max rules where weight changes depend on the reward R. Weight w_{ij} is increased if higher than average postsynaptic activity leads to reward,

$$\Delta w_{ij} = \mathrm{pre}_j \times (\mathrm{post}_i - \overline{\mathrm{post}_i}) \times R\,. \tag{1.11}$$

However, learning rules of this class do not suffer from roving conditions and therefore fails to explain the experiments. The second class of reinforcement learning rules have the form

$$\Delta w_{ij} = \mathrm{pre}_j \times \mathrm{post}_i \times (R - \bar{R})\,. \tag{1.12}$$

Rules of this shape rely on a correct estimate of the mean reward $\bar{R}$. If this is not given, the learning rules suffers from a so-called unsupervised bias term; weights updates are unrelated to the correlation of actual reward and activity in a given trial. As a consequence this can drift weights in a potentially unrewarded region (Frémaux et al., 2010). In the bisection stimuli task, the mean reward for each stimulus type can indeed not correctly estimated because both stimulus types can come with a different reward. Therefore, the unsupervised bias term is not zero and learning is impaired.

CHAPTER 2

Hypothesis and Aim

A journey to a thousand miles begins with a single step.

— Lao Tzu

One of the big open questions in neuroscience is how synaptic weights are changed during learning. This question can surely not be answered within the scope of a PhD thesis but our aim is to shed light on this question by developing a computational model that provides insights to a few aspects of learning and memory. The model should establish ties between multiple time scales, but also between multiple spatial scales. It should describe learning that lasts several minutes and its effects that last several days. The aim is to have a model which can reproduce data from electrophysiological recordings while keeping it as simple as possible.

In particular, we aim to describe synaptic plasticity more accurately by incorporating multicomponent synaptic weights in our model. Thereby, biological plausible answers to the plasticity-stability dilemma and the phenomenon of catastrophic forgetting should be found as wells as synaptic consolidation and learning disruption during roving conditions better explained.

Multicomponent synaptic weights

Many computational models exist that try to explain learning, most of these models use scalar variables as weights. This is an oversimplification because it is well known that complex biochemical processes occur permanently in individual synapses. From several experiments made in animals we know that synaptic weights can change on different timescales (Frey et al., 2009), proteins are delivered to the synapse in three sequential temporal phases (Bosch et al., 2014) and at least two distinct phases of long-term potentiation can be identified (Krug et al., 1984; Frey et al., 1988; Bliss and Collingridge, 1993). Thus, most computational models could miss essential functional components as they do not represent this complexity. Hinton and Plaut (1987) wrote that there have been relatively few attempts to investigate computational advantages of giving each synapse several weights that change at different speeds. More than 30 years later, this statement still holds. We hypothesize that incorporating weights with multiple components into a computational model, synaptic plasticity relevant for associative learning can be more accurately described without needing to model detailed biochemical processes.

Plasticity-stability dilemma

The pace of synaptic weights change is crucial for learning and memory. The plasticity-stability dilemma illustrates well this conundrum. Synapses should be plastic to learn new tasks and stable to keep old memories (Carpenter and Grossberg, 1988; McClelland et al., 1995; Fusi et al., 2005; Kumaran et al., 2016). Synapses that are too stable prevent the nervous system to learn new tasks, synapses that change too easily lead to catastrophic forgetting. Recently, several attempts were made to find a solution to this dilemma (Fusi et al., 2005; Benna and Fusi, 2016; Kirkpatrick et al., 2017; Zenke et al., 2017; Aljundi et al., 2018). All these models postulate hidden states that do not contribute to the effective synaptic weight. We aim to model consolidation in a biological plausible way without the need of hidden states, and by doing so, finding a new mechanism how the plasticity-stability dilemma can be overcome.

Synaptic consolidation

Timing plays a key role when studying consolidation. While consolidation can happen through multiple processes on different timescales (Dudai, 2004), the question about

the optimal time of each process is relevant. Early fixation of synaptic weights can result in overloading the nervous system with irrelevant association which can cause a loss in the ability to generalize concepts or even the loss of the ability to learn. If consolidation sets in late, the acquired transient memories may already infer with new inputs. In isolated experiments triggers for consolidation could be found (Reymann and Frey, 2007; Redondo and Morris, 2011). We speculate that by incorporating these triggers in a model of learning, we allow for continual learning and prevent interference of memories.

Roving phenomenon

An often ignored aspect when developing computational models is the fact that human strategies are not optimal in all environments, although evolution drives the brain to get more reward (Schultz, 2015). An example is learning under roving conditions (Parkosadze et al., 2008), c.f. Section 1.6. We hypothesize that fast weight changes are optimal if contexts change slowly (some ten minutes) but if contexts change quickly, e.g. roving conditions, learning is hampered. Our intuition is that fast weight changes have a large variance which can only be corrected for slowly changing contexts.

CHAPTER 3

Results

I want to know why the universe exists, why there is something greater than nothing.

—Stephen Hawking

This chapter presents the following submitted, but not yet published, article:

Synaptic weight decay with selective consolidation enables fast learning without catastrophic forgetting.

Author contribution

The manuscript was written by Prof. Dr. Walter Senn and myself with contributions by Prof. Dr. Michael Herzog[1]. Analytical derivations and model simulation were carried out by me. All illustrations were drawn by me and combined with simulation results into the resulting figures.

[1]Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL)

Synaptic weight decay with selective consolidation enables fast learning without catastrophic forgetting

Pascal Leimer[1], Michael Herzog[2], Walter Senn[1],

[1] Department of Physiology, University of Bern, Switzerland
[2] Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Abstract

Learning can interfere with pre-existing memories that in classical neural networks may lead to catastrophic forgetting. We present a gradient-based 2-component synaptic plasticity model that enables context-dependent fast learning without catastrophic forgetting. A highly plastic and quickly decaying fast component enables swift learning in a given context. Part of the decaying fast component is selectively secured in a long-lasting slow component that protects context-specific information. While the synaptic decay is governed by a sparsity constraint on the voltage deflections and the fast synaptic component, the consolidation jointly selects events associated with consistent suprathreshold voltage deflections and suprathreshold fast synaptic components. The two components produce a phenomenology that can be interpreted as an early decay of both long-term potentiation and depression that is (cross-) consolidated by a tag & capture mechanism as observed in biological synapses. For reinforcement learning in multiple contexts, the plasticity rule maximizes the expected reward while minimizing interferences between subsequent contexts. As a downside of these mechanisms, learning is hampered when consolidation is triggered prematurely by interleaving easy and difficult tasks, consistent with human psychophysical experiments.

Introduction

Depending on the context, similar situations may require different actions. For instance, when we wait at a crosswalk in continental Europe we first look to the left, but in the UK we look to the right. After living a while at a new place, the reliably acquired

behavior causes a short moment of uncertainty when crossing the road the first time back home. In general, learning to correctly act in one situation may interfere with learning in a similar situation. Retaining context information imposes a challenge to the underlying neuronal network, in particular if contexts switch quickly. On the one hand, synaptic connection strengths, which were important in a previous context, need to be protected from being overwritten. On the other hand, new associations have to be learned for similar inputs that require a different action in another context. This problem is known as the stability-plasticity dilemma, and mechanisms at the network and synapse level have been proposed to tackle it (Abraham and Robins, 2005; Fusi et al., 2005; Carpenter and Grossberg, 1988).

Catastrophic forgetting is a problem in neuronal networks, but not so much in human continual learning (French, 1999; Kumaran et al., 2016). In classical neuronal network models, the connection strengths are adapted to learn new associations in the current context regardless of the importance of the connections in previous contexts. For strong overlaps of the input patterns, ongoing learning results in memories which are forgotten on a catastrophically short time scale. To prevent catastrophic forgetting, memories can be consolidated at the level of the system or the synapses. In systems consolidation, memories are suggested to be transferred to other brain areas for long-term storage such as the hippocampus (Squire and Alvarez, 1995; McClelland et al., 1995; Roxin and Fusi, 2013). In synaptic consolidation, the fast decay of the so-called early long-term potentiation/depression (early LTP/LTD) was shown to be prevented by a synaptic tagging & capture mechanism (Frey and Morris, 1997; Sajikumar and Frey, 2004; Morris, 2006; Redondo and Morris, 2011; Shires et al., 2012; Bosch et al., 2014). Synaptic models with a cascade of internal states of progressively longer retention times were shown to prevent an exponentially fast forgetting while being continuously exposed to stimuli (Fusi et al., 2005; Benna and Fusi, 2016; Kaplanis et al., 2018). Forgetting can also be counterbalanced by reducing plasticity for weight configurations that are important in previous contexts, but assessing the importance of a synaptic weight requires additional information and memories (Kirkpatrick et al., 2017; Zenke et al., 2017; Aljundi et al., 2018).

The passive decay of the early LTP/LTD is typically seen as a weakness of the synapses that needs to be counteracted. Instead, forgetting may itself be beneficial (Brea et al., 2014). We suggest that this passive forgetting is the expression of minimizing the interference between subsequent contexts in the presence of fast learning. The fast synaptic plasticity may enable the retention of information in working mem-

ory (Mongillo and Denève, 2008) that must fade out in time to prevent cross-talks in a subsequent context where similar stimuli may need to be differently processed. The passive forgetting further allows for selectively retaining relevant information by the tag & capture consolidation (Frey and Morris, 1997). Various phenomenological models reproducing the tag & capture mechanisms exist (Clopath et al., 2008; Barrett et al., 2009; Ziegler et al., 2015). But how can the ideas of fast learning, passive weight decay and synaptic consolidation for minimizing cross-talks be captured in a simple normative theory of synaptic plasticity and learning, derived from optimality principles?

Here we suggest a 2-stage model of synaptic modifications and consolidation that maximizes the expected reward in a reinforcement learning context through stochastic gradient ascent. For simplicity, and different from previous models (Fusi et al., 2005; Clopath et al., 2008; Barrett et al., 2009; Benna and Fusi, 2016; Kaplanis et al., 2018; Ziegler et al., 2015), we do not consider internal synaptic variables. Instead, our fast ('early') and slow ('consolidated') components both affect the synaptic strength. Early LTP/LTD promotes fast learning and thereby enables exploitation of reward in a reinforcement learning scenario. It necessarily includes a fast decay to prevent inferences with future contexts where similar sensory inputs require different responses. Hence, early LTP/LTD and synaptic tagging/consolidation are seen as a means to address the stability-plasticity dilemma. We suggest a normative theory of synaptic consolidation that functionally reproduces the phenomena of the multi-state models of tagging & capture (Barrett et al., 2009; Clopath et al., 2008; Ziegler et al., 2015). Our theory asserts that fast learning can be achieved in a given context with minimal interference with other contexts, provided that only strong synaptic modifications associated with peak postsynaptic depolarizations are retained. Weak synaptic modifications associated with low postsynaptic depolarizations must be extinguished on the timescale of the context duration, relating to the decay of the early LTP/LTD.

While our 2-component plasticity model boosts learning for typical context switches, it may hamper learning when the task-difficulties change too quickly. Such phenomena are in fact observed in psychophysical experiments involving stimulus or task mixing (Tartaglia et al., 2009a; Flesch et al., 2018). We postulate that the passive weight decay with the selective consolidation that allows for fast context-dependent learning is the reason why humans, unlike neural networks with simpler plasticity models, show reduced performances in these mixing experiment.

Results

Reward maximizing learning rule with selective consolidation

We hypothesize that early LTP and its consolidation mechanism is one of nature's choices to deal with the stability-plasticity dilemma. To keep the benefit of fast synaptic plasticity while avoiding contextual interference, the network should only consolidate a minimum number of synapses while still be able to change enough synapses if required by novel learning. A particularly pronounced form of the stability-plasticity dilemma becomes apparent when similar tasks are learned each after the other. To illustrate this problem, we consider a learning task that extends across two subsequent contexts (Fig. 7A). The context defines the criteria according to which sensory patterns are classified, and patterns in the two contexts are similar, but not identical. They consist in written words that have to be classified in the first context according to the color of the letters, and in the second context according to the meaning of the word. Learning takes place in a single-layer network that classifies the correlated input patterns by a winner-take-all (WTA) dynamics; the first spike of an output neuron suppresses other output neurons via global inhibition (Fig. 7B).

If the task is learned with a classical (1-component) reward-based learning rule, the synaptic modifications in the first context are likely to be undone during learning in the second context (Fig. 7C). To prevent this, we consider a 2-component plasticity model that consolidates appropriately selected weight changes and protects them from erasure (Fig. 7D). A fast weight component, w^f, enables quick memory acquisition and acts on a short timescale, and a slow component, w^s, allows for keeping a selected memory across a longer timescale. The sum of the two components determines the total synaptic strength, $w = w^f + w^s$.

We postulate that the dynamics of the fast weight component follows approximately the gradient of the utility function U defined by the expected reward and two penalty terms,

$$U = \langle R \rangle - \frac{\lambda_w}{2} \sum_{ij} \|w_{ij}^f\|^2 - \lambda_V \sum_i |V_i| \,, \tag{3.1}$$

where R is the binary reward signal released in response to the action of the network, V_i is the voltage of neuron i, and w_{ij}^f is the fast weight component of the synapse from the presynaptic neuron j to the postsynaptic neuron i. Beside the expected reward

$\langle R \rangle$, the utility function has two penalty terms. The first term punishes high values of the fast weight components. This term results in a passive decay of the fast component such that, in the absence of new plasticity events, it converges to baseline value $w_{ij}^f = 0$. The second term punishes voltage deflections $|V_i|$ from rest at 0, and takes energy costs into account. At the same time this term helps to reduce interferences within contexts. The positive factors λ_w and λ_V represent a weighting of the two penalty terms.

To calculate the gradient of the utility function U with respect to w_{ij}^f we assume that the postsynaptic neuron fires with instantaneous Poisson rate $\varphi(V_i)$, where the postsynaptic voltage is the sum of the input rates x_j weighted by the effective synaptic strengths w_{ij},

$$V_i = \sum_j w_{ij} x_j \,, \text{ with } \; w_{ij} = w_{ij}^f + w_{ij}^s \,. \tag{3.2}$$

The firing rates range between 0 and $\varphi_{\max}$. Stochastic gradient ascent on the utility function U leads to the update of the fast weight component at repetitive time steps (assumed to be every 30 s, see Methods),

$$\Delta w_{ij}^f = \eta_i \, (R - \bar{R}) \Big(\varphi_i^{\mathrm{WTA}} - \varphi(V_i) \Big) \, x_j - \lambda_w w_{ij}^f - \lambda_V \mathrm{sign}(V_i) x_j \,, \tag{3.3}$$

with $\bar{R}$ being the mean reward, and $\varphi_i^{\mathrm{WTA}} = \varphi_{\max}$ if neuron i is the winner neuron ($i = k$) and $\varphi_i^{\mathrm{WTA}} = 0$ else. This time-discrete update implicitly assumes a time step $\Delta t = 1$ (suppressed in Eq. 3.3 and below) that corresponds to the biological time of $\frac{1}{2}$ minutes. The decay term $-\lambda_V w_{ij}^f$, for the optimal values for context learning, causes a passive decay in the order of $1/(2\lambda_w) = 42$ biological minutes.

The activity φ_i^{WTA} represents the target value for the instantaneous firing rate $\varphi(V_i)$. For positive reward ($R = 1$), the weights driving the winner neuron k are strengthened to push $\varphi(V_k)$ towards $\varphi_{\max}$, and the weights driving the non-winning neurons are weakened to push $\varphi(V_i)$ towards 0, with the opposite changes in the absence of reward ($R = 0$). The learning rate η_i incorporates a positive factor necessary for the gradient property to hold (Methods). The voltage penalty causes a decrease or increase of an activated weight ($x_j > 0$), depending on whether the postsynaptic voltage is above or below rest ($V_i = 0$), respectively. In essence, $-\mathrm{sign}(V_i) x_j$ is an anti-Hebbian term that subtracts away the non-informative overlaps in the presynaptic activity patterns.

To store long-term memories, the fast component needs to be consolidated. We postulate that this is done when two conditions are met: (1) the instantaneous value of the fast component, $|w^f|$, exceeds a weight threshold, and (2) the low-pass filtered

postsynaptic voltage, $\bar{V}$, exceeds a voltage threshold. Formally, the decay $-\lambda_w w_{ij}^f$ of the fast component is transferred to the slow component provided the two conditions are simultaneously satisfied,

$$\Delta w_{ij}^s = \lambda_w w_{ij}^f \Theta(|w_{ij}^f| - \theta_w)\, \Theta(\bar{V}_i - \theta_V)\,. \tag{3.4}$$

where the Heaviside step function $\Theta(.)$ is 1 if the argument is positive and 0 else.

Preventing catastrophic forgetting with the 2-component rule

We next test the learning rule when the network performs the above mentioned learning task that is distributed across two subsequent contexts each 100 biological minutes long, interleaved by a 50 biological minutes break. To conceptualize the idea, we consider four input classes $\bar{x}^l$ ($l = 1...4$), with each class requiring its desired output defined by the unique activity of a specific output neuron. Input patterns are defined as noisy samples x^l around $\bar{x}^l$ with independent Gaussian noise on the components. The mean input patterns are $\bar{x}^1 = (1,\, 1-\epsilon,\, \alpha,\, \alpha)$, $\bar{x}^2 = (1-\epsilon, 1,\, \alpha,\, \alpha)$, $\bar{x}^3 = (\alpha,\, \alpha,\, 1,\, 1-\epsilon)$, $\bar{x}^4 = (\alpha,\, \alpha,\, 1-\epsilon, 1)$, see Fig. 7A. Patterns from the first two classes are presented within the first context (c1), and patterns from the second two classes within the second context (c2). Within each context, the classes are only distinguishable by a small ϵ in one component, and across contexts classes have strong overlap α (close to 1). Patterns from the two classes are presented many times (200, corresponding to 100 minutes) in random order until the context switches and patterns from the other two classes are randomly presented. In response to a pattern presentation one output neuron fires first, and if this matches the desired output neuron of the class, a global reward signal ($R = 1$) is given. Otherwise reward is omitted ($R = 0$, Fig. 7B).

We first performed the learning experiment with the 1-component rule. This is the rule with only the fast component w^f governed by Eq. 3.3 while setting the decay parameter λ_w to zero and thus also $w^s = 0$ (Fig. 7C). The other parameters were optimized for highest performance (Methods). Because the 1-component rule follows approximately the reward gradient, the associations can be learned separately within each context (Fig. 8A). However, because in the second context similar patterns are associated with different outputs, the associations learned in the first context are overwritten. The larger the overlap (α), the stronger the interference. When retesting the associations of the first context after learning in the second context, performance decayed roughly to a third of the original one (Fig. 8A and B, orange).

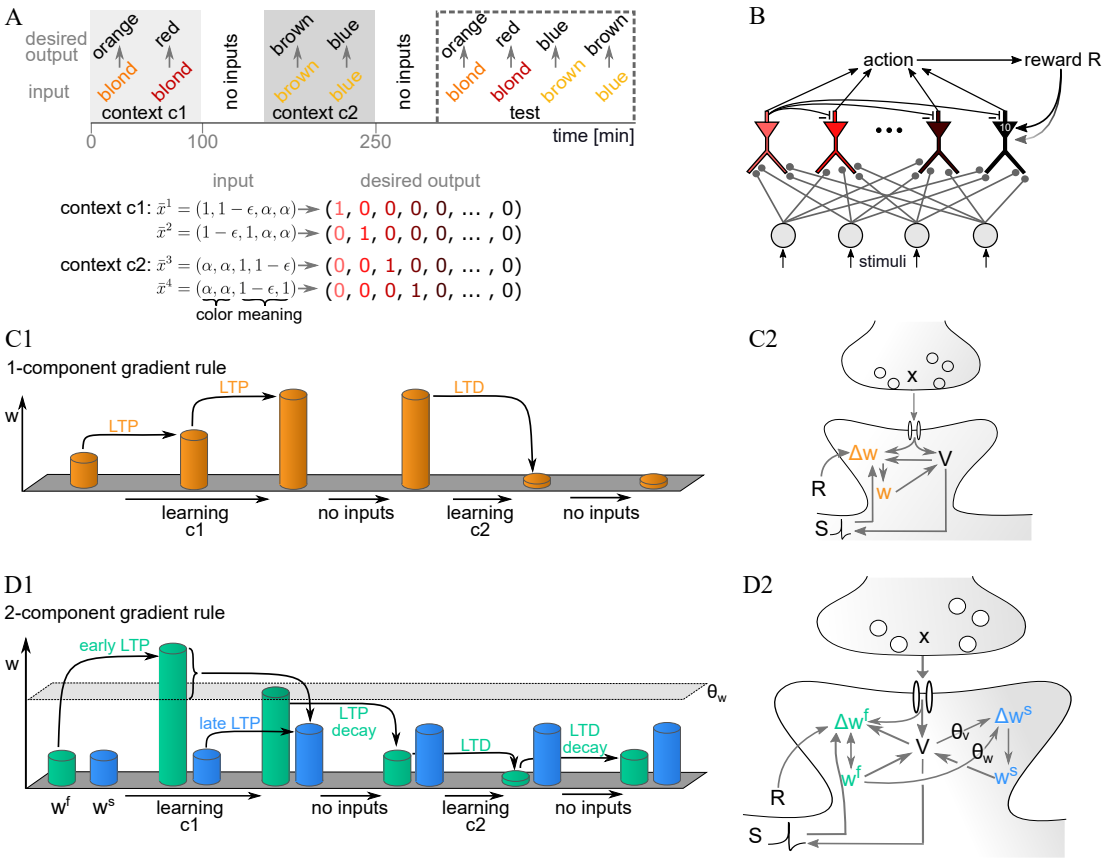

Figure 7: Classifying similar input patterns in different contexts fails with a 1-component plasticity model but succeeds with a 2-component model. (A) Input patterns have to be classified according to different criteria (e.g. color or letters), depending on the context. Patterns presented in one context may look similar (encoded by ϵ) and have strong overlaps across contexts (encoded by α). When being sequentially exposed to the contexts, learning in the second context interferes with the memory acquired in the first context. (B) Model network to solve the classification task using reward-based learning. The synapses from the input to the output layer are plastic (dots at line endings). Lateral inhibition in the output layer (only shown for first output neuron) enforces a winner-take-all mechanism according to which a first spike of a neuron in the output layer suppresses the possible spikes in the other output neurons right at the axonal initial segment. The winner defines an action, and a global reward signal ($R = 1$) is fed back to the network if the action is the desired one (otherwise $R = 0$). (C1) In the 1-component rule, long-term potentiation (LTP) induced in a first context c1 may be undone by long-term depression (LTD) in the second context c2, leading to forgetting of the previous weight change. (C2) The weight change Δw depends on the presynaptic activity x, the postsynaptic voltage V, a possible postsynaptic spike, and the reward signal R. (D1) In the 2-component rule, an early LTP that pushes the fast weight component w^f across a threshold θ_w can become consolidated in the slow weight component w^s upon a later strong postsynaptic activation (crossing a voltage threshold θ_V, late LTP). This slow component is protected against LTD in the second context. (D2) Dependencies of the weight changes for the 2-component rule (cf. Eqs 3.3 and 3.4).

The forgetting of the previous associations can be prevented if the fast weight changes are selectively consolidated in the slow component. The informative weight changes are identified when the two conditions for the consolidation are satisfied, a large absolute value of the fast component ($|w^f| > \theta_w$) and a large low-pass filtered voltage ($\bar{V} > \theta_V$). These are the filter criteria for saving the passive decay of the fast component ($-\lambda_w w^f$) into the non-decaying slow component (Eq. 3.4). Because the non-informative weight changes decay without consolidation, they do not contribute to erroneous activation of the postsynaptic neurons, and this prevents the interference-induced forgetting (Fig. 8A and B, blue).

Another benefit of the selective consolidation is that it filters out imbalanced plasticity inductions originating in the randomness of the pattern presentations and the noisy inputs. As a consequence, the slow components are less noisy than the fast components. During the delayed test period, when the fast components decayed back to baseline, we therefore obtain a better recall performance than immediately after the learning (Fig. 8A, black arrow). Hence, a break after learning improves the performance due to synaptic consolidation (that includes the decay of the last, stochastically triggered fast components), reminiscent to memory consolidation during sleep (Stickgold, 2005).

To investigate how learning depends on the pattern dissimilarity ϵ and context overlap α we repeated the experiment for many values $\epsilon \in [0, 1]$ and $\alpha \in [0, 1]$ and measured the average recall performance over all four input patterns during the test period (Fig. 8C). With the 1-component rule, already a small overlap between the patterns in the two contexts ($\alpha > 0.2$) erased the associations learned with the first context when the second context was presented. In contrast, with the 2-component rule, the performance for the first context stayed at the original level also after exposure to the second context, even when the overlap is maximal ($\alpha = 1$), assuming that the patterns are dissimilar enough within a single context ($\epsilon > 0.4$). For more similar patterns, both learning rules show a performance decline. Extending the learning experiment by more contexts with new but similar associations supports our previous results: with 1-component rule the network gradually forgets old memories when learning new ones, but with the 2-component rule no forgetting is observed (Fig. 8D).

Competition-agnostic synapses allow for high learning rates

Stochastic gradient rules have the intrinsic problem that the learning rate (η, effectively being a plasticity scaling factor) should be very small to estimate the gradient of the

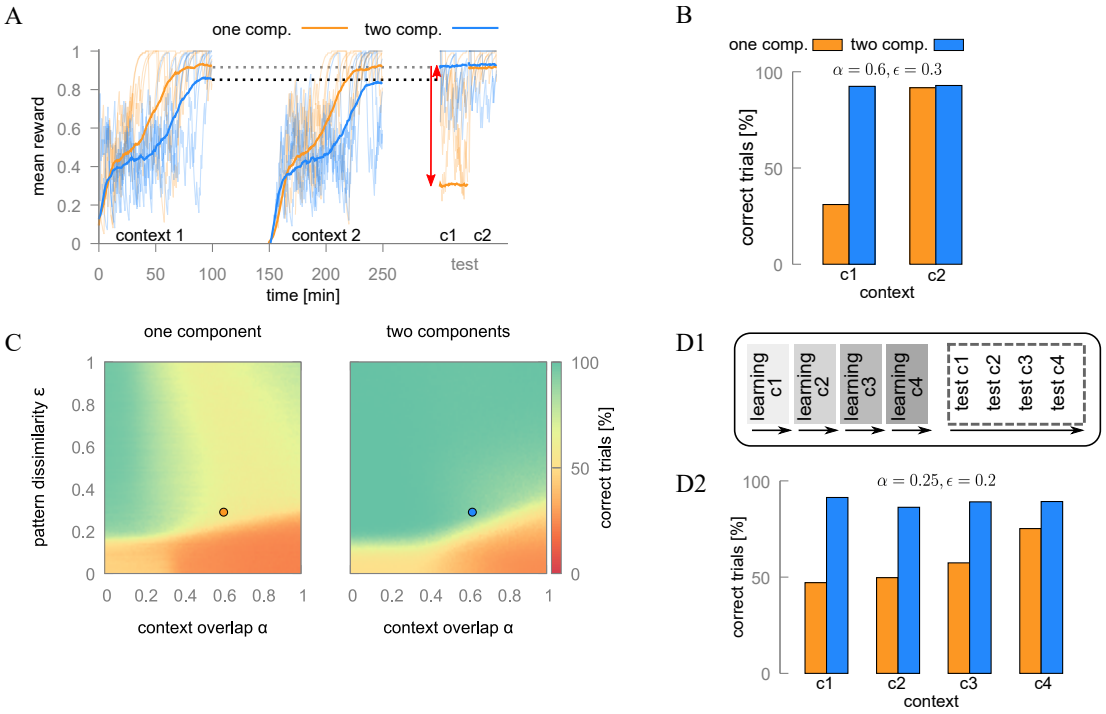

Figure 8: Learning similar associations in sequence causes catastrophic forgetting with a classical 1-component rule. (A) Five example traces of low-pass filtered reward $\bar{R}$ (light) and the average (bold) in two subsequent contexts, interleaved by a break without inputs. With the 1-component rule (orange) learned memories from the first context are forgotten after exposure to the second context (red downward arrow, c1), while the memories from the second context are still intact (c2). With the 2-component rule (blue) the memory performance for both contexts is improved after the second break due to the consolidation mechanisms (red upward arrow) for both contexts c1 and c2. (B) Average performance for the two contexts during the test period (c1 and c2 from panel A). (C) Average retrieval performance of all four memories for values ϵ and α. Dots show the parameters used in panel A and B. Catastrophic forgetting is mainly observed for strong context overlaps ($\alpha > 0.2$) when using the 1-component rule. (D) When appending more contexts, each with two associations to be learned, memory performance at the final test decreases the further back context lays.

expected reward. Learning can therefore not be accelerated by arbitrarily increasing the plasticity factor. In practice, increasing the plasticity factor for one pattern leads to overwriting of previously acquired weight changes induced by other patterns. Not all gradient rules are equally sensitive to this scaling problem. A learning rule that for small η would follow the reward gradient in a WTA network, is particularly vulnerable to a large increase of the plasticity factor. With a large η the voltage response in the output layer becomes close to an all-or-none structure, and this reduces the difference between the voltage and the stochastically generated WTA spike structure. Since it is this difference that drives plasticity, learning quickly saturates for the presented pattern, and so does it also for the similar other patterns.

As a remedy to the saturation problem for high plasticity factors, the difference term can be enhanced by suppressing the competition term in the voltage accessible to the synapses, without introducing a deteriorating sign change in the plasticity rule. Even though the output spike pattern is established based on the lateral competition, the learning rule ignores this and calculates the difference to the target as the somatic voltage would have been produced only by the feedforward input without lateral inhibition. This leads to a competition-agnostic learning rule that teaches the feedforward afferents to reach the target by them alone – it is a 'learn to do it yourself' signal (Fig. 9).

On the implementation level the separation of the lateral inhibition from the forward drive is achieved by deferring the WTA-mechanism with the lateral inhibition to the axon initial segment. The synapses will then predict the spike activity based on the unperturbed local dendritic voltage alone. In fact, because soma and dendrites represent a deep current sink viewed from the axon initial segment, the dendritic voltage is only barely influenced by the voltage in the axonal segment. Yet, the spikes generated there backpropagate to the dendrites where they can be read out by the synapses to identify the output spike rate (φ_i^{WTA}, see Eq. 3.3 and Fig. S1 and Suppl. Mat.). Inhibition of the axon initial segment that shunts the spike trigger mechanism has been found at various types of neurons (Somogyi et al., 1983; Douglas and Martin, 1990).

Consolidation by selecting informative learning events

High plasticity factors lead to fast learning, but in general also to fast forgetting by overwriting synaptic weights relevant for previous memories (Fusi et al., 2005). To keep the relevant information in the synaptic weight structure, our rule selects informative

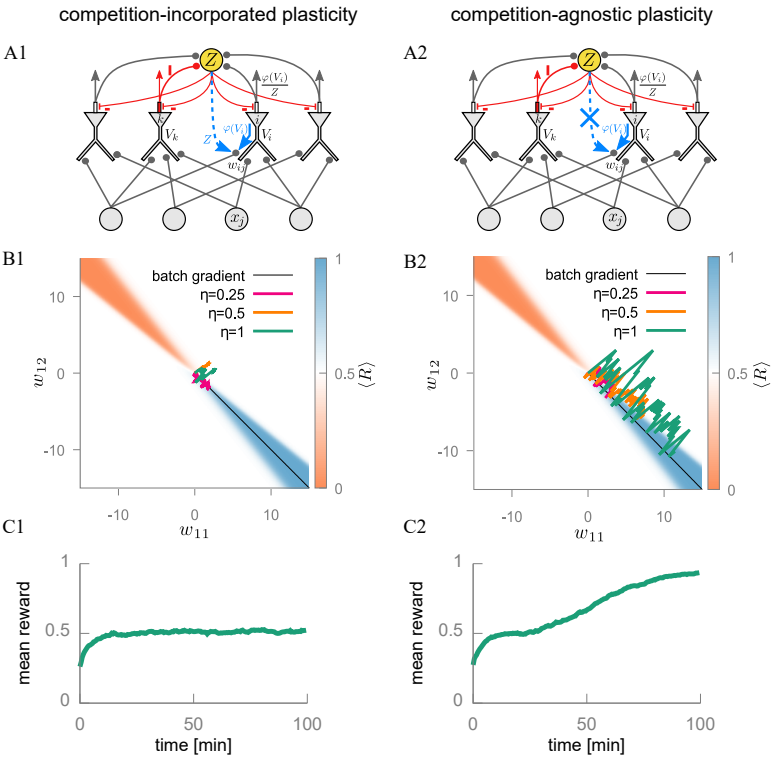

Figure 9: Competition-agnostic plasticity allows for fast learning without saturation. (A) The winner-take-all mechanism is implemented via global inhibition at the axon initial segment, triggered by the first spike of an output neuron (here neuron k, red) that inhibits via a global inhibitory neuron (orange) the others from spiking. This leads to the output rate normalized by the total relative firing rate Z across all output neurons (Eq. 3.11) for the competition-incorporated rule (A1, dashed blue arrow) but not in the competition-agnostic rule (A2, crossed dashed blue arrow). (B) Evolution of the weights gets stuck for the competition-incorporated (B1) but not for the competition-agnostic rule (B2). (C) Correspondingly, for high learning rates η performance gets stuck for the competition-incorporated (C1), but not for the competition-agnostic rule (C2).

events for writing over the quickly decaying fast weight components ($-\lambda_w w^f$) into the non-decaying components w^s. This is done by two criteria, one on the strength of the fast weight components themselves, and one on the average postsynaptic voltage (Fig. 10).

To motivate the first selection criterion, we note that changes of the fast component, summed across the pattern presentations in time, $w_{ij}^f(t) = \sum_{t'<t} \Delta w_{ij}^f(t')$, yield a more robust gradient estimate than individual updates Δw_{ij}^f themselves. One idea to improve the estimate is therefore to consider the fast weight components as samples of the gradient, and update the slow component not by Δw_{ij}^f but by the potentially better gradient estimates w_{ij}^f. These summed estimates are only better if they occurred often enough in the same direction, and this can be expressed by the criterion $\|w_{ij}^f\| \geq \theta_w$. This justifies the criterion for the slow weight update (cf. Eq. 3.4)

$$\Delta w_{ij}^s \propto w_{ij}^f \, \Theta(\|w_{ij}^f\| - \theta_w) \,. \tag{3.5}$$

To motivate the second selection criterion we consider the penalty term in the utility function that punishes strong voltage deflections on $|V_i|$, see Eq. 3.1. At the level of the gradient this leads to the forgetting term in the weight update of the form, $\Delta w_i^f \propto ... - \text{sign}(V_i) x$, where w_i^f is the weight vector targeting the postsynaptic neuron i. It subtracts away common components in the inputs x from w_i^f until, on average, this term vanishes, i.e. $\langle \text{sign}(V_i) \rangle \approx 0$, where $\langle . \rangle$ denotes the average over a long sequence of pattern presentations. As a consequence, w_i^f becomes roughly orthogonal to the average input pattern, $w_i^f \langle x \rangle \approx 0$, making the voltage $V_i^f = w_i^f x$ informative about pattern differences (Fig. 10B).

Imposing a voltage threshold for the transcription into the slow component selects the appropriate neuron for which synapses are consolidated. Applying the threshold condition on the temporal mean instead of the instantaneous voltage, favors the consolidation of associations that are encountered in the temporal vicinity of strong and lasting depolarizations. This becomes important when contexts are switched and patterns that were similar in the previous context now cause a strong but erroneous depolarizations that should not be consolidated (Fig. 10C, black dots). Fortunately, due to the fast learning in w^f, an incorrect strong depolarization will only arise a few times in a row, and by imposing the threshold onto a consistently high depolarization, only the correctly activated neurons become selected. This justifies the voltage-based

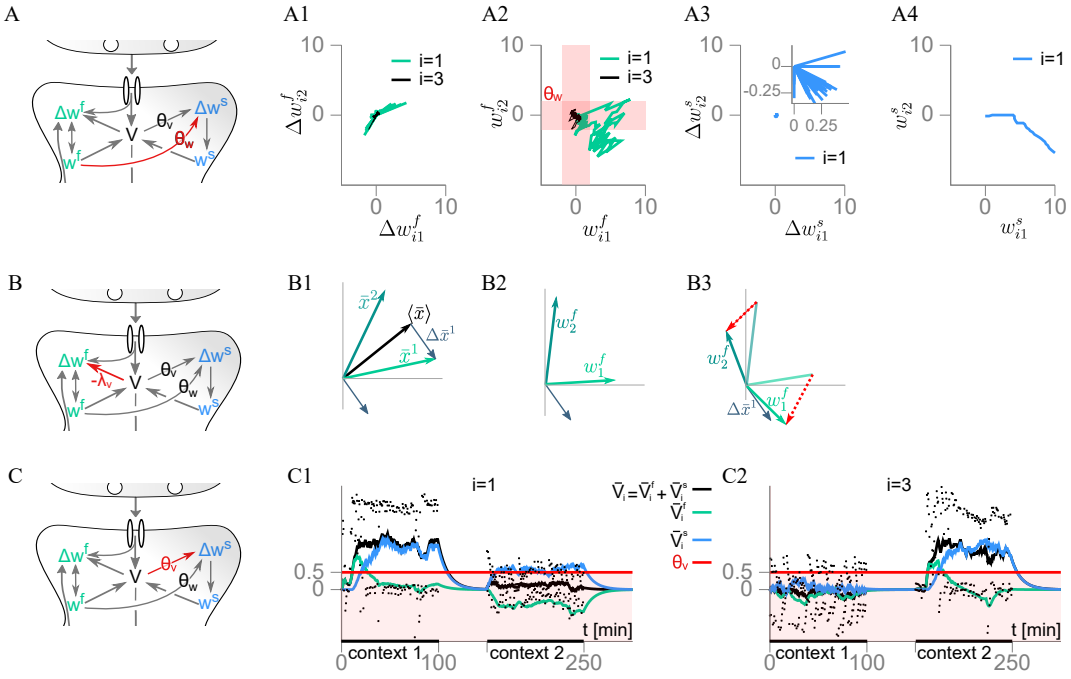

Figure 10: Selecting informative events for consolidation. (A) Only large enough fast weight components (the 'tagged' ones) are eligible for consolidation. (A1) The first 100 weight changes Δw_i^f induced by presenting patterns from class $\bar{x}^{(1)}$ and $\bar{x}^{(2)}$, for the two output neurons $i = 1$ (green) and $i = 3$ (black) and the presynaptic neurons $j = 1, 2$; see also Fig. 7A, B. (A2) The weight changes are summed up, and if a component exceeds the threshold $|w_{ij}^f| \geq \theta_w$ (i.e. if outside the corresponding red shaded stripe), it will be transcribed into the slow component. (A3) The slow weight changes induced by the suprathreshold fast components, $\Delta w_{ij}^s = \lambda_w w_{ij}^f$ if $|w_{ij}^f| \geq \theta_w$ (Inset: zoom-in). (A4) Summing up the slow weight changes Δw_i^s yields an improved gradient estimate w_i^s as compared to w_i^f, see A2. (B) The voltage penalty minimizes weight overlaps. (B1) The first two input classes $\bar{x}^{(1)}$ and $\bar{x}^{(2)}$ (projected to the first two dimensions) have a large overlap, as expressed by the large mean vector $\langle \bar{x} \rangle = \frac{1}{n} \sum_{i=1}^{n} \bar{x}^{(i)}$ leading to the decomposition $\bar{x}^{(i)} = \langle \bar{x} \rangle + \Delta \bar{x}^{(i)}$ (shown for $i = 1$). (B2) Without voltage penalty term in the energy function ($\lambda_V = 0$), the weight vectors w_i^f inherit some overlap from the input patterns. (B3) By penalizing systematic voltage deflections in one direction ($\lambda_V > 0$), the weight vectors become roughly orthogonal to the mean pattern and instead align with the deviations form the mean, $\Delta \bar{x}^{(i)}$. (C) Only events with strong enough mean depolarization (producing 'plasticity related proteins') are informative about the context and eligible for consolidation. (C1) The low-pass filtered voltage of neuron 1 ($\bar{V}_1$, black) exceeds the consolidation threshold (red) during context 1 as it should respond to input class 1, but it remains sub-threshold (red shaded region) during context 2 and hence the new context does not touch the previously consolidated weights of neuron 1. This would be different if the condition for consolidation were imposed on the non-averaged voltage (V_1, black dots), or if the voltage were driven by the slow weight component alone ($\bar{V}_1^s$, blue). Yet, because neuron 1 is never the correct winner, the fast component quickly learns that neuron 1 should be inactive in context 2 (averaged voltage induced by this fast component, V_1^f, green, is negative) and its consolidation is prevented. (C2) Same as in C1 but for neuron 3 that undergoes weight consolidation only in context 2.

consolidation criterion,

$$\Delta w_{ij}^s \propto w_{ij}^f \, \Theta(\bar{V}_i - \theta_V) \, . \tag{3.6}$$

Synaptic tagging, capture and consolidation

We next interpret the consolidation mechanisms in terms of biological quantities arising in the synaptic tagging & capture framework (Frey and Morris, 1997; Redondo and Morris, 2011). We postulate that whenever the absolute value of the fast weight component crosses a threshold, $|w^f| \geq \theta_w$, a synaptic tag in that synapse is set that remains active as long as $|w^f|$ is above threshold. Next, if the low-pass filtered postsynaptic voltage is above a threshold, $\bar{V}_i \geq \theta_V$, the translation of plasticity-related proteins (PRPs) is triggered that move up the dendritic tree. A synapse with an active tag captures PRPs and a fraction of the fast weight component w^f is transcribed into the slow component (Fig. 11A).

Based on this interpretation, we can reproduce results from synaptic tagging experiments *in vivo*, where electrodes were placed bilaterally in CA3 to stimulate independent synaptic inputs targeting a common population of postsynaptic neurons in CA1 (Fig. 11B, Shires et al. (2012)). A strong tetanus protocol applied on pathway 1 (such that the postsynaptic voltage $\bar{V}$ crosses threshold and strong postsynaptic firing is elicited, Methods) directly consolidates the early LTP in the that same pathway, provided the synaptic tag was activated (i.e. $|w_1^f| \geq \theta_w$, Fig. 11C, top). If pathway 2 is stimulated weakly (such that $\bar{V}$ does not cross θ_V, but postsynaptic spikes still triggered) the induced early LTP decays back to baseline again (Fig. 11C, middle). However, if within 30 minutes after (or before) the weak stimulation the second pathway is strongly stimulated such that now $\bar{V} \geq \theta_V$ (and PRPs are expressed), also the first pathway gets consolidated, provided there the tag is activated by the previous (or following) weak stimulation (Fig. 11C, bottom).

The synaptic tagging and consolidation experiments are astonishingly symmetric with respect to LTP and LTD (Sajikumar and Frey (2003, 2004), see Fig. 12). This may be caused by the fact that in all experiments plasticity was induced by presynaptic stimulations of excitatory afferents, although for LTD the overall stimulation was weaker. Whether LTP or LTD is induced may depend on the elicited postsynaptic spike rate. In our model, the sign of the plasticity induction depends on whether the somatic spiking activity is higher or lower than the dendritic prediction (yielding a positive or negative sign in the error term $(\varphi_i^{\mathrm{WTA}} - \varphi_i)$ entering in the update Δw_{ij}^f, see Eq. 3.3).

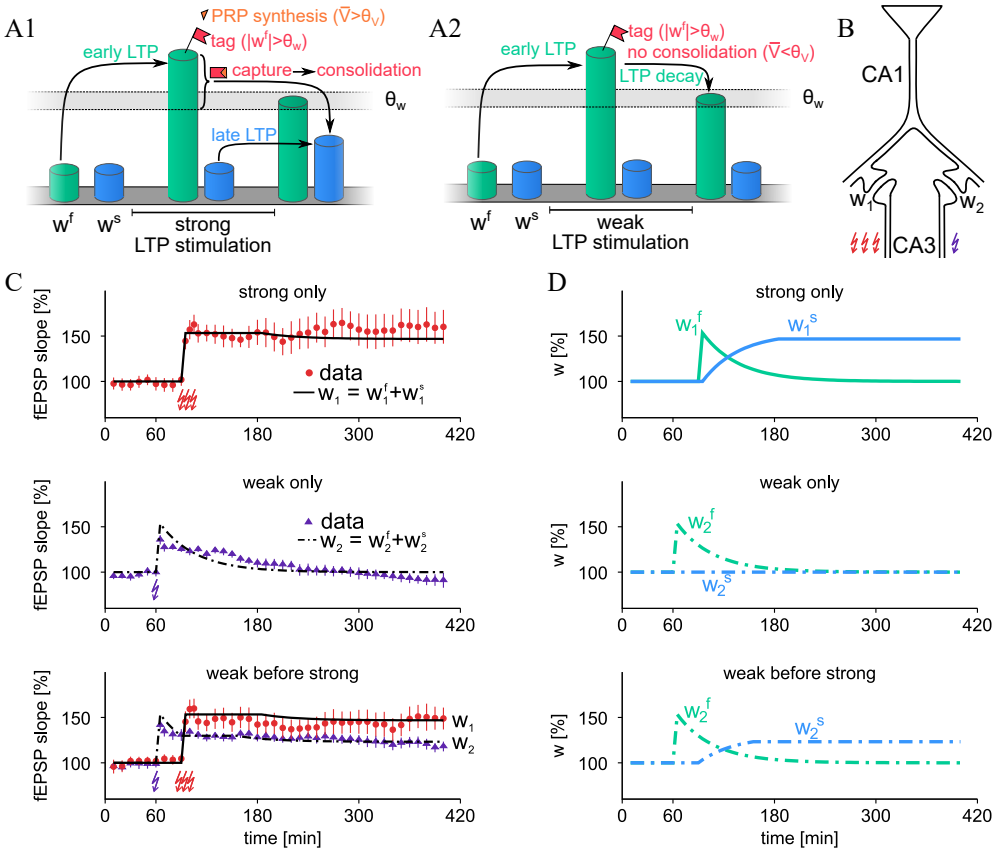

Figure 11: Model accounts for synaptic tagging *in vivo*. (A1) Early LTP in the fast weight component w^f, e.g. elicited by weak or strong tetanus, sets a local synaptic tag if w^f crosses a threshold, $|w^f| \geq \theta_w$. Strong and lasting postsynaptic depolarization triggers the synthesis of plasticity-related proteins (PRPs), $\bar{V} \geq \theta_V$. When a tag captures available PRPs (red & orange symbols), the fast weight component is consolidated in the slow component w^s (late LTP, Eq. 3.4). (A2) When no PRPs are available, the early LTP decays without consolidation (Eq. 3.3). (B) Experimental setup from Shires et al. (2012). Two afferent pathways projecting to a common CA1 pyramidal neuron (modeled with synaptic strengths w_1 and w_2) are stimulated by a strong tetanus (3 red flashes, such that $\bar{V}$ crosses threshold) and weak tetanus (1 purple flash, without θ_V crossing but still triggering postsynaptic spikes). (C) Top: Strong tetanization of one pathway only results in long-lasting changes in the data and the model. Middle: For weak tetanization of a single pathway, the synaptic strength (that is the sum of the fast and slow component) decays back to baseline. Bottom: If the weak tetanization of pathway 2 (purple) is followed by a strong tetanization of pathway 1 thirty minutes later (red), the decay of the synaptic strengths in pathway 2 is stopped. (D) Evolution of the separately shown fast and slow weight components of the strongly (solid, w_1) and weakly (dashed, w_2) stimulated pathways.

To describe the unsupervised character of the experiments, the reward prediction error $(R - \bar{R})$ was set to a constant value $(= 1)$. To reproduce the additional experimental data we assumed that, for simplicity, no postsynaptic spike activity was triggered during the LTD stimulations (in contrast to the LTP stimulations), while both the strong LTD and LTP stimulations generated a local low-pass filtered dendritic voltage that was suprathreshold (in contrast to weak LTD and LTP – weak LTP stimulations may trigger more global depolarization that still causes postsynaptic spiking).

Figure 12 summarizes the various experiments that were reproduced by our plasticity model of Eqs. 3.3 and 3.4. These experiments also include the so-called cross-tagging according to which a strong LTP protocol consolidates weakly induced LTD, and a strong LTD protocol consolidates a weakly induced LTP, be the strong protocol applied before or after the weak induction (Fig. 12C, D).

Fast learning deteriorates pattern discriminability when mixing easy and difficult tasks

So far we have shown how fast learning in different contexts with similar inputs is possible while reducing context interference. However, fast learning comes with a price. If in a given context classification tasks of unequal difficulties are mixed, learning slows down. This is indeed a well known phenomenon in psychophysics, called roving (Tartaglia et al., 2009b; Parkosadze et al., 2008) or interleaved learning (Flesch et al., 2018).

To exemplify these phenomena, we consider a perceptual bisection task with two different stimulus types characterized by offsets (ϵ) from the middle line chosen around larger (easy) or smaller (difficult) means $(\bar{\epsilon}_1 > \bar{\epsilon}_2$, Fig. 13A). After each stimulus presentation, participants have to tell in which direction they perceived the offset. To model this task we set up a network with 100 input neurons and two output neurons (Fig 13B) with an all-to-all connectivity from input to output neurons. The firing rates of either the first or second half of the input neurons are sampled around the mean $1 + \bar{\epsilon}_d$, while the other half is sampled around $1 - \bar{\epsilon}_d$ with $d = 1$ or 2. The mean offset $\bar{\epsilon}_1$ or $\bar{\epsilon}_2$ codes for the easy and difficult stimulus type, respectively. If the first half of the input neurons fire on average with a higher firing rate, the first output neuron needs to be activated to get reward $(R = 1)$, if the second half of the neurons fire on average more, the second output neuron needs to be activated. If incorrectly classified, reward is omitted $(R = 0)$.

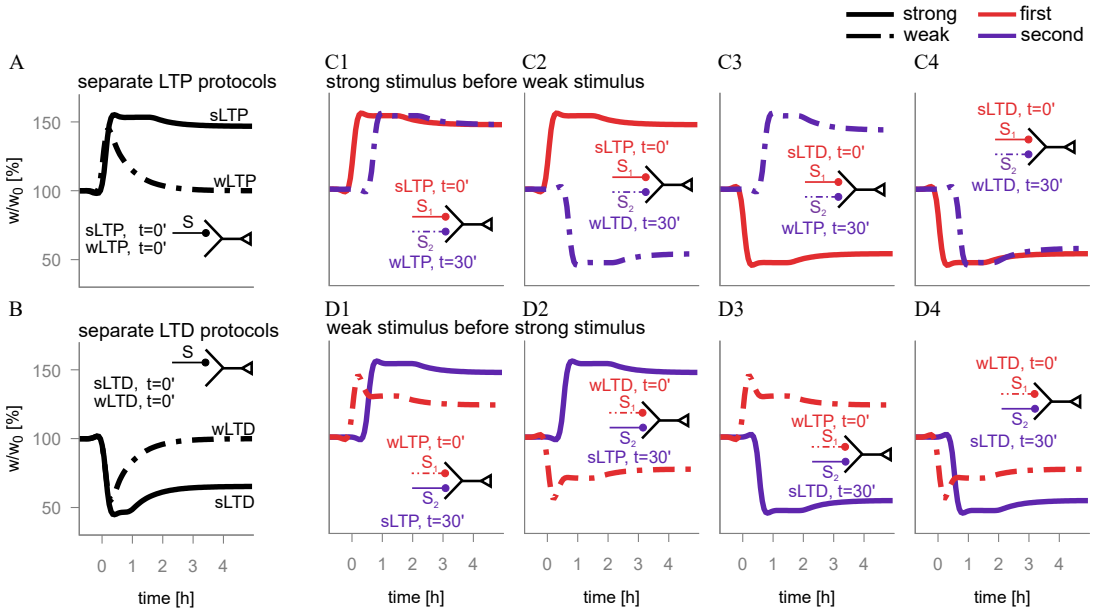

Figure 12: The model also accounts for LTD tagging and cross-tagging. For the corresponding experimental data see below. (A) Strong (solid) and weak (dashed) LTP stimulation applied in separate 'experiments' causes a strong lasting and a weaker decaying weight change, respectively (shown is $w = w^f + w^s$). (B) Same as in A, but for LTD protocols. (C) Strong stimulation applied 30 minutes before weak stimulation on different pathways rescues the weight decay after weakly induced early LTP and LTD. (D) Same as in C, but for the strong stimulation applied 30 minutes after the weak plasticity induction. Insets summarize the stimulation protocols that have also been applied in the experiments. A: Frey and Morris (1997), B: Sajikumar and Frey (2003), C1: Frey and Morris (1997), C2-C4: Sajikumar and Frey (2004), D1: Frey and Morris (1998), D2-D4: Sajikumar and Frey (2004)

If the easy stimuli are presented in a first block, followed by the difficult stimuli in a second block, learning is possible for both the 1- and 2-component rules (Fig. 13C, similar results for reversed block order, not shown). The same learning rates were used as the ones optimized for the context learning task (cf. Fig. 7). If the easy and difficult stimuli are presented randomly interleaved, the 2-component rule performs less well, unlike the 1-component rule (Fig. 13D), analogously to human behaviour (Parkosadze et al., 2008; Clarke et al., 2014). The performance decrease with roving (mixing) occurs because weight changes, in response to an easy stimulus, point further away from the optimal weight due to the larger variance of presynaptic rates around the mean for these easy stimuli. When presented in a block, the large deviations are corrected during the learning of the difficult stimuli that themselves show lower variance (inset of Fig. 13C). If the easy and difficult stimuli are mixed, however, the large variances remain throughout learning and prevent a convergence towards the optimal weight (inset of Fig. 13D). Weight updates also become smaller during learning because the modulating reward prediction error typically becomes small ($\langle R - \bar{R} \rangle \approx 0$ as $R = 1$ most of the time). Learning in the roving condition with the 1-component rule is better because the learning rate optimized for the previously studied context learning is smaller, paying out for roving, but also leading to catastrophic forgetting.

Interferences between stimulus types diminish when the two stimulus types become more similar in difficulty. Since reinforcement learning follows the principle of exploration and exploitation, a certain level of exploration in the weight space, for instance caused by mixing of different stimulus types, may be beneficial. This is in fact what occurs when the difficulties for the two stimulus types become similar and they are randomly mixed to boost stochastic weight fluctuations (Fig. 13E, star). Consequently, the high variance weight updates from the easier stimuli, that would hinder discrimination for the difficult stimuli, now rather improve the performance in the roving scenario for the 2-component rule as compared to block learning (Fig. 13F, blue).

Discussion

We demonstrated that catastrophic forgetting in continual learning tasks can be prevented by using two synaptic weight components, a fast and a slow one. The dynamics of the fast component are derived from a stochastic gradient procedure on a reward-based utility function. The utility incorporates the expected reward and a sparsity constraint on both, the postsynaptic voltage and the fast component. Since the change

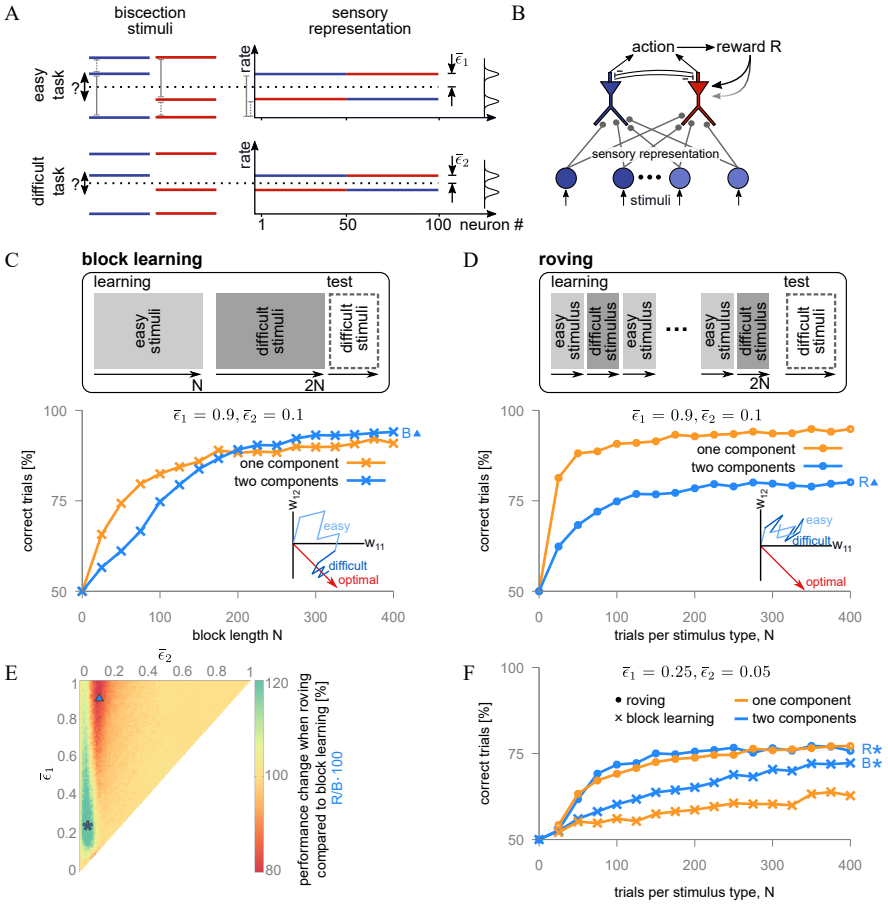

Figure 13: 2-component rule accounts for roving. (A) In a bisection task participants have to decide whether the midline is offset to the top or bottom (left). In the model (right), either the first (1-50, blue) or second (51-100, red) half of the sensory neurons have a higher firing rates. Two stimulus types are considered, an easy one with a large midline offset ($\bar{\epsilon}_1$, right top) and a difficult one with a small offset ($\bar{\epsilon}_2$, right bottom). (B) The model network has 100 sensory neurons and two action neurons (coding for the up- or down shift from the midline) that mutually inhibit each other with their 'first spike'. If the first (second) half of the input neurons have a higher mean firing rate, the first (second) output neuron has to be active to get global reward $R = 1$, otherwise $R = 0$. (C) If the easy and difficult stimulus types are learned block-wise (N trials per block), both the 1- and 2-component rules learn the task. Performance is shown only for the difficult stimulus type. Inset: Example of a 2-component weight trajectory. During the first block of easy stimuli (light blue) the weight jumps strongly around the optimal weight (red), and converges during the block with the difficult stimuli (dark blue, $N = 7$). (D) Easy and difficult bisection stimuli are presented in random order (roving). Compared to block learning, performance for the difficult stimulus type improves more strongly than with the 1-component rule and drops with the 2-component rule, in accordance with the psychophysical experiments. Inset: The large fluctuations caused by weight updates in response to the more varying easy stimuli (light blue) cannot be corrected by the interleaved difficult stimuli (dark blue) and the fluctuations around the optimal weight (red) remain larger as compared to the block-wise learning in C. (E) Performance change for roving compared to block-learning, for any pair of task difficulties $(\bar{\epsilon}_1, \bar{\epsilon}_2)$, using the 2-component rule. (F) If the difficulty of both types is increased while becoming more similar, both the 1- and 2-component rules predict that roved learning outperforms block learning.

of the fast component follows the utility gradient, this weight component maximizes the expected reward under the constraint of small weight modifications and small voltage deflections. The sparsity constraint on the voltage implies an orthogonalization of the pattern representation, and the sparsity constraint on the weight implies a passive decay of the fast component. Both the orthogonalization and the decay help to reduce interference when new associations for similar stimuli have to be learned. To not forget the old associations, part of the decaying fast weight component is consolidated in a slow weight component. An additional selection process on the strength of the fast weight component and the amplitude of the voltage deflections insures that only the 'informative' events are memorized. This selection process nonlinearly amplifies the two sparsity constraints on the weight and the voltage: only when the fast weight component exceeds a threshold, and only when the averaged postsynaptic voltage is simultaneously above a voltage threshold, the fast component is consolidated.

We show that the 2-component gradient rule, unlike 1-component rule(s), prevents catastrophic forgetting across subsequent contexts in which classification tasks with similar input patterns but different outputs have to be learned. Intriguingly, after a pause (of roughly an hour biological time) the test performance of the model network is even better than before the pause. This is explained by the decay of the fast component that leads to a forgetting of the sample-specific noise, while the class-relevant information is consolidated akin to the consolidation and the semantization of memory during sleep (Stickgold, 2005). We also show that the consolidation criteria on the strong weight component and the large voltage deflection can be interpreted in terms of synaptic tagging and capture (Frey and Morris, 1997; Redondo and Morris, 2011). The synaptic tag & capture hypothesis posits that a weak synaptic stimulation that triggers a tag during the induction of early LTP/LTD – in our model the threshold-crossing by potentiating/depressing the fast component – is captured by a strong stimulation within some minutes before or after – in our model the threshold-crossing of the low-pass filtered postsynaptic voltage causing consolidation. Notably, our model implicates that reward and consolidation arise also from weak LTP to strong LTD and from weak LTD to strong LTP, just in the same way as in the real experiments dopamine is implicated and cross-tagging between potentiation and depression is observed.

The strategy of consolidating only large weight components associated with high voltages leads to an overweighting of strong stimuli. A similar overweighting of large magnitudes is also observed when humans solve a categorical decision task under limited cognitive resources , e.g., induced by time constraints (Spitzer et al., 2017). In eco-

nomics, the selective integration that discards low-value options is known as economic irrationality (Tsetsos et al., 2016), reminiscent of discarding low-activity events when learning associations across time. In a similar way as economic irrationality remains optimal under constraints, our learning rule remains hill-climbing on the utility function in the competition-agnostic version that discards winner-take-all information among the output neurons. Upon reward, the competition-agnostic rule considers the postsynaptic activity as a target, irrespectively of how the activity is produced, and by this virtue prevents an early saturation of learning even for very high learning rates. Turning reinforcement learning into a target-based learning appears as a trick to speed up learning. The ultra-fast learning rate for the competition-agnostic learning rule is only possible with the 2-components that enable the selective decay of the 'non-informative' weight changes while consolidating the 'informative' ones.

Fast learning with selective consolidation comes with a price. Fast learning is enabled by a transient synaptic memory buffer that has an intrinsic time constant. Its downside is exposed when intermixing tasks of unequal difficulties that fill up the buffer with unspecific information. Randomly mixing easy and difficult stimuli ('roving') in a classification task hampers learning with the 2-component rule, but not with the rate-optimized 1-component rule. A similar phenomenon is observed in perceptual discrimination tasks when bisection stimuli of unequal difficulties are randomly intermixed (Otto et al., 2006; Parkosadze et al., 2008; Clarke et al., 2014). In our model performance degrades because the easy stimuli simultaneously trigger suprathreshold weight changes and suprathreshold voltages that are then prematurely consolidated, even when these changes are not accurate enough to correctly classify the difficult stimuli. A previous model was explaining the roving phenomenon by an imprecise critic in the context of reinforcement learning (Herzog et al., 2012). When mixing tasks of different difficulties, the critic induces a synaptic drift towards either LTD for the difficult and LTP for the easy task, or vice verse. In our model, learning is hampered not due to a systematic drift, but due to the early consolidation of erratic weight changes triggered by the easy task.

The model we presented can be seen as one of many attempts to fight catastrophic forgetting during continual learning with sophisticated synaptic consolidation mechanisms. The synaptic cascade model (Fusi et al., 2005; Benna and Fusi, 2016; Kaplanis et al., 2018) postulates hidden states that, different from our model, do not contribute to the effective synaptic strength. Others, more globally defined learning rules have been suggested that keep synaptic parameter changes small if the same parameters were

previously engaged in a performance increase (Kirkpatrick et al., 2017; Zenke et al., 2017). A yet more global criterion for learning rate adaptation has been suggested that prevents catastrophic interference using conceptors that seek to orthogonalize the pattern representation in different contexts (He and Jaeger, 2018).

Our gradient-based 2-component plasticity model yields a synaptic explanation of well-known learning strategies. (1) Context switches should not be faster than the decay time of early LTP/LTD (some ten minutes) to prevent interferences induced by the not-yet-decayed fast weight components. (2) A learning break without stimulus exposures helps to semanticize memories through forgetting of non-systematic, random stimuli that did neither trigger suprathreshold fast component changes nor suprathreshold neuronal activities. (3) Mixing tasks of unequal difficulties may block the learning of the difficult task by premature weight consolidations triggered during the easy task. On the synaptic level we predict that consolidation thresholds are dynamically adapted to the statistics of the fast weight- and postsynaptic voltage-distributions, preventing catastrophic forgetting in a volatile environment, and endowing individual synapses in deep networks with a local consolidation mechanism.

Methods

Throughout the paper a single-layer network with an all-to-all connectivity between input and output layer is used. Each postsynaptic neuron receives from each presynaptic neuron input $x_j(t)$ through synapses with weight $w_{ij}(t)$. The postsynaptic voltage $V_i(t)$ can elicit a spike which mediated by strong lateral inhibition suppresses all other postsynaptic neurons from spiking upon onset of a new input. The probability that a spike is observed in neuron k is given by the normalized Poisson firing rate,

$$P_w(k|x) = \frac{\varphi(V_k)}{\sum_i \varphi(V_i)}, \quad V_i(t) = \sum_j w_{ij}(t) x_j(t)\,, \tag{3.7}$$

where φ denotes the logistic function, $\varphi(x) = \varphi_{\max}(1+\exp(-(x-a)/\beta))^{-1}$. The strong lateral inhibition generates a WTA dynamics. As a result one gets for each presented input pattern an output vector of the form $y^k(t) = (0,\ldots,0,1,0,\ldots,0)$ where 1 is at position k. The environment is evaluating the output and issues a global binary reward signal R to the network. For each input class there is exactly one output which results in $R = 1$, all other responses give $R = 0$.

The competition-incorporated rule is gradient ascent

Using the 2-component rule, each synaptic weight consists of two components, a fast component $w_{ij}^f(t)$ and a slow component $w_{ij}^s(t)$. The total synaptic weight $w_{ij}(t)$ is the sum of the fast and the slow component, $w_{ij}(t) = w_{ij}^f(t) + w_{ij}^s(t)$.

The update of the fast component, Eq. 3.3, is derived from the gradient ascent rule of the utility function in Eq. 3.1. The expected reward is given by

$$\langle R \rangle = \sum_l \sum_k P(x^l) P_w(k|x^l) R_{x^l,k} \,,$$

where the sums run over all input pattern indices l and potential winning units k. $R_{x^l,k}$ denotes the value of reward given for input-response pair (x^l, k). The gradient of $\langle R \rangle$ with respect to w_{ij} can be calculated by using the log trick,

$$\frac{\partial \langle R \rangle}{\partial w_{ij}} = \sum_l \sum_k P(x^l) P_w(k|x^l) R_{x^l,k} \frac{\partial}{\partial w_{ij}} \log P_w(k|x^l) \,.$$

The derivative term can be computed with $P_w(k|x^l)$ from Eq. 3.7 as

$$\frac{\partial}{\partial w_{ij}} \log P_w(k|x^l) = \left(\delta_{ik} - \frac{\varphi(V_i)}{\sum_i \varphi(V_i)} \right) \frac{\varphi'(V_i)}{\varphi(V_i)} x_j^l \,,$$

where k being the winning unit and the Kronecker delta δ_{ij} is 1 if $i = k$ and 0 otherwise. To minimize the variance of the reward estimate we center the effective reward around its approximate mean $\bar{R}$,

$$\frac{\partial \langle R \rangle}{\partial w_{ij}} = \sum_l \sum_k P(x^l) P_w(k|x^l) (R_{x^l,k} - \bar{R}) \left(\delta_{ik} - \frac{\varphi(V_i)}{\sum_i \varphi(V_i)} \right) \frac{\varphi'(V_i)}{\varphi(V_i)} x_j^l \,. \tag{3.8}$$

The terms involving $\bar{R}$ do not originate from the derivative but since they add to 0 they can be added for convenience. If (x^l, k)-pairs are sampled with probability $P(x^l)P_w(k|x^l)$, the learning rule based on gradient ascent of U is given by

$$\Delta w_{ij}^f \sim (R - \bar{R}) \left(\delta_{ik} - \frac{\varphi(V_i)}{\sum_i \varphi(V_i)} \right) \frac{\varphi'(V_i)}{\varphi(V_i)} x_j^l - \lambda_V \mathrm{sign}(V_i) x_j^l - \lambda_w w_{ij}^f \,.$$

We next assume that the neuronal transfer function $\varphi(V_i)$ saturates for large ar-

guments at $\varphi_{\max}$. By setting $\varphi_i^{\mathrm{WTA}} = \delta_{ik}\varphi_{\max}$ (i.e. $\varphi_i^{\mathrm{WTA}} = \varphi_{\max}$ if i is the winner, $i{=}k$, and $\varphi_i^{\mathrm{WTA}} = 0$ for $i{\neq}k$), $\eta_i = \eta\varphi_i'/(\varphi_i\varphi_{\max})$ for some global learning rate η, and $Z = \sum_i \varphi_i/\varphi_{\max}$, we end up with the learning rule

$$\Delta w_{ij}^f = \eta_i(R-\bar{R})\left(\varphi_i^{\mathrm{WTA}} - \frac{\varphi(V_i)}{Z}\right)x_j^l - \lambda_V \mathrm{sign}(V_i)x_j^l - \lambda_w w_{ij}^f\,. \tag{3.9}$$

Time-averaged quantities as $\bar{R}$ and $\bar{V}_i$ are calculated according to $\bar{R}(t) = \sum_{t'=0}^{t} R(t')e^{-(t-t')/\tau_R}$, and analogously for $\bar{V}_i$. We set $\tau_R = \tau_V = 10$ trials, with a trial duration of 30 s. The results do not qualitatively change when choosing half or twice as large time constants. The fast and slow weight components are updated every 30 s biological time (including the 'breaks' when $R = 0$ and $x = 0$ by definition).

The competition-agnostic rule is hill-climbing

The reward-based component of the competition-agnostic learning rule,

$$\Delta w_{ij}^{\mathrm{ca}} = \eta_i\,(R-\bar{R})\Big(\varphi_i^{\mathrm{WTA}} - \varphi_i\Big)\,x_j\,, \tag{3.10}$$

is stochastic hill-climbing on the expected reward $\langle R\rangle$. This is because the reward-based component of the competition-incorporated learning rule

$$\Delta w_{ij}^{\mathrm{ci}} = \eta_i\,(R-\bar{R})\Big(\varphi_i^{\mathrm{WTA}} - \frac{\varphi_i}{Z}\Big)\,x_j$$

according to Eq. 3.8 is stochastic gradient ascent on $\langle R\rangle$, and the two updates are always within 90°, $\Delta w^{\mathrm{ca}}\,\Delta w^{\mathrm{ci}} > 0$. The latter scalar product is positive because for both rules the winner is the same, and $\mathrm{sign}\Big(\varphi_i^{\mathrm{WTA}} - \varphi_i\Big) = \mathrm{sign}\Big(\varphi_i^{\mathrm{WTA}} - \frac{\varphi_i}{Z}\Big)$ for all i. Hence, the averaged update vectors are still within 90°, while the averaged competition-incorporated update points in the direction of the reward gradient, $\langle\Delta w^{\mathrm{ci}}\rangle \propto \frac{\partial\langle R\rangle}{\partial w}$. Adding the two vectors Δw^{ca} and Δw^{ci} to the same penalty gradient $\{-\lambda_V\mathrm{sign}(V_i)x_j^l - \lambda_w w_{ij}^f\}_{ij}$ does only decrease the angle. We conclude that the competition-agnostic rule is hill-climbing on the utility function U.

One-component rule

For comparison, all simulations were repeated with a 1-component rule that consisted of the fast component $w = w^f$ only (Eq. 3.3), but with $\lambda_w = 0$ and hence $w^s =$

0. Parameters of the 1-component rule were optimized independently from the 2-component rule.

Associative task

In the associative task, we consider input classes $\bar{x}^l$ with each class requiring its desired output defined by the exclusive activity of a specific output neuron. Input patterns, which are defined as noisy samples x^l around $\bar{x}^l$, are presented each after the other. The goal is to change the synaptic weights such that for each input class the desired output is obtained. For a pattern of the class l the instantaneous Poisson firing rate of the presynaptic neuron j is

$$x_j^l = \begin{cases} 1+\xi & \text{if } j = l \\ 1-\epsilon+\xi & \text{if } j = l{+}1 \text{ for } l \text{ odd, and } j = l{-}1 \text{ for } l \text{ even} \\ \alpha+\xi & \text{else .} \end{cases}$$

where ξ is independent Gaussian noise with zero mean. Two input classes form together a context. Contexts are learned in succession without repetition. Within a context, the input patterns of the two classes are presented repeatedly and in random order. The number of presynaptic neurons in the network corresponds to the number of classes, the number of postsynaptic neurons is fixed at 10. Reward is given if for input pattern x^l output neuron l is exclusively active ($R = 1$). Otherwise the reward signal is omitted ($R = 0$). After learning a context, a phase with a null input but unchanged dynamics was on. Performance of learning is evaluated during a test period at the end of the task in which all input patterns are presented again though now without synaptic plasticity. Performance is measured by the percentage of correct trials during the test period. Parameters were set to $\eta = 1$, $\lambda_w = 0.012$, $\lambda_V = 0.45$, $\theta_w = 2$, $\theta_V = 2$.

Tagging experiments

In tagging experiments a network with five presynaptic neurons and one postsynaptic neuron was used. Only one of the presynaptic neurons was stimulated at a time. Simulation were run with four stimulation protocols: strong LTP, weak LTP, strong LTD and weak LTD. Stimulation strength was regulated by the input duration. Strong LTP and LTD lasted 15 minutes biological time and weak LTP and LTD lasted 5 minutes biological time. The stimulation amplitude was kept fixed for all four protocols.

For LTP stimulations the teacher term φ^{WTA} in Eq. 3.3 was constantly set to $\varphi_{\max}$ whereas it was set to 0 for LTD stimulations. In all tagging experiments, $R - \bar{R}$ was set to 1. If two protocols were paired, the second stimulation set in 30 minutes after the first stimulation on a second presynaptic neuron.

In experiments, late-LTP was induced *in vivo* by a strong tetanization consisted of three trains of 50 pulses at 250 Hz, with a 5-min intertrain interval; weaker induction of LTP was investigated by a weak tetanus consisting of 1 train of 50 pulses at 100Hz (Shires et al., 2012). For LTD, see Sajikumar and Frey (2003): Late-LTD was induced using a low-frequency stimulus protocol (LFS) of 900 small bursts (one burst consisted of three stimuli at 20 Hz, interburst interval 1 s, i.e. f=1 Hz, stimulus duration 0.2 ms per half-wave, a total of 2700 stimuli). This stimulation pattern produced a stable long-term depression in vitro for at least 8 h. In experiments in which a weaker induction of LTD was investigated, a transient early-LTD was induced using LFS consisting of 900 pulses (1 Hz, impulse duration 0.2 ms per half-wave, a total of 900 stimuli).

Roving

To explore the phenomenon of roving a bisection task with two stimulus types were used, an easy and a difficult type. The bisection task consisted of two input patterns which needed to be discriminated. The stimulus types only differed in the difficulty to discriminate these input patterns. For each input pattern, half of the population had mean Poisson firing rate of $1 + \bar{\epsilon}$, the other half of $1 - \bar{\epsilon}$, where $\bar{\epsilon}$ is a stimulus type specific parameter. Each input was modulated by an additive Gaussian noise ξ,

$$x_j^1 = \begin{cases} 1 + \bar{\epsilon} + \xi_j & j = 1...50 \\ 1 - \bar{\epsilon} + \xi_j & j = 51...100 \end{cases}$$

$$x_j^2 = \begin{cases} 1 - \bar{\epsilon} + \xi_j & j = 1...50 \\ 1 + \bar{\epsilon} + \xi_j & j = 51...100\,. \end{cases}$$

The network consisted of 100 presynaptic neurons and two postsynaptic neurons representing the two possible actions in the experiment of pressing left or right. Reward was given if neuron 1 was activated and the first half of the input population had a higher activity or if neuron 2 was activated and the second half of the population had a higher activity. Two learning scenarios were considered: block learning and roving. In the block learning scenario, easy input patterns were first trained followed by the

training of the difficult patterns. During roving, patterns of the easy and difficult type were alternated presented in random order. The total number of pattern presentations was for both learning scenarios equal. In the test period, which followed a phase with null inputs, the percentage of correct trials using patterns of the difficult type was evaluated.

Simulation details

Simulations were run on the HPC cluster of the University of Bern. For the associative tasks (Fig. 8), all parameters were separately optimized for the 1- and 2-component rule with a multivariate bisection method. The objective function was the sum of the expected reward after 200 simulation runs. Each optimization run was performed with new random values of α and ϵ and a random number of contexts between two an six. Keeping parameters fixed, simulations were done for combinations with $\alpha \in [0, 1]$ and $\epsilon \in [0, 1]$ using step size 0.01. The percentages of correct trials are average values over 200 simulations.

For the tagging experiments (Fig. 11 and 12), parameters were adapted to best fit the *in vivo* results by Shires et al. (2012) reproduced in Figure 11. For additional protocols and pairings shown in Figure 12 the parameters were kept unchanged.

In the roving scenario (Fig. 13), the same parameter were used as for the associative task. In panels 13C-E each point represents a separate simulation set. Shown values are averages over 200 simulations.

Acknowledgments

This work was supported by the Swiss National Science Foundation (SNSF, Sinergia grant CRSII2-147636 to MH and WS), and by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 785907 (Human Brain Project, SGA2).

Supplementary Material

Competition-incorporated plasticity quickly saturates

To formally capture the described problem of large plasticity factors we consider the exact gradient of the expected reward that includes a global normalization Z (Methods),

$$\frac{\mathrm{d}\langle \bar{R} \rangle}{\mathrm{d}w_{ij}^f} \propto \left\langle \eta_i \left(R - \bar{R}\right) \left(\varphi_i^{\mathrm{WTA}} - \frac{\varphi_i}{Z}\right) x_j \right\rangle \text{ , with } Z = \sum_i \varphi_i / \varphi_{\max} \tag{3.11}$$

for $\varphi_i = \varphi(V_i)$. As compared to the approximate stochastic gradient (Eq. 3.3), the normalization by Z appears. This normalization makes the ratio φ_i/Z identical to $\varphi_{\max}$ times the probability that neuron i is the winner (with k being the index of the winner neuron and $\varphi_i^{\mathrm{WTA}} = \varphi_{\max}$ for $i = k$ and 0 else). If for a synaptic update a sample of the right-hand side of Eq. 3.11 is used with a high plasticity factor, $\Delta w_{ij}^f = \eta_i \left(R - \bar{R}\right) \left(\varphi_i^{\mathrm{WTA}} - \varphi_i/Z\right) x_j$, the winner probability quickly becomes all-or-none, $\varphi_i/Z \approx \varphi_i^{\mathrm{WTA}}$, and learning will saturate. Leaving out the normalization by Z, however, keeps Δw_{ij} away from 0, without changing the sign of Δw_{ij}. Overall, this yields a plasticity rule (Eq. 3.3) that allows for speeding up the learning by increasing η_i. On average, the update still stays within 90° of the true gradient as the synaptic bias does not invert the sign of the weight change, and thus the competition-agnostic rule remains hill-ascending on the utility function (Methods, Fig. S1).

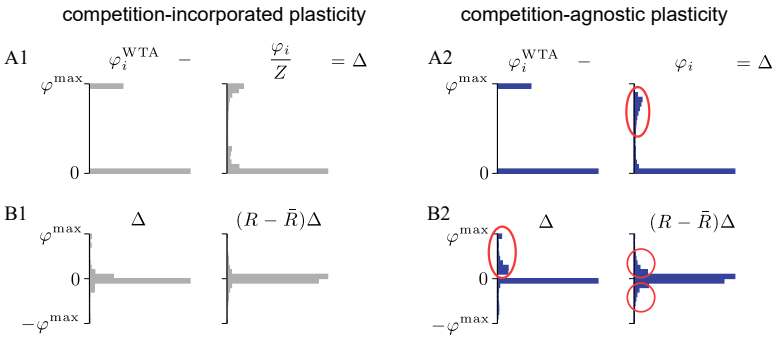

Figure S1: Competition-agnostic plasticity create non-vanishing learning signals. (A1, B1) Learning stalls for high η since the error in the competition-incorporated rule, $\Delta = \varphi_i^{\mathrm{WTA}} - \varphi(V_i)/Z$, mostly vanishes. This arises because for fast learning the output quickly converges to a one-hot representation, such that $\varphi(V_i)/Z$ gets close to 1 (for $i = k$) or 0 (for $i \neq k$), as this is by construction the case for φ_i^{WTA}. (A2, B2) This is different for the error term $\varphi_i^{\mathrm{WTA}} - \varphi(V_i)$ of the competition-agnostic rule where $\varphi(V_i)$ is not normalized (A2, oval), leading to non-vanishing errors Δ (B2 left, oval). Although the error term Δ is now transiently unbalanced, the product $(R - \bar{R})(\varphi_i^{\mathrm{WTA}} - \varphi(V_i))$ across synapses becomes balanced again (B2 right, two circles) because $(R - \bar{R})$ is balanced, preventing a deteriorating weight drift. In fact, the competition-agnostic rule remains hill-climbing (Methods).

Bibliography

Abraham, W. C. and Robins, A. (2005). Memory retention - The synaptic stability versus plasticity dilemma. *Trends in Neurosciences*, 28(2):73–78.

Aljundi, R., Babiloni, F., Elhoseiny, M., Rohrbach, M., and Tuytelaars, T. (2018). Memory Aware Synapses: Learning what (not) to forget. In Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, W., editors, *The European Conference on Computer Vision (ECCV)*, pages 144–161. Springer.

Barrett, A. B., Billings, G. O., Morris, R. G. M., and Van Rossum, M. C. W. (2009). State based model of long-term potentiation and synaptic tagging and capture. *PLoS Computational Biology*, 5(1).

Benna, M. K. and Fusi, S. (2016). Computational principles of biological memory. *Nature Neuroscience*, 19(12):1697–1706.

Bosch, M., Castro, J., Saneyoshi, T., Matsuno, H., Sur, M., and Hayashi, Y. (2014).

Structural and molecular remodeling of dendritic spine substructures during long-term potentiation. *Neuron*, 82(2):444–459.

Brea, J., Urbanczik, R., and Senn, W. (2014). A Normative Theory of Forgetting: Lessons from the Fruit Fly. *PLoS Computational Biology*, 10(6):1–9.

Carpenter, G. A. and Grossberg, S. (1988). *The ART of adaptive pattern recognition by a self-organizing neural network*. MIT Press.

Clarke, A., Grzeczkowski, L., Mast, F. W., Gauthier, I., and Herzog, M. H. (2014). Deleterious effects of roving on learned tasks. *Vision Research*, 99:88–92.

Clopath, C., Ziegler, L., Vasilaki, E., Büsing, L., and Gerstner, W. (2008). Tag-trigger-consolidation: A model of early and late long-term-potentiation and depression. *PLoS Computational Biology*, 4(12).

Douglas, R. J. and Martin, K. A. C. (1990). Control of Neuronal Output by Inhibition at the Axon Initial Segment. *Neural Computation*, 2:282–292.

Flesch, T., Balaguer, J., Dekker, R., Nili, H., and Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences*, 115(44):201800755.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4):128–135.

Frey, U. and Morris, R. G. M. (1997). Synaptic tagging and long-term potentiation. *Nature*, 385:533–536.

Frey, U. and Morris, R. G. M. (1998). Weak before strong: Dissociating synaptic tagging and plasticity-factor accounts of late-LTP. *Neuropharmacology*, 37(4-5):545–552.

Fusi, S., Drew, P. J., and Abbott, L. F. (2005). Cascade models of synaptically stored memories. *Neuron*, 45(4):599–611.

He, X. and Jaeger, H. (2018). Overcoming Catastrophic Interference using Conceptor-Aided Backpropagation. In *International Conference on Learning Representations*, pages 1–11.

Herzog, M. H., Aberg, K. C., Frémaux, N., Gerstner, W., and Sprekeler, H. (2012). Perceptual learning, roving and the unsupervised bias. *Vision Research*, 61:95–99.

Kaplanis, C., Shanahan, M., and Clopath, C. (2018). Continual Reinforcement Learning with Complex Synapses. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, pages 2497–2506. PMLR.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., and Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *PNAS*, 114(13):3521–3526.

Kumaran, D., Hassabis, D., and McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, 20(7):512–534.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (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. *Psychological review*, 102(3):419–457.

Mongillo, G. and Denève, S. (2008). Online Learning with Hidden Markov Models. *Neural Computation*.

Morris, R. G. M. (2006). Elements of a neurobiological theory of hippocampal function: The role of synaptic plasticity, synaptic tagging and schemas. *European Journal of Neuroscience*, 23:2829–2846.

Otto, T. U., Herzog, M. H., Fahle, M., and Zhaoping, L. (2006). Perceptual learning with spatial uncertainties. *Vision Research*, 46:3223–3233.

Parkosadze, K., Otto, T. U., Malania, M., Kezeli, A., and Herzog, M. H. (2008). Perceptual learning of bisection stimuli under roving : Slow and largely specific. *Journal of Vision*, 8(5):1–8.

Redondo, R. L. and Morris, R. G. M. (2011). Making memories last: the synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, 12:17–30.

Roxin, A. and Fusi, S. (2013). Efficient Partitioning of Memory Systems and Its Importance for Memory Consolidation. *PLoS Computational Biology*, 9(7).

Sajikumar, S. and Frey, J. U. (2003). Anisomycin inhibits the late maintenance of long-term depression in rat hippocampal slices in vitro. *Neuroscience Letters*, 338(2):147–150.

Sajikumar, S. and Frey, J. U. (2004). Late-associativity, synaptic tagging, and the role of dopamine during LTP and LTD. *Neurobiology of Learning and Memory*, 82:12–25.

Shires, K. L., Da Silva, B. M., Hawthorne, J. P., Morris, R. G. M., and Martin, S. J. (2012). Synaptic tagging and capture in the living rat. *Nature Communications*, 3(1246).

Somogyi, P., Nunzi, M. G., Gorio, A., and Smith, A. D. (1983). A new type of specific interneuron in the monkey hippocampus forming synapses exclusively with the axon initial segments of pyramidal cells. *Brain Research*, 259(1):137–142.

Spitzer, B., Waschke, L., and Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, 1(8):1–8.

Squire, L. R. and Alvarez, P. (1995). Retrograde amnesia and memory consolidation:a neurobiological perspective. *Current Opinion in Neurobiology*, 5:169–177.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437:1272–1278.

Tartaglia, E. M., Aberg, K. C., and Herzog, M. H. (2009a). Perceptual learning and roving: Stimulus types and overlapping neural populations. *Vision Research*, 49(11):1420–1427.

Tartaglia, E. M., Bamert, L., Mast, F. W., and Herzog, M. H. (2009b). Human Perceptual Learning by Mental Imagery. *Current Biology*, 19(24):2081–2085.

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., and Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences*, 113(11):3102–3107.

Zenke, F., Poole, B., and Ganguli, S. (2017). Continual Learning Through Synaptic Intelligence. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3987–3995. PMLR.

Ziegler, L., Zenke, F., Kastner, D. B., and Gerstner, W. (2015). Development/Plasticity/Repair Synaptic Consolidation: From Synapses to Behavioral Modeling. *The Journal of Neuroscience*, 35(3):1319–1334.

CHAPTER 4

Discussion

The larger the island of knowledge, the longer the shoreline of mystery.
— Unknown author

The main results of our research were discussed in our publication (Chapter 3). In this chapter, we want to briefly recap the key findings and provide further insights. We discuss how our model compares with other related ones and provide ideas on how to improve the learning rule in future work.

4.1 Key findings

Our novel computational model shows that describing synaptic plasticity with two weight components, a fast and a slow one, diminishes the stability-plasticity dilemma and solves the problem of catastrophic forgetting in continual learning tasks.

We postulate that the fast component is the dominant part during learning. Its main driving force is the goal to maximize reward. A fast component which is considered to be important is consolidated into the slow component and is stored there for a longer time, whereas unimportant fast components decay back to baseline. The learning rule is reward-based, is derived from optimal learning and does not contain any hidden variables.

The model describes fast transients during early LTP and most of the tagging & capture data that have been addressed by phenomenological models. As a price for fighting catastrophic forgetting the model shows slowdowns in learning during stimulus mixing (roving) experiments similar as observed with humans.

4.2 Further insights

4.2.1 Catastrophic forgetting

Our learning rule prevents catastrophic forgetting in continual learning since the fast component can be altered temporarily without influencing the slow component. Generally, after a context switch, neurons which were active in the previous context will still be active at first in the new context if input patterns in both contexts overlap. This causes many non-rewarded trials and thus synaptic suppression. The bigger the overlap of input patterns across contexts, the stronger this effect. With a classical 1-component rule, this results in losing memories of the previous context. With the 2-component rule, on the other hand, only the fast components are suppressed, keeping the memory of the previous context still present in the slow components. To avoid that the fast components overwrites the slow components, only fast components informative about the input pattern should be selected for consolidation. This can be achieved by imposing two consolidating criteria. The first one is a threshold on the absolute value of the fast component $\Theta(|w_{ij}^f| - \theta_w)$. It ensures that a weight update has to occur often enough in the same direction before consolidation is triggered. The second criterion is a threshold on the low-pass filtered postsynaptic voltage $\Theta(\bar{V}_i - \theta_V)$. It selects neurons for consolidation which are relevant for the current context. Using the low-pass filtered voltage $\bar{V}$ instead the voltage V itself improves this selection, because also irrelevant neurons sometimes reach a high voltage due to neuronal noise.

4.2.2 Small pattern distinctions

Within a context, input patterns can be very similar which causes an additional challenge besides the problem of catastrophic forgetting. While catastrophic forgetting can not be solved by extending the learning time, separation of very similar patterns can be gradually improved by increasing the number of trials.

For a fixed number of trials, separation of similar patterns is achieved by punishing strong voltage deflections. The term $(-\lambda_V \text{sign}(V_i) x_j)$ in the learning rule of the fast

component (Equation 3.3) removes common components of the input pattern, thus ensuring that deviations from the mean input pattern are learned instead of the input patterns themselves. If components are identical within and across contexts, the punishment term also helps to avoid catastrophic forgetting because context overlaps get reduced.

4.2.3 Fast learning

To model fast transient LTP/LTD as observed in experiments, weights need to change strongly in a short time. As a consequence, the standard approach of stochastic gradient descent learning fails because the gradient can not be sampled often enough in a short time to get a precise estimate. To enable fast learning anyway, we use two mechanisms which are helpful not only in continual learning tasks but also if tasks are learned in an isolated way.

First and foremost, the consolidation process ensures that the total weight changes are a good estimate of the gradient despite using a high learning rate for the fast component. The original changes of the fast component Δw_{ij}^f are a rough estimate of the gradient. However, since we consider online learning, i.e. the weights are updated after each pattern presentation, Δw_{ij}^f can strongly deviate from the gradient. A better estimate is the fast component itself, which is formed by summing up all past changes, $w_{ij}^f(t) = \sum_{t'<t} \Delta w_{ij}^f(t')$. Due to the high learning rate, this is still an imprecise estimate of the gradient. To improve the estimate, we consider the fast component at different time steps as samples of the gradient, and update the slow component not by Δw_{ij}^f but by the better gradient estimates w_{ij}^f. By doing so, the weight updates become a precise estimate of the gradient. This mechanism in similar to batch learning in machine learning. In batch learning, weight contribution of each pattern are stored in a hidden variable. Only after a set of patterns are presented, the value of the hidden variable is used to update the weight. The hidden variable corresponds to the fast component in our model, but, differently from batch learning, the fast component influences the total weight.

The second mechanism to enable fast learning is to make the synapses agnostic to competition. If the learning rule is derived by calculating the derivative of the utility function, learning happens based on the difference between the target activity φ_i^{WTA} imposed by the winner-take-all dynamics and the neuronal activity normalized across the whole output layer, $\varphi(V_i)/\sum_i \varphi(V_i)$. This leads to fast saturation which

is problematic in particular if input patterns are similar and thus target activity and normalized neuron activity will remain similar over a longer time. Hence, the term $(\varphi_i^{\mathrm{WTA}} - \varphi(V_i)/\sum_i \varphi(V_i))$ stays small and weight changes are extinguished even for wrong outputs. In contrast, if the normalization is ignored, each synapse adapts their weight independently of the activity of other neurons and thus saturation does not occur prematurely. While this is a small modification of the gradient approach, we found out that this strongly helps to speed up learning.

4.3 Biological relevance

The model provides insights to learning and memory in neural substrates and, despite its simple formulation, several aspects of the model can be related to biological processes.

4.3.1 Protein dependent and independent processes

Synaptic weight changes are influenced by many different processes evolving at different speeds (Frey et al., 2009). The underlying biochemical processes can be assigned to two classes corresponding to the two components in the model. The fast component can be associated with the phosphorylation of glutamate-receptor-gated ion channels while the slow component represents protein synthesis dependent processes.

Differentiation between protein synthesis independent and dependent processes is well known from the synaptic tagging & capture hypothesis (Frey and Morris, 1997; Redondo and Morris, 2011). The hypothesis claims that consolidation is only triggered if plasticity related proteins and a local synaptic tag are simultaneously present. Without triggering consolidation, early LTP decays back to baseline. Plasticity related proteins need to be synthesized which requires strong neural activity (Redondo and Morris, 2011), implied in our model by the threshold $\Theta(\bar{V}_i - \theta_V)$. The use of the low-pass filtered postsynaptic voltage relates to the observations that plasticity related proteins have a lifetime of a few ten minutes (Shires et al., 2012). The biological basis of tag-setting is unclear. We modeled it with the threshold $\Theta(|w_{ij}^f| - \theta_w)$ which is in accordance with all considered experiments.

4.3.2 Memory consolidation during breaks

Several studies show that sleep improves consolidation of newly acquired memories (Stickgold, 2005; Diekelmann and Born, 2010). In some experiments, test subjects performed a task better if they slept between learning and testing than staying awake (Fenn et al., 2003; Rasch et al., 2007). One hypothesis, among several ones, proposes sleep consolidation happens due to synaptic homeostasis. (Tononi and Cirelli, 2006; Daniel et al., 2011). In brief, net synaptic strength is increased during wakefulness and downscaled during sleep. Molecular and electrophysiological evidences for synaptic homeostasis were found in rodents (Vyazovskiy et al., 2008). The dynamics imposed by the 2-component rule match with this hypothesis. During learning, the fast weight components contribute to a net increase of the synaptic strengths while during a break, i.e. a phase of no input stimuli, the fast weight components decay and thus the synaptic strength is downscaled. Intriguingly, the downscaling in the network improves task performance because sample-specific noise in the fast components is removed.

4.3.3 Roving

Since the 2-component rule is optimized for learning when contexts change slowly, it causes subprime results when they change fast. This corresponds to observations made in psychophysical experiments with humans. In these experiments it was shown that learning to discriminate bisection stimuli is slowed down under roving conditions ('mixing') compared to block-wise learning (Otto et al., 2006; Parkosadze et al., 2008; Clarke et al., 2014). However, most learning rules, even the most simplest one, are able to learn the bisection task under roving conditions as the task is linearly separable. We found a possible explanation for the learning impairment which is related to the high learning rate for the fast component. While the high learning rate is beneficial when contexts are changed on the same time scale as early LTP decays, it hinders learning if contexts or tasks are switched more quickly. A confirmation of our theory with experimental data is difficult as there are no experiments that measured synaptic weights while learning under roving conditions.

4.4 Comparison with other models

In section 1.8, models dealing with the same questions as the 2-component rule were presented. Here, a selection of these models gets compared with the 2-component rule.

All discussed models which tackle the problem of catastrophic forgetting postulate internal 'hidden' variables, i.e. variables which do not contribute to the synaptic weight. In Fusi et al. (2005) and Benna and Fusi (2016) the hidden variables are cascades of states associated with each synapse. In Kirkpatrick et al. (2017), Zenke et al. (2017) and Aljundi et al. (2018) the importance weights associated with each synaptic weight are hidden variables. In contrast, all variables in the 2-component rule are 'visible' and can thus directly be observed.

Additionally, these models and the 2-component rule differ in the assumptions about the environment. The cascade models (Fusi et al., 2005; Benna and Fusi, 2016) are based on the hypothesis that forgetting can never be completely prevented but an optimal decay rate can be found. They consider uncorrelated inputs where forgetting happens due to capacity restrictions. While in our model capacity restrictions are not taken into account, we consider a task where catastrophic forgetting occurs due to the interference of similar inputs requiring the activation of different outputs. The goal of the 2-component rule is to have no forgetting at all for weights informative about the context and complete forgetting for non-informative weights, a distinction which can not be made if only uncorrelated inputs are considered.

Kirkpatrick et al. (2017), Zenke et al. (2017) and Aljundi et al. (2018) claim that an optimal weight configuration for a new task can always be found in the region of a previously established configuration. They try to avoid strong weight changes to not disturb the already established one. If many tasks are learned in sequence, plasticity is however more and more reduced. In our model strong weight changes are implemented on purpose to find a new weight configuration fast, though the strong changes are only transiently stored in the fast component.

The roving phenomenon was previously described by Herzog et al. (2012). They claim that humans cannot estimate the correct reward during learning of a bisection task with different difficulties. This results in a bias term which causes a synaptic drift and hinders learning. While a bias term also exists in our model, it can not explain the learning impairments for us. The 1-component version of our model has the same bias term as the 2-component rule, but learning is impaired only with the 2-component rule.

One observation made in roving experiments is neither explained by Herzog et al. (2012) nor by the 2-component rule: Clarke et al. (2014) could show that roving does

not only slow down learning but the learning performance even drops below a previously achieved level.

4.5 Model predictions

The model was designed to reach maximal performance in continual learning tasks. Using the optimized learning rule, several predictions about synaptic consolidation can be made.

1. *Protein synthesis depends on the post-synaptic somatic voltage.*
   Several studies confirmed that proteins are synthesized centrally in the soma (Frey and Morris, 1997). Other findings indicate that the synthesis may occur also in dendrites and that for the establishment of late LTP both dendritic and somatic protein synthesis is required (Casadio et al., 1999; Bradshaw et al., 2003) The proof that protein synthesis depends on the post-synaptic somatic voltage is however missing. An adaption of the model to dendritic voltage or activity is straightforward.

2. *A local synaptic tag is set if the synaptic weight is strong during the early phase of LTP.*
   Experiments verified that a minimal stimulus strength is necessary to set a tag and to prevent the decay of early LTP (Frey and Morris, 1997). Open, however, is the question whether a tag is set if a strong stimulus is applied but the synaptic weight stays small, e.g. due to a negative reward prediction error.

3. *Consolidation does not require dopamine. However, dopamine is necessary for protein synthesis independent plasticity.*
   In experiments it was observed that dopamine enhances memory consolidation (Bernabeu et al., 1997; Schott et al., 2005). In our model, dopamine as a reward signal is only implicitly necessary for consolidation; only due to reward the fast component can cross the threshold $\Theta(|w_{ij}^f| - \theta_w)$. The subsequent processes in the model do not rely on a reward signal. This is in contradiction to the study of Huang and Kandel (1995) where strong weights are artificially induced and dopamine receptor agonists are required to observe long-lasting changes. In newer studies done *in vivo*, however, it was shown that despite dopamine inhibition consolidation can occur (Shires et al., 2012).

Concerning learning in general, the 2-component rule predicts that a strong stimulus is necessary to enable robust learning. This stimulus can either be the task stimulus or a secondary stimulus applied to the same neuron during the same time window. Important, it is however, that one of the stimuli activates protein synthesis. Memories induced by weak stimuli can be learned but they decay on the timescale of a few hours. As pointed out by Redondo and Morris (2011), this might be an explanation for 'flashbulb memories'. While in general we forget unimportant events, we can remember them if they happen around the time of an significant life event.

Moreover, the model predicts recall improvements during breaks. In the absence of external stimuli, the fast component decays back to baseline. Since the slow component has a small update rate and is not explicitly driven by external stimuli, it provides a more accurate knowledge about the task after learning than the sum of both components.

4.6 Future work & model improvements

The 2-component learning rule is derived from higher order principles with little assumptions about the underlying biochemical processes and thus is easily expandable and implementable in other frameworks.

4.6.1 Towards more biological plausibility

The model contains two simplifications which are biologically implausible:

- synaptic weight changes are based on firing rates,
- synaptic weights can switch sign, thus violating Dale's principle according to which neurons are either excitatory or inhibitory.

These two simplifications strongly keep the model easy and understandable. However, if the goal is to create a unified theory of the nervous system, these two points need to be addressed by formulating the learning rule in terms of spikes instead of rates and by restricting weights to one sign. A spiking version of the model would additionally allow to run faster and more energy-efficient simulations on dedicated hardware.

More biological plausibility could also be achieved by implementing more than two weight components in the model. In the limit, one could even have for each processes underlying synaptic plasticity an own weight component. Such a model could reveal

new computational advantages but it would go along with a loss in generality and simplicity.

4.6.2 New directions

Reinforcement learning is a popular framework due to its high biological plausibility. However, especially in machine learning other frameworks are often considered. In future work, the consequences of the 2-component rule in supervised and unsupervised learning could be investigated. Like reinforcement learning, these frameworks suffer from catastrophic forgetting and we expect, with small modifications on the fast component, that the 2-component rule can also be applied to these frameworks. The slow component was not conceived for reinforcement learning in particular and can thus directly be applied to other frameworks.

For applications in machine learning, continual learning becomes more and more important. The steady growth of data makes it necessary that tasks are split up into smaller parts and are learned each after the other. Hence, a solution for catastrophic forgetting must be found. With this perspective, having an implementation of our 2-component rule in deep neural networks is desirable. Depending on the task, different consolidation threshold might be necessary. An extension of our rule could implement adaptive thresholds which react to the mean input strength such that the consolidation rate stays constant.

A dependency of the thresholds on the mean reward could avoid unnecessary consolidation events. This is a relevant aspect if energy costs in biology are taken into account. Not only do synaptic changes consume energy but also maintaining the strength of a synapse costs energy Harris et al. (2012).

Future work should investigate catastrophic forgetting resulting from capacity limitations. So far we focused on catastrophic forgetting caused by highly correlated tasks learned in sequence. If tasks are uncorrelated but the number of synapses is limited, it is necessary to find an efficient encoding of information. The selective consolidation proposed by our model could be used as a possible starting point.

Bibliography

Abbott LF. Synaptic Depression and Cortical Gain Control. Science. 1997; 275(5297):221–224. doi: 10.1126/science.275.5297.221.

Abraham WC, Robins A. Memory retention - The synaptic stability versus plasticity dilemma. Trends in Neurosciences. 2005; 28(2):73–78. doi: 10.1016/j.tins.2004.12.003.

Aljundi R, Babiloni F, Elhoseiny M, Rohrbach M, Tuytelaars T. Memory Aware Synapses: Learning what (not) to forget. In: Ferrari V, Hebert M, Sminchisescu C, Weiss W, editors. *The European Conference on Computer Vision (ECCV)* Springer; 2018. p. 144–161.

Amarasingham A. Spike Count Reliability and the Poisson Hypothesis. Journal of Neuroscience. 2006; 26(3):801–809. doi: 10.1523/JNEUROSCI.2948-05.2006.

Amari SI. Natural Gradient Works Efficiently in Learning. Neural Computationomputation. 1998; 10(2):251–276.

Anderson RB. The power law as an emergent property. Memory and Cognition. 2001; 29(7):1061–1068. doi: 10.3758/BF03195767.

Baddeley A. Working memory: Looking back and looking forward. Nature Reviews Neuroscience. 2003; 4(10):829–839. doi: 10.1038/nrn1201.

Barrett AB, Billings GO, Morris RGM, Van Rossum MCW. State based model of long-term potentiation and synaptic tagging and capture. PLoS Computational Biology. 2009; 5(1). doi: 10.1371/journal.pcbi.1000259.

Barto AG, Sutton RS. Reinforcement learning: An introduction; 2018. doi: 10.1109/TNN.1998.712192.

Bell CC, Han VZ, Sugawara Y, Grant K. Synaptic plasticity in a cerebellum-like structure depends on temportal order. Nature. 1997; 387:278–281. doi: 10.1038/nature02013.

Benna MK, Fusi S. Computational principles of biological memory. Nature Neuroscience. 2016; 19(12):1697–1706. doi: 10.1038/nn.4401.

Bernabeu R, Bevilaqua LRM, Ardenghi P, Bromberg E, Schmitz P, Bianchin M, Izquierdo I, Medina JH. Involvement of hippocampal cAMP/cAMP-dependent protein kinase signaling pathways in a late memory consolidation phase of aversively motivated learning in rats. Proceedings of the National Academy of Sciences of the United States of America. 1997; 94:7041–7046.

Bi Gq, Poo Mm. Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. The Jourcnal of Neuroscience. 1998; 18(24):10464–10472. doi: 10.1038/25665.

Bliss TVP, Collingridge GL. A synaptic model of memory: long-term potentiation in the hippocampus. Nature. 1993; 361(6407):31–39. doi: 10.1038/361031a0.

Bliss TVP, Lømo T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. The Journal of Physiology. 1973; 232(2):331–356. doi: 10.1113/jphysiol.1973.sp010273.

Bosch M, Castro J, Saneyoshi T, Matsuno H, Sur M, Hayashi Y. Structural and molecular remodeling of dendritic spine substructures during long-term potentiation. Neuron. 2014; 82(2):444–459. doi: 10.1016/j.neuron.2014.03.021.

Bradshaw KD, Emptage NJ, Bliss TVP. A role for dendritic protein synthesis in hippocampal late LTP. European Journal of NeuroscienceNeuroscience. 2003; 18:3150–3152. doi: 10.1046/j.1460-9568.2003.03054.x.

Brashers-Krug T, Shadmehr R, Bizzi E. Consolidation in human motro memory. Nature. 1996; 383:252–255. doi: 10.1038/nature03031.1.

Brea J, Urbanczik R, Senn W. A Normative Theory of Forgetting: Lessons from the Fruit Fly. PLoS Computational Biology. 2014; 10(6):1–9. doi: 10.1371/journal.pcbi.1003640.

Carpenter GA, Grossberg S. The ART of adaptive pattern recognition by a self-organizing neural network. MIT Press; 1988. doi: 10.1109/2.33.

Casadio A, Martin KC, Giustetto M, Zhu H, Chen M, Bartsch D, Bailey CH, Kandel ER. A Transient, Neuron-Wide Form of CREB-Mediated Long-Term Facilitation Can Be Stabilized at Specific Synapses by Local Protein Synthesis. Cell. 1999; 99:221–237.

Castellucci V, Pinsker H, Kupfermann I, Kandel ER. Neuronal mechanisms of habituation and dishabituation of the gill withdrawal reflex in Aplysia. Science. 1970; 167(3926):1745–1748. doi: 10.1126/science.167.3926.1745.

Clarke A, Grzeczkowski L, Mast FW, Gauthier I, Herzog MH. Deleterious effects of roving on learned tasks. Vision Research. 2014; 99:88–92. doi: 10.1016/j.visres.2013.12.010.

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

Colombo M. Deep and beautiful. The reward prediction error hypothesis of dopamine. Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences. 2014; 45(1):57–67. doi: 10.1016/j.shpsc.2013.10.006.

Corkin S. What's new with the amnesic patient H. M.? Nature Rev Neurosci. 2002; 3(February):153–160. doi: 10.1038/nrn726.

Crick F. The recent excitement about neural networks. Nature. 1989; 337:129–132.

Curtis DR, Eccles JC. Synaptic Action During and After Repetitive. J Physiol. 1960; 150:374–398.

Daniel B, Giulio T, Chiara C. Sleep and Synaptic Homeostasis : Structural Evidence in Drosophila. Science. 2011; 332(June):1576–1581. doi: 10.1126/science.1202839.

Dayan P, Abbott LF. Theoretical neuroscience: computational and mathematical modeling of neural systems. MIT Press; 2001.

De Carlos JA, Borrell J. A historical reflection of the contributions of Cajal and Golgi to the foundations of neuroscience. Brain Research Reviews. 2007; 55(1):8–16. doi: 10.1016/j.brainresrev.2007.03.010.

De Robertis E, Bennett HS. Some features of the submicroscopic morphology of synapses in frog and earthworm. The Journal of biophysical and biochemical cytology. 1955; I(1):47 – 56.

Diekelmann S, Born J. The memory function of sleep. Nature Reviews Neuroscience. 2010; 11(2):114–126. doi: 10.1038/nrn2762.

Douglas RJ, Martin KAC. Control of Neuronal Output by Inhibition at the Axon Initial Segment. Neural Computation. 1990; 2:282–292. doi: 10.1162/neco.1990.2.3.283.

Dudai Y. The Neurobiology of Consolidations, Or, How Stable is the Engram? Annual Review of Psychology. 2004; 55(1):51–86. doi: 10.1146/annurev.psych.55.090902.142050.

Dudek SM, Bear MF. CAl N-methyl-D-aspartate. Neurobiology. 1992; 89(May):4363–4367.

Dunwiddie T, Lynch G. Long-term potentiation and depression of synaptic responses in the rat hippocampus: Localization and frequency dependency. J Physiol. 1978; (276):353–367.

Eichenbaum H. The hippocampus and declarative memory: cognitive mechanisms and neural codes. Behav Brain Res. 2001; 127(1-2):199–207.

Fenn KM, Nusbaum HC, Margoliash D. Consolidation during sleep of perceptual learning of spoken language. Nature. 2003; 425(6958):614–616. doi: 10.1038/nature01951.

Flesch T, Balaguer J, Dekker R, Nili H, Summerfield C. Comparing continual task learning in minds and machines. Proceedings of the National Academy of Sciences. 2018; 115(44):201800755. doi: 10.1073/pnas.1800755115.

Frankland PW, Bontempi B. The organization of recent and remote memories. Nature Reviews Neuroscience. 2005; 6(2):119–130. doi: 10.1038/nrn1607.

Frémaux N, Sprekeler H, Gerstner W. Development/Plasticity/Repair Functional Requirements for Reward-Modulated Spike-Timing-Dependent Plasticity. The Journal of Neuroscience. 2010; 30(40):13326–13337. doi: 10.1523/JNEUROSCI.6249-09.2010.

French RM. Catastrophic forgetting in connectionist networks. Trends in Cognitive Sciences. 1999; 3(4):128–135. doi: 10.1016/S1364-6613(99)01294-2.

Frey MC, Sprengel R, Nevian T. Cellular/Molecular Activity Pattern-Dependent Long-Term Potentiation in Neocortex and Hippocampus of GluA1 (GluR-A) Subunit- Deficient Mice. The Journal of Neuroscience. 2009; 29(17):5587–5596. doi: 10.1523/JNEUROSCI.5314-08.2009.

Frey U, Huan YY, Kandel ER. Effects of cAMP simulate a late stage of LTP in hippocampal CA1 neurons. Science. 1993; 260:1661–1664.

Frey U, Krug M, Reymann KG, Matthies H. Anisomycin, an inhibitor of protein synthesis, blocks late phases of LTP phenomena in the hippocampal CA1 region in vitro. Brain Research. 1988; 452(1-2):57–65. doi: 10.1016/0006-8993(88)90008-X.

Frey U, Morris RGM. Synaptic tagging and long-term potentiation. Nature. 1997; 385:533–536. doi: 10.1038/385533a0.

Frey U, Morris RGM. Synaptic tagging: Implications for late maintenance of hippocampal long- term potentiation. Trends in Neurosciences. 1998; 21(5):181–188. doi: 10.1016/S0166-2236(97)01189-2.

Frey U, Morris RGM. Weak before strong: Dissociating synaptic tagging and plasticity-factor accounts of late-LTP. Neuropharmacology. 1998; 37(4-5):545–552. doi: 10.1016/S0028-3908(98)00040-9.

Fusi S, Drew PJ, Abbott LF. Cascade models of synaptically stored memories. Neuron. 2005; 45(4):599–611. doi: 10.1016/j.neuron.2005.02.001.

von Gerlach J. Von dem Rückmark. In: *Handbuch der Lehre von den Geweben des Menschen und der Thiere* Engelmann; 1871.p. 665–693.

Gerstner W, Richard K, Leo van Hemmen J, Wagner H. A neuronal learning rule for sub-millisecond temportal coding. Nature. 1996; 383:76–78.

Gerstner W, Sprekeler H, Deco G. Theory and Simulation in Neuroscience. Science. 2012; 338:60–65. doi: 10.1126/science.1227356.

Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.

Grossberg S. Competitive learning: From interactive activation to adaptive resonance. Cognitive Science. 1987; 11:23–63. http://www.sciencedirect.com/science/article/pii/S0364021387800253, doi: 10.1016/s0364-0213(87)80025-3.

Harris JJ, Jolivet R, Attwell D. Synaptic Energy Use and Supply. Neuron. 2012; 75:762–777. doi: 10.1016/j.neuron.2012.08.019.

Hayashi Y, Shi SH, Esteban J, Piccini A, Poncer JC, Malinow R. Driving AMPA Receptors into Synapses by LTP and CaMKII: Requirement for GluR1 and PDZ Domain Interaction. Science. 2000; 287(5461):2262–2267. doi: 10.1126/science.287.5461.2262.

He X, Jaeger H. Overcoming Catastrophic Interference using Conceptor-Aided Backpropagation. In: *International Conference on Learning Representations*; 2018. p. 1–11. https://openreview.net/forum?id=B1al7jg0b.

Hebb DO. The Organization of Behavior. John Wiley and Sons; 1949.

Herzog MH, Aberg KC, Frémaux N, Gerstner W, Sprekeler H. Perceptual learning, roving and the unsupervised bias. Vision Research. 2012; 61:95–99. doi: 10.1016/j.visres.2011.11.001.

Hinton G, Plaut DC. Using Fast Weights to Deblur Old Memories. In: *The Ninth Annual Conference of the Cognitive Science Society*; 1987. p. 177–186.

Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve, vol. 52; 1952. doi: 10.1007/BF02459568.

Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience. 1998; 1(4):304–309. doi: 10.1038/1124.

Huang Yy, Kandel ER. D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the CAl region of the hippocampus. PNAS. 1995; 92:2446–2450. doi: 10.1073/pnas.92.7.2446.

Isele D, Cosgun A. Selective Experience Replay for Lifelong Learning. In: *The Thirty-Second AAAI Conference on Artificial Intelligence*; 2018. p. 3302–3309. http://arxiv.org/abs/1802.10269.

Jones EG. Santiago Ramon y Cajal and the croonian lecture, March 1894. Trends in Neurosciences. 1994; 17(5):192–193. doi: 10.1016/0166-2236(94)90101-5.

Kaplanis C, Shanahan M, Clopath C. Continual Reinforcement Learning with Complex Synapses. In: Dy J, Krause A, editors. *Proceedings of the 35th International Conference on Machine Learning* PMLR; 2018. p. 2497–2506.

Karmarkar UR, Buonomano DV. Temporal Specificity of Perceptual Learning in an Auditory Discrimination Task. Learning & Memory. 2003; 10:141–147. doi: 10.1101/lm.55503.

Karni A, Sagi D. Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. PNAS. 1991; 88:4966–4970.

Kazdin AE. Encyclopedia of Psychology. Oxford University Press; 2000.

Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D, Hadsell R. Overcoming catastrophic forgetting in neural networks. PNAS. 2017; 114(13):3521–3526. doi: 10.1073/pnas.1611835114.

Kossel A, Bonhoeffer T, Bolz J. Non-Hebbian synapses in rat visual cortex. Developmental Neuroscience. 1990; 1(2):115–118.

Krug M, Lössner B, Ott T. Anisomycin blocks the late phase of long-term potentiation in the dentate gyrus of freely moving rats. Brain Research Bulletin. 1984; 13(1):39–42. doi: 10.1016/0361-9230(84)90005-4.

Kullmann DM, Nicoll RA. Long-term potentiation is associated with increases in quantal content and quantal amplitude. Nature. 1992; 357(6375):240–244. doi: 10.1038/357240a0.

Kumaran D, Hassabis D, McClelland JL. What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. Trends in Cognitive Sciences. 2016; 20(7):512–534. doi: 10.1016/j.tics.2016.05.004.

Li YX, Rinzel J, Equations for InsP3 receptor-mediated [Ca2+]i oscillations derived from a detailed kinetic model: a Hodgkin-Huxley like formalism.; 1994. doi: 10.1006/jtbi.1994.1041.

Lynch MA. Long-term potentiation and memory. Physiological reviews. 2004; 84(1):87–136. doi: 10.1152/physrev.00014.2003.

Markram H, Lubke J, Frotscher M, Sakmann B. Regulation of Synaptic Efficacy by Coindence of Postsynaptic APs and EPSPs. Science. 1997; 275(January):213–215.

Markram H, Tsodyks M, Redistribution of synaptic efficacy between neocortical pyramidal neurons; 1996. doi: 10.1038/382807a0.

Masanori M, Honkura N, Ellis-Davies GCR, Kasai H. Structural basis of long-term potentiation in single dendritic spines. Nature. 2004; 429.

McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review. 1995; 102(3):419–457. doi: 10.1037/0033-295X.102.3.419.

McCloskey M, Cohen NJ. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. Psychology of Learning and Motivation - Advances in Research and Theory. 1989; 24(C):109–165. doi: 10.1016/S0079-7421(08)60536-8.

McGaugh JL. Memory - A century of consolidation. Science. 2000; 287(5451):248–251. doi: 10.1126/science.287.5451.248.

Meunier C, Segev I. Playing the Devil's advocate: is the Hodgkin–Huxley model useful? Trends in Cognitive Sciences. 2002; 25(11):558–563.

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D. Human-level control through deep reinforcement learning. Nature. 2015; 518:529–533. doi: 10.1038/nature14236.

Mongillo G, Denève S. Online Learning with Hidden Markov Models. Neural Computation. 2008; doi: 10.1162/neco.2008.10-06-351.

Montague PR, Dayan P, Sejnowski T. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. The Journal of Neuroscience. 1996; 16(5):1936–1947. doi: 10.1523/JNEUROSCI.16-05-01936.1996.

Morris RGM. Elements of a neurobiological theory of hippocampal function: The role of synaptic plasticity, synaptic tagging and schemas. European Journal of Neuroscience. 2006; 23:2829–2846. doi: 10.1111/j.1460-9568.2006.04888.x.

Murakoshi H, Yasuda R. Postsynaptic signaling during plasticity of dendritic spines. Trends in Neurosciences. 2012; 35(2):135–143. doi: 10.1016/j.tins.2011.12.002.

Oja E. Simplified neuron model as a principal component analyzer. Journal of Mathematical Biology. 1982; 15:267–273. doi: 10.1007/BF00275687.

Okada Ki, Toyama K, Inoue Y, Isa T, Kobayashi Y. Different Pedunculopontine Tegmental Neurons Signal Predicted and Actual Task Rewards. Journal of Neuroscience. 2009; 29(15):4858–4870. doi: 10.1523/JNEUROSCI.4415-08.2009.

O'Keefe J, Nadel L. The Hippocampus as a Cognitive Map. Oxford University Press; 1978. doi: 10.1097/00005053-198003000-00018.

O'Neil EB, Newsome RN, Li IHN, Thavabalasingam S, Ito R, Lee ACH. Examining the Role of the Human Hippocampus in Approach-Avoidance Decision Making Using a Novel Conflict Paradigm and Multivariate Functional Magnetic Resonance Imaging. Journal of Neuroscience. 2015; 35(45):15039–15049. doi: 10.1523/JNEUROSCI.1915-15.2015.

Otto TU, Herzog MH, Fahle M, Zhaoping L. Perceptual learning with spatial uncertainties. Vision Research. 2006; 46:3223–3233. doi: 10.1016/j.visres.2006.03.021.

Oyama K, Tateyama Y, Hernádi I, Tobler PN, Iijima T, Tsutsui KI. Discrete coding of stimulus value, reward expectation, and reward prediction error in the dorsal striatum. Journal of Neurophysiology. 2015; 114(5):2600–2615. doi: 10.1152/jn.00097.2015.

Palay SL. Synapses in the central nervous system. The Journal of biophysical and biochemical cytology. 1956; 2(4):193–202.

Parkosadze K, Otto TU, Malania M, Kezeli A, Herzog MH. Perceptual learning of bisection stimuli under roving : Slow and largely specific. Journal of Vision. 2008; 8(5):1–8. doi: 10.1167/8.1.5.

Pereda AE. Electrical synapses and their functional interactions with chemical synapses. Nature Reviews Neuroscience. 2014; 15(4):250–263. http://dx.doi.org/10.1038/nrn3708, doi: 10.1038/nrn3708.

Pitts W, McCulloch WS. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics. 1943; 5:115–133. doi: 10.1007/BF02478259.

Pospischil M, Toledo-Rodriguez M, Monier C, Piwkowska Z, Bal T, Frégnac Y, Markram H, Destexhe A. Minimal Hodgkin-Huxley type models for different classes of cortical and thalamic neurons. Biological Cybernetics. 2008; 99(4-5):427–441. doi: 10.1007/s00422-008-0263-8.

Rasch B, Buchel C, Gais S, Born J. Odor Cues During Slow-Wave Sleep Propt Declarative Memory Consolidation. Science. 2007; 315(March):1426–1429. doi: 10.1126/science.1138581.

Redondo RL, Morris RGM. Making memories last: the synaptic tagging and capture hypothesis. Nature Reviews Neuroscience. 2011; 12:17–30. doi: 10.1038/nrn2963.

Reymann KG, Frey JU. The late maintenance of hippocampal LTP: Requirements, phases, 'synaptic tagging', 'late-associativity' and implications. Neuropharmacology. 2007; doi: 10.1016/j.neuropharm.2006.07.026.

Robins A, McCallum S. Catastrophic forgetting and the pseudorehearsal solution in Hopfield-type networks. Connection Science. 1998; 10(2):121–135. doi: 10.1080/095400998116530.

Rolnick D, Ahuja A, Schwarz J, Lillicrap TP, Wayne G. Experience Replay for Continual Learning; 2018, http://arxiv.org/abs/1811.11682.

Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. . 1958; 65(6):386–408. doi: 10.1037/h0042519.

Roxin A, Fusi S. Efficient Partitioning of Memory Systems and Its Importance for Memory Consolidation. PLoS Computational Biology. 2013; 9(7). doi: 10.1371/journal.pcbi.1003146.

Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986; 323:533–536.

Sajikumar S, Frey JU. Anisomycin inhibits the late maintenance of long-term depression in rat hippocampal slices in vitro. Neuroscience Letters. 2003; 338(2):147–150. doi: 10.1016/S0304-3940(02)01400-3.

Sajikumar S, Frey JU. Late-associativity, synaptic tagging, and the role of dopamine during LTP and LTD. Neurobiology of Learning and Memory. 2004; 82:12–25. doi: 10.1016/j.nlm.2004.03.003.

Sajikumar S, Navakkode S, Sacktor TC, Frey JU. Synaptic Tagging and Cross-Tagging: The Role of Protein Kinase M in Maintaining Long-Term Potentiation But Not Long-Term Depression. Journal of Neuroscience. 2005; 25(24):5750–5756. doi: 10.1523/JNEUROSCI.1104-05.2005.

Schott BH, Düzel E, Heinze HJ, Wittmann BC, Guderian S, Frey JU. Reward-Related fMRI Activation of Dopaminergic Midbrain Is Associated with Enhanced Hippocampus- Dependent Long-Term Memory Formation. Neuron. 2005; 45:459–467. doi: 10.1016/j.neuron.2005.01.010.

Schultz W. Predictive reward signal of dopamine neurons. Journal of neurophysiology. 1998; 80:1–27. doi: 10.1152/jn.1998.80.1.1.

Schultz W. Neuronal Reward and Decision Signals: From Theories to Data. Physiological Reviews. 2015; 95(3):853–951. doi: 10.1152/physrev.00023.2014.

Schultz W, Dayan P, Montague PR. A Neural Substrate of Prediction and Reward. Science. 1997; 275(5306):1593–1599. doi: 10.1126/science.275.5306.1593.

Shires KL, Da Silva BM, Hawthorne JP, Morris RGM, Martin SJ. Synaptic tagging and capture in the living rat. Nature Communications. 2012; 3(1246). doi: 10.1038/ncomms2250.

Somogyi P, Nunzi MG, Gorio A, Smith AD. A new type of specific interneuron in the monkey hippocampus forming synapses exclusively with the axon initial segments of pyramidal cells. Brain Research. 1983; 259(1):137–142. doi: 10.1016/0006-8993(83)91076-4.

Spitzer B, Waschke L, Summerfield C. Selective overweighting of larger magnitudes during noisy numerical comparison. Nature Human Behaviour. 2017; 1(8):1–8. doi: 10.1038/s41562-017-0145.

Squire LR, Alvarez P. Retrograde amnesia and memory consolidation:a neurobiological perspective. Current Opinion in Neurobiology. 1995; 5:169–177. doi: 10.1016/0959-4388(95)80023-9.

Squire LR, Kandel ER. Memory: From Mind to Molecules. Robers and Company; 2009.

Squire LR, Kosslyn S, Zola-Morgan S, Haist F, Musen G. Memory and the Hippocampus: A Synthesis From Findings With Rats, Monkeys, and Humans. Psychological Review. 1992; 99(2):195–231. doi: 10.1037/0033-295X.99.3.582.

Stein W, De Hoz L, Gerstner W, Frémaux N. Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules. . 2016; 9. doi: 10.3389/fncir.2015.00085.

Stevens CF, Wang Y. Facilitation and depression at single central synapses. Neuron. 1995; 14(4):795–802. doi: 10.1016/0896-6273(95)90223-6.

Stickgold R. Sleep-dependent memory consolidation. Nature. 2005; 437:1272–1278. doi: 10.1038/nature04286.

Sutton RS, Barto AG. Toward a modern theory of adaptive networks: Expectation and prediction. Psychological Review. 1981; 88(2):135–170. `{%}3CGo{%}0Ato`, doi: 10.1016/j.envpol.2004.09.009.

Szepesvari C. Algorithms for Reinforcement Learning; 2010. doi: 10.2200/S00268ED1V01Y201005AIM009.

Takeuchi T, Duszkiewicz AJ, Morris RGM. The synaptic plasticity and memory hypothesis : Encoding , storage and persistence. Phil Trans R Soc B. 2014; 369. doi: 10.1098/rstb.2013.0288.

Tartaglia EM, Aberg KC, Herzog MH. Perceptual learning and roving: Stimulus types and overlapping neural populations. Vision Research. 2009; 49(11):1420–1427. doi: 10.1016/j.visres.2009.02.013.

Tartaglia EM, Bamert L, Mast FW, Herzog MH. Human Perceptual Learning by Mental Imagery. Current Biology. 2009; 19(24):2081–2085. doi: 10.1016/j.cub.2009.10.060.

Tononi G, Cirelli C. Sleep function and synaptic homeostasis. Sleep medicine reviews. 2006; 10(1):49–62. doi: 10.1016/j.smrv.2005.05.002.

Tsetsos K, Moran R, Moreland J, Chater N, Usher M, Summerfield C. Economic irrationality is optimal during noisy decision making. Proceedings of the National Academy of Sciences. 2016; 113(11):3102–3107. doi: 10.1073/pnas.1519157113.

Tzounopoulos T, Kim Y, Oertel D, Trussell LO. Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. Nature Neuroscience. 2004; 7(7):719–725. doi: 10.1038/nn1272.

Viola H, Moncada D. The Tagging And Capture Hypothesis From Synapse to Memory. In: *Molecular Basis of Memory* Academic Press; 2014.p. 391–423.

Vyazovskiy VV, Cirelli C, Pfister-Genskow M, Faraguna U, Tononi G. Molecular and electrophysiological evidence for net synaptic potentiation in wake and depression in sleep. Nature Neuroscience. 2008; 11(2):200–208. doi: 10.1038/nn2035.

Wixted JT. The Psychology and Neuroscience of Forgetting. Annual Review of Psychology. 2004; 55(1):235–269. doi: 10.1146/annurev.psych.55.090902.141555.

Yu C, Klein SA, Levi DM. Perceptual learning in contrast discrimination and the (minimal) role of context. Journal of Vision. 2004; 4(3):4. doi: 10.1167/4.3.4.

Ze H, Senior A, Schuster M. Statistical parametric speech synthesis using deep neural networks. In: *IEEE International Conference on Acoustics, Speech and Signal Processing* IEEE; 2013. p. 7962–7966. doi: 10.1109/ICASSP.2013.6639215.

Zenke F, Poole B, Ganguli S. Continual Learning Through Synaptic Intelligence. In: *Proceedings of the 34th International Conference on Machine Learning* PMLR; 2017. p. 3987–3995.

Zhang JY, Kuai SG, Xiao LQ, Klein SA, Levi DM, Yu C. Stimulus coding rules for perceptual learning. PLoS Biology. 2008; doi: 10.1371/journal.pbio.0060197.

Zhou Q, Homma KJ, Poo Mm, Berkeley B. Shrinkage of Dendritic Spines Associated with Long-Term Depression of Hippocampal Synapses. Neuron. 2004; 44:749–757.

Ziegler L, Zenke F, Kastner DB, Gerstner W. Development/Plasticity/Repair Synaptic Consolidation: From Synapses to Behavioral Modeling. The Journal of Neuroscience. 2015; 35(3):1319–1334. doi: 10.1523/JNEUROSCI.3989-14.2015.

List of publications

Peer-reviewed journal papers (unpublished)

- P. Leimer, M. Herzog, W. Senn *Synaptic weight decay with selective consolidation enables fast learning without catastrophic forgetting*, to be announced.

Pre-print submissions

- P. Leimer, M. Herzog, W. Senn *Synaptic weight decay with selective consolidation enables fast learning without catastrophic forgetting*, bioRxiv, 2019.

Conference submissions

- P. Leimer, W. Senn *Fast learning without forgetting by synaptic consolidation*, Cosyne 2018, Denver, USA.

Declaration of originality

Last name, first name	Leimer, Pascal
Matriculation number	09-110-578

I hereby declare that this thesis represents my original work and that I have used no other sources except as noted by citations. All data, tables, figures and text citations which have been reproduced from any other source, including the internet, have been explicitly acknowledged as such. I am aware that in case of non-compliance, the Senate is entitled to withdraw the doctorate degree awarded to me on the basis of the present thesis, in accordance with the "Statut der Universität Bern (Universitätsstatut; UniSt)", Art. 69, of 7 June 2011.

Bern, May 21, 2019