Cognitive Sciences EPrint Archive (CogPrints) 4012, (2004)
http://cogprints.org/4012/

# What is the functional role of adult neurogenesis in the hippocampus?

Laurenz Wiskott[1], Malte J. Rasch[1,*], and Gerd Kempermann[2,3]

[1]Institute for Theoretical Biology
Humboldt-University Berlin, Germany
http://itb.biologie.hu-berlin.de/

[2]Max Delbrück Center for Molecular Medicine
Berlin-Buch, Germany
[3]Dept. of Experimental Neurology, Charité
University Medicine Berlin, Germany

**Abstract**

The dentate gyrus is part of the hippocampal memory system and special in that it generates new neurons throughout life. Here we discuss the question of what the functional role of these new neurons might be. Our hypothesis is that they help the dentate gyrus to avoid the problem of catastrophic interference when adapting to new environments. We assume that old neurons are rather stable and preserve an optimal encoding learned for known environments while new neurons are plastic to adapt to those features that are qualitatively new in a new environment. A simple network simulation demonstrates that adding new plastic neurons is indeed a successful strategy for adaptation without catastrophic interference.

Keywords: hippocampus, dentate gyrus, adult neurogenesis, network model, catastrophic interference.



*Now at the Institute for Theoretical Computer Science, Technical University of Graz, Austria.

# Contents

# 1 Introduction

The hippocampus is a structure in the mammalian brain that is instrumental for acquiring episodic and declarative memories (see Sec. 2). If the hippocampus is lesioned, no new episodes or facts can be stored in long term memory. Skills, however, can still be learned (procedural memory). Because of its importance for learning and memory the hippocampus has been being extensively studied anatomically, physiologically, psychophysically, and by means of computational models.

One curious fact about the hippocampus is that one of its substructures, the dentate gyrus, generates new neurons even in the adult brain (see Sec. 2.4). This is very unusual and has been found to this extent only in one other brain structure, namely the olfactory bulb or rather the subventricular zone (Carleton et al., 2002) (there is some evidence of weak neurogenesis also in cortical areas (Gould & Gross, 2002)). In the olfactory bulb there seems to be a turnover of neurons, so that the new neurons simply replace old ones. In the dentate gyrus, on the other hand, the new neurons seem to be added to the existing network of neurons and somehow enhance its performance. What exactly is the functional role of these new neurons that are generated in the dentate gyrus throughout life? What makes the dentate gyrus so special? This question is not only interesting in itself but also offers a new approach for investigating the hippocampus as a whole.

We follow the view that the dentate gyrus performs an encoding operation and assume that this encoding has to adapt to a changing environment the animal lives in. Our hypothesis is that the new neurons are necessary to avoid the negative side effects of this adaptation, referred to as catastrophic interference. In this paper we lay out our arguments in detail and present a simple network model that illustrates our hypothesis and shows that adding new neurons is indeed a reasonable strategy for avoiding catastrophic interference.

In the following section the neurobiology of the hippocampus is briefly summarized. Then we will introduce relevant concepts from the theory of neural networks. The basic model of hippocampal function we have adopted is described in Section 4 and motivated by the facts and concepts of the preceding two section. Within this basic model we present our hypothesis in Section 5. The computational model illustrating our hypothesis is defined in Section 6, which also includes the simulation results. We conclude with a discussion. The basic ideas presented here have earlier been sketched in (Kempermann & Wiskott, 2004). The model simulations have been described in detail in (Rasch, 2003).

# 2 The hippocampal formation

## 2.1 Neuroanatomy

The hippocampal formation is part of the medial temporal lobe and consists of the entorhinal cortex (EC), the hippocampus, and the subicular complex (Amaral & Witter, 1989; Hammond, 2001). The hippocampus is composed of the dentate gyrus (DG) and the Ammon's horn (*cornu ammonis*), which is sometimes referred to as hippocampus proper. The Ammon's horn can be further subdivided into four subregions, CA1–CA4 (CA stands for *cornu ammonis*); however, often only CA1 and CA3 are distinguished. The subicular complex includes the subiculum, presubiculum, and parasubiculum.

The entorhinal cortex receives highly integrated multimodal information from many associative cortical areas via the perirhinal and parahippocampal cortices (mainly into the superficial layers) and also sends information back (mainly from the deep layers) (Lavenex & Amaral, 2000; Witter et al., 2000). It therefore seems to serves as an interface between the hippocampus and other cortical areas.

The different regions of the hippocampal formation can be ordered based on their connectivity in a loop like EC - DG - CA3 - CA1 - subiculum - presubiculum/parasubiculum - EC (Amaral & Witter, 1989; Chrobak et al., 2000; Lavenex & Amaral, 2000). From the superficial layers II/III of the entorhinal cortex originates the perforant path projecting to the dentate gyrus (mainly from II), the CA regions, and the subiculum (mainly from III). CA3 also receives input from dentate gyrus through the mossy fibers, which are not so numerous as the perforant path input fibers but have very large and strong synapses (Henze et al., 2000). The mossy fiber input is often thought to be the primary input into CA3, but see (Urban et al., 2001) for a critical discussion of this view. CA3 projects to CA1 via the Schaffer collaterals; CA1 projects to the subiculum and somewhat weaker also to the deep layers V/VI of the entorhinal cortex; the subiculum projects to the entorhinal cortex, partly via the pre- and parasubiculum. The direct projections from subiculum to entorhinal cortex terminate in the deep layers V/VI while those coming from pre- and parasubiculum terminate in the superficial layers II/III.

The dentate gyrus as well as the CA3-region also have some recurrent connectivity, i.e. from DG to DG cells and from CA3 to CA3 cells (Amaral & Witter, 1989; Witter, 1993). This recurrent connectivity is more direct and particularly strong in CA3 and has been estimated to be around 2% in rats (Amaral et al., 1990).

The connections between different regions are typically formed between principal cells (Lavenex & Amaral, 2000), which are excitatory. The principal cells in the dentate gyrus are granular cells, which are particularly small and densely packed; in the Ammon's horn the principal cells are pyramidal cells (Patton & McNaughton, 1995; Hammond, 2001). There are also various types of locally-connected inhibitory interneurons present in these regions (Freund & Buzsáki, 1996; Jones & Yakel, 1999). They are much less numerous than the principal cells.

The number of neurons varies between the different regions. In humans West & Slomianka (1998) have estimated that the entorhinal cortex has about 8.1 million neurons, which are distributed over the layers as follows (in millions): II: 0.7, III: 3.7, V: 1.2, and VI: 2.5. The number of neurons within the hippocampus and subicular complex have been estimated to be (in millions): $11 \pm 3$ in DG, $1.1 \pm 0.3$ in CA4, $2.3 \pm 0.6$ in CA2–3, $6.1 \pm 1.6$ in CA1, $4.6 \pm 1.1$ in subiculum, and $9.8 \pm 2.4$ in presubiculum (Harding et al., 1998). In rats it has been estimated that the number of neurons (in millions) is about 0.2 in EC layer II, 1.0 in DG, 0.33 in CA3, 0.42 in CA1, and 0.13 in the subiculum (Amaral et al., 1990, Fig. 1).

## 2.2 Psychophysics and lesion studies

Lesion studies have been instrumental in investigating the functional role of the hippocampal formation (Gluck & Myers, 2001, chap. 2). In humans lesions to the hippocampus and related regions lead to anterograde amnesia, i.e. an inability to form new memories, and graded retrograde amnesia, i.e. a loss of past memories which is more severe for recent events than for events long ago. This only applies to memories of facts and events, which constitute the declarative memory; nondeclarative memory, such as skill learning, seems to be unaffected. The effect of graded retrograde amnesia has led to the hypothesis that the hippocampus is required to acquire and consolidate memory, but that after some time (many years in humans) declarative memory is independent of the hippocampus, but see (Nadel & Moscovitch, 1997) for a critical discussion of this view.

In monkeys it has been found that lesions to the hippocampal formation lead to deficits in the delayed nonmatch to sample (DNMS) task, in which the monkey has to memorize an object and then after some delay (usually 5 to 60 seconds) avoid this object in a choice situation (Gluck & Myers, 2001). In lesioned monkeys performance is normal with little or no delay but poor if the delay is longer than a few seconds. Thus short-term memory seems to be intact but new memories that last for more than a few seconds cannot be formed. The same pattern has been found in humans and rats (Gluck & Myers, 2001).

In rats it has been shown that lesions to the hippocampal formation lead to impaired spatial learning and navigation, for example in the water maze task, in which the rat has to learn to swim to a platform hidden right below the surface in a pool of milky water (Morris et al., 1982; D'Hooge & De Deyn, 2001, sec. 3.2). Even though this task is very different from the DNMS task, it shares with it that the rat has to memorize and combine views of the surrounding (landmarks) in order to navigate successfully. Thus the reason for impairment may again be the inability to form appropriate memories (Gluck & Myers, 2001, sec. 2.3).

## 2.3 Neurophysiology

Much of the neural activity in the hippocampal formation is characterized by oscillatory firing patterns (Chrobak & Buzsáki, 1998; Chrobak et al., 2000; Buzsáki, 2002). During exploratory behavior and REM sleep the hippocampus shows so called gamma oscillations (40-100 Hz) amplitude modulated by slower theta oscillations (4-12 Hz). In the inactive state and during slow-wave sleep, sharp waves (50-150 ms long large amplitude field potentials) can be observed with associated short high-frequency bursts (200 Hz) referred to as ripples. The theta-modulated gamma oscillations seem to synchronize the input path as they propagate from the superficial layers II/III of entorhinal cortex (where they are generated by input from the medial septum and supramammillary nucleus) to the dentate gyrus and the CA-regions; the sharp waves/ripples seem to synchronize the output path as they originate in CA3, become most visible in CA1, and can also be observed in the deep layers V/VI of entorhinal cortex. It has been suggested that these two different physiological states correspond to a write-in or storage mode and a read-out or retrieval mode, respectively. Interestingly, one finds that in the sharp waves/ripples state, neural firing patterns are spontaneously reactivated that were earlier observed during exploration and in the theta/gamma state (Sutherland & McNaughton, 2000), which supports the idea of consolidation of memory during (slow-wave) sleep.

In rats it has been found that some cells in the hippocampus fire only when the rat is at a particular location (O'Keefe, 1979; Moser & Paulsen, 2001); such cells are referred to as 'place cells'. Similarly, cells selective to a particular view have been found in monkeys (Rolls, 1999). These types of cells are thought to be instrumental for spatial navigation, which we know is hippocampus dependent.

Experimental data suggest that overall the patterns of neural activity in the hippocampus are sparse compared to those of the entorhinal cortex or the subiculum (Barnes et al., 1990; Jung & McNaughton, 1993). O'Reilly & McClelland (1994) have estimated from such data activity levels of 0.4% in dentate gyrus, 2.5% in CA1 and CA3, but 7% in entorhinal cortex and 9.2% in subiculum.

The hippocampal formation is highly plastic (Kandel, 2000; Henze et al., 2000; Abraham & Williams, 2003). Depending on the neural activities (in experiments usually governed by the stimulation protocol) the synaptic efficacies can increase or decrease. This effect can be long lasting (several hours up to weeks) and is then referred to as long-term potentiation (LTP) and long-term depression (LTD), respectively. LTP and LTD can occur in the same synapse (e.g. Dudek & Bear, 1992). LTP in the perforant path (DG) and Schaffer collaterals (CA1) is associative, i.e. it requires pre- and postsynaptic activity while LTP in the mossy fiber synapses (CA3) can also be nonassociative, i.e. it only depends on presynaptic activity, if a long-lasting high-frequency stimulus is given. LTP and LTD are generally considered to be the underlying mechanism of memory formation (Martin & Morris, 2002), but see (Hölscher, 1999) for a critical discussion of this view.

## 2.4 Adult neurogenesis

One peculiar property of the dentate gyrus is that it generates new neurons throughout life, a phenomenon referred to as adult neurogenesis (Altman & Das, 1965; Gould & Gross, 2002; Kempermann et al., 2004). The new cells arise from precursor cells and differentiate into granular cells that appear to become fully functional and integrated over a time course of several weeks (van Praag et al., 2002; Jessberger & Kempermann, 2003). However, young granular cells seem to differ from old ones in that they show enhanced long-term potentiation (LTP), which has a lower threshold of induction (Schmidt-Hieber et al., 2004) and cannot be inhibited by GABA (Wang et al., 2000; Snyder et al., 2001). It has also been argued that it is reasonable to assume that new cells form synaptic connections more rapidly (Gould & Gross, 2002).

The number of newly added neurons depends on the rate of proliferation, i.e. cell generation, and the probability of cell survival, since most of the newly generated cells actually die again after a short period of time (apoptosis) (Eriksson et al., 1998; Gould et al., 2001; Kempermann et al., 2003). If they survive the initial phase, they can last for a long time, e.g. at least 1 year in mice (Kempermann et al., 2003), 8 months in rats (Altman & Das, 1965), and 2 years in humans (Eriksson et al., 1998). Cell proliferation decreases with age (Altman & Das, 1965; Seki & Arai, 1995; Kempermann et al., 1998) and can also be reduced by stressful experiences (Gould et al., 1997, 1998; Tanapat et al., 2001). Both might be explained by the dependency of the proliferation on hormones (glucocorticoids); other hormones (ovarian steroids / estrogen) stimulate the proliferation (see Gould & Gross, 2002, for an overview). Other factors that increase neurogenesis are dietary restriction (rat) (Lee et al., 2000), voluntary physical activity (mouse, running but not swimming) (van Praag et al., 1999a,b; Kronenberg et al., 2003), enriched environments (mouse, rat) (Kempermann et al., 1998; Nilsson et al., 1999; van Praag et al., 1999b), and some hippocampal-dependent learning-tasks (Gould et al., 1999), the former factors effecting more the proliferation the latter ones effecting more the cell survival. Genetics also has a strong influence on neurogenesis (Kempermann et al., 1997; Kempermann & Gage, 2002). The total number of new cells generated per day by proliferation in the dentate gyrus of young adult rats (9 weeks old) has been estimated to be 9000 (Cameron & McKay, 2001), only about 60% of which survive the first four weeks. The effect of neurogenesis seems to be accumulative, but since its rate decreases quickly with age the overall addition of new neurons over the whole lifetime is relatively small. Kempermann et al. (1998) have estimated an overall increase in mice of about 40,000 (approx. 10%); more recent results indicate that it might be more in the range of 30% (Kempermann et al., unpublished data).

Like behavioral and environmental factors have an effect on the generation and survival of new neurons, there is also some evidence that neurogenesis affects performance in some hippocampal-dependent learning-tasks, such as trace conditioning (Shors et al., 2001, 2002) or the acquisition in the water maze task (Kempermann & Gage, 2002). Reduced neurogenesis might contribute to major depression (Santarelli et al., 2003; Kempermann & Kronenberg, 2003).

# 3 Artificial neural networks and catastrophic interference

## 3.1 Feedforward and Hopfield networks

Like physiological neural networks, artificial neural networks consist of a number of relatively simple computational units that are mutually connected (see Hertz et al., 1991, for an introduction). Each unit integrates the activity it receives from other units via the connections, performs some simple stereotype operation on it, and sends its own activity to other units it is connected to. The connections have weights that determine how much one unit contributes to the activity of another unit and thereby what kind of computation the whole network performs. It is characteristic for neural networks that the computation is distributed over many weights, which has advantages and disadvantages. There are a number of learning rules that adapt the weights such that the network performs a certain desired computation, such as pattern recoding or memory storage and retrieval. Feedforward networks (see Bishop, 1995, for an introduction) and Hopfield networks (see Amit, 1989, for an introduction) are two classes of artificial neural networks that are commonly used for modeling hippocampal function.

Feedforward networks have an input layer of units, possibly several intermediate (so called hidden) layers, and an output layer. As the name indicates connections only go from earlier layers to later layers. Thus the network has no particular dynamics but simply computes an output based on an input. It is therefore equivalent to a mathematical function with vectors as an input and vectors as an output. Feedforward networks are usually trained in a supervised fashion, which means that the network gets told explicitly what output it should generate for a given input. Thus the training patterns are actually input-output pairs, for each input there is a desired output and the training procedure is designed such that the difference between desired output and actual output is minimized.

Hopfield networks are very different. Every computational unit serves as an input as well as an output unit and, in its simplest form, can only assume the values 0 or 1. All units are reciprocally connected, so that the network has a highly recurrent connectivity. Since each unit has an influence on every other unit, such a network can have quite a complex dynamics. However, it is typically designed such that after some time the network activity settles in a stable state (a certain combination of zeros and ones) and stays there. Training a Hopfield network means choosing the connection weights such that the stable states the network can settle in correspond to a given set of desired patterns. These are the stored patterns. If one initializes the activity of the network with a corrupted version of a stored pattern, the dynamics of the network will typically arrive at the stable state that corresponds to the uncorrupted pattern. This effect is known as auto-association and is the main advantage of a Hopfield network.

## 3.2 Catastrophic interference

Due to the distributed nature of the computation performed in a neural network, learning a new pattern typically requires to change all weights. This poses a problem. Assume the weights have been trained for input pattern **a**, so that the network behaves as desired. Then training pattern **b** is given and all weights change to be optimal for input **b**. This will typically degrade the performance on pattern **a**. If a third pattern **c** is learned, the performance on **a** will degrade even further and very quickly pattern **a** will be forgotten completely. This effect can be quite dramatic and is referred to as *catastrophic interference* (McCloskey & Cohen, 1989, cited in McClelland et al., 1995).

In feedforward networks the problem of catastrophic interference is typically solved by interleaved training, which means that patterns **a**, **b**, and **c** in the example above are presented repeatedly in an interleaved fashion, like **abcabcabcabc** or more irregularly **abbcbaaccbacccaabacbb**, with a relatively small learning rate, which makes the weights change only little per pattern presentation. The patterns are therefore effectively learned simultaneously and no pattern dominates another one. Interleaved learning, however, requires that all training patterns are available all the time. A sequence like **aaaabbbbcccc** would not work.

One way to avoid catastrophic interference in a feedforward network even if the training sequence is of the form **aaaabbbbcccc** is to use a dual network (e.g. Ans & Rousset, 1997; French, 1997). Such an architecture takes advantage of the fact that some networks can spontaneously reproduce patterns that they have learned earlier. In a dual network two such networks are coupled and (partially) train each other. Assume both networks have learned pattern **a** already. During training of the next pattern, **b**, network 1 is trained and network 2 is used to generate samples of previously learned patterns for network 1, in this case $\mathbf{a}_2$, where the index 2 indicates that $\mathbf{a}_2$ is not an original training pattern but one that has been spontaneously reproduced by network 2. Thus network 1 may receive the training sequence $\mathbf{a}_2\mathbf{b}\mathbf{a}_2\mathbf{a}_2\mathbf{b}\mathbf{a}_2\mathbf{b}\mathbf{b}\mathbf{a}_2$. If no training of new patterns happens, network 1 generates training patterns for network 2, in this case something like $\mathbf{a}_1\mathbf{b}_1\mathbf{b}_1\mathbf{a}_1\mathbf{a}_1\mathbf{b}_1\mathbf{a}_1\mathbf{a}_1\mathbf{b}_1$, with index 1 indicating that the patterns have been generated by network 1. Next network 1 learns new pattern $\mathbf{c}$ and is trained with a sequence like $\mathbf{a}_2\mathbf{b}_2\mathbf{c}\mathbf{b}_2\mathbf{c}\mathbf{a}_2\mathbf{b}_2\mathbf{b}_2\mathbf{c}\mathbf{a}_2\mathbf{b}_2\mathbf{c}\mathbf{a}_2$. Then network 2 is updated again and trained with a sequence like $\mathbf{a}_1\mathbf{b}_1\mathbf{a}_1\mathbf{c}_1\mathbf{b}_1\mathbf{c}_1\mathbf{c}_1\mathbf{a}_1\mathbf{c}_1\mathbf{a}_1\mathbf{b}_1\mathbf{c}_1$. Thus networks 1 and 2 train each other in turns by interleaved training and new patterns enter network 1. None of the networks has the problem of catastrophic interference. However, even in this architecture there may be some degradation of learned patterns in the course of learning new patterns, because after the initial learning of a new pattern the stabilization is always done with a spontaneously reproduced version of it, which may be somewhat corrupted.

Hopfield networks solve the problem of catastrophic interference very differently. They can be easily setup in such a way that orthogonal patterns, i.e. patterns that are uncorrelated, do not interfere with each other at all. Thus if a set of patterns has to be stored in a Hopfield network, it is good to first encode them such that the transformed patterns are orthogonal to each other. They can then be stored without any interference. Of course, one has to also memorize the encoding and in addition the encoding must be invertible, otherwise one could not retrieve patterns from the memory anymore.

One way to make patterns orthogonal is to generate a very sparse representation of them (Bentz et al., 1989). This means that they are encoded such that only few units are significantly active at a time. If only very few units are active, there is a good chance that two different patterns have no active units in common and are therefore orthogonal. This technique might require to increase the number of units for representing the patterns, though. To compensate for this increase one can compress the patterns first, so that irrelevant dimensions are eliminated and only relevant dimensions are represented.

Even with a sparse representation of the patterns, if there are more and more patterns stored in the Hopfield network, at some point the capacity will be exhausted, and new patterns will interfere with old ones. This effect again can be quite dramatic, so that all in a sudden none of the patterns can be retrieved anymore, even those that have been stored recently. Measures can be taken to avoid also this type of catastrophic interference, at least to some extent. For example with the learning of each new pattern all previous weights could be decreased a little bit, so that old patterns are gradually forgotten and make room for new patterns.

When discussing the problem of catastrophic interference, which results from the distributed representation in neural networks, one should not forget the advantages that distributed representations provide. For instance, they are much more robust with respect to damage like the loss of units or connections, a property known as *graceful degradation*. More importantly, feed forward networks can generalize. Only a limited number of training patterns can be learned exactly. As the number of training patterns increases and the network cannot memorize all of them individually, the network has to find an input-output function that is a compromise and provides a reasonably good performance on all training patterns. In finding this compromise the network discovers and employs regularities in the data which results in the nice effect that the network will generate reasonable responses also for new patterns never seen before. This property is referred to as *generalization*.

The problem of catastrophic interference is somewhat related to the *stability-plasticity dilem-*

*ma* (Carpenter & Grossberg, 1987, cited in Gerstner & Kistler, 2002). The latter refers to the problem of setting the right learning rate. If the rate is too low the network cannot learn new patterns; if the rate is too high the network can learn new patterns well, but old patterns are forgotten quickly, catastrophic interference occurs.

Catastrophic interference is very distinct from the *bias-variance dilemma* (Geman et al., 1992), which refers to the problem of choosing the right network complexity (does not apply to Hopfield networks). If the network is too simple (large bias case) it will not even be able to represent the training patterns well and will produce a large error; if it is too complex (large variance case) it can memorize all training patterns individually and does not need to discover regularities, thus it will not generalize well.

# 4 A model of hippocampal function

In our considerations we adopt a standard model of hippocampal function as proposed and discussed previously (Treves & Rolls, 1994; McClelland et al., 1995) and summarized in this section.

We have seen above that neural networks tend to have the problem of catastrophic interference and that measures have to be taken to avoid it. It is clear that the cortex is neither simply a feedforward network nor a Hopfield network. However, conceptually it shares important properties with artificial neural networks and will most likely have the same problem of catastrophic interference. If in particular we have in mind its ability to learn from examples by exploiting their regularities and thereby generalizing to new situations, we see that it would be useful to store examples temporarily and permit the cortex to perform some variant of interleaved training. The function as a temporary storage of patterns is usually assigned to the hippocampus with the entorhinal cortex serving as an interface between cortex and hippocampus.

Simply storing input patterns can be done in a Hopfield network. Characteristic for a Hopfield network is its highly recurrent connectivity. This fits well to the recurrent connectivity in the CA3 region of the hippocampus, and it is mainly for this reason that CA3 is often considered the actual site of storage in the hippocampus.

We have seen above that a Hopfield network has two versions of the catastrophic-interference problem. The first can be solved by making the input patterns orthogonal, which can, for instance, be done by generating a sparse representation. This role is usually assigned to the dentate gyrus for two reasons. Firstly, it provides the main input to CA3 and secondly it has a very sparse activity.

In order to use the CA3 memory efficiently, it is assumed that dentate gyrus must also learn to select the important dimensions of the patterns and discard others that are redundant, thereby performing a lossy compression.

There are different views on whether CA3 also suffers from the second version of the problem of catastrophic interference. Theoretically, it is simply a question of capacity. Either one assumes that CA3 has sufficient capacity for lifelong storage of patterns, then there is no interference problem due to an exhausted capacity, or the capacity is not sufficient, then the measures described above have to be taken, so that old patterns are gradually forgotten to make room for new patterns. This issue does not effect our ideas about the functional role of neurogenesis much, so we will not discuss it further.

If one wants to retrieve the encoded patterns in CA3, one also needs a decoding network. This role is commonly assigned to CA1 (and the subiculum), mainly for the reason that there are strong connections from CA3 via CA1 and subiculum back to the entorhinal cortex, which is thought to be the interface between cortex and hippocampus.

In summary we have the following picture. The hippocampus serves as a temporary (or even permanent) storage that can store new patterns quickly. CA3 serves as the actual storage site and works like a Hopfield network. Dentate gyrus performs an encoding of the patterns to make them suitable for storage. This includes compression and sparsification. CA1 and subiculum perform a decoding of patterns retrieved from CA3. The entorhinal cortex serves as an interface between hippocampus and cortex. In the following, patterns of activity arriving at the entorhinal cortex will be indicated by $\mathbf{x}$; patterns encoded by the dentate gyrus and possibly stored in CA3 will be indicated by $\mathbf{x}'$; encoded and possibly stored patterns decoded again by CA1 and subiculum and sent back to the entorhinal cortex will be indicated by $\mathbf{x}''$. In general $\mathbf{x}''$ should be similar to $\mathbf{x}$ for successful retrieval of patterns from CA3. When we refer to specific environments the animal lives in, we will use $\mathbf{a}$ and $\mathbf{b}$ instead of $\mathbf{x}$.

There are a number of important issue we do not consider in this paper, such as the processing of temporal pattern sequences, the details of the storage and retrieval process, the function of the various oscillations one finds in the hippocampus, and the consequences of having spike trains instead of graded responses. However, we believe that these issues do not interfere too much with the issue of neurogenesis the way addressed here.

# 5 Why new neurons?

What might be the role of new neurons in the dentate gyrus in the overall framework given above? Why are they generated only in the dentate gyrus and not in CA3 or CA1? Why is it sufficient to add only about 30% new neurons over the whole lifetime? Why should adult neurogenesis be regulated by activity or richness of the environment?

## 5.1 Avoiding catastrophic interference in the dentate gyrus with new neurons

In the framework presented above, the dentate gyrus has the role of encoding input patterns coming from entorhinal cortex such that they can be efficiently stored in CA3. This includes sparsification and compression. The optimal encoding depends, of course, on the distribution of the input patterns or, in other words, on the environment the animal lives in. When the environment changes, the encoding should change, too, to guarantee an optimal encoding and efficient storage of new patterns.

However this adaptation of the dentate gyrus to the environment conflicts with the requirement that also older patterns should be retrievable from CA3. Imagine an animal lives in environment $A$ for a while and has its dentate gyrus fully adapted to the distribution of stimuli it encounters in $A$. Consider one stimulus that gives rise to pattern $\mathbf{a}$ in the entorhinal cortex. This pattern is encoded by the dentate gyrus to produce pattern $\mathbf{a}'$ suitable for storage in CA3. Now imagine the animal moves to environment $B$ with a different distribution of stimuli. Again dentate gyrus should adapt to perform an optimal encoding of patterns $\mathbf{b}$ in this new environment. However, after adaptation the old pattern $\mathbf{a}$ will be encoded differently and produce a pattern of activity $\mathbf{a}'_{\text{new}}$ that differs significantly from the originally stored pattern $\mathbf{a}'$, thereby making it impossible to retrieve pattern $\mathbf{a}$ from CA3.

Thus we see that the dentate gyrus also has the problem of catastrophic interference. As it adapts to a new environment, the encoding learned in the preceding environment gets lost quickly. The solution of interleaved training cannot be applied, because the animal does not have arbitrary access to the different stimuli to generate an interleaved stimulus sequence nor does the hippocampus have an intermediate buffer, which in any case would have the same problem of catastrophic interference again. Thus, the dentate gyrus must have another strategy to be able to adapt to a new environment without forgetting the encoding learned in the old environment.

Our hypothesis is that the problem of catastrophic interference in the dentate gyrus can be avoided by extending the encoding by adding new neurons instead of changing it by modifying existing synaptic weights; old neurons and their synaptic weights are fixed and preserve the old code while new neurons are added and provide those features that are novel. In Section 6 it will be shown with a computational model that this is indeed a possible strategy to avoid catastrophic interference.

## 5.2 Why no new neurons in CA3 and CA1?

If the dentate gyrus has the problem of catastrophic interference and can solve it by adding new neurons, wouldn't neurogenesis also be helpful in CA3 and CA1?

If we take a Hopfield network as a good model of CA3, then it is clear it does not have the problem of catastrophic interference in the same way as the dentate gyrus. Adding a few new neurons in a Hopfield network would neither help much in making patterns orthogonal nor would it increase the capacity significantly. Therefore neurogenesis does not seem a good strategy in CA3.

The situation is different in CA1. Dentate gyrus and CA1 are complementary networks. One performs the encoding; the other performs the decoding approximately inverting the encoding. It is indeed the case that, due to the close relationship between en- and decoding, CA1 has the problem of catastrophic interference to the same extent as the dentate gyrus. However, for exactly this reason it is sufficient to solve the problem in the dentate gyrus alone. If the encoding is stabilized, the optimal decoding learned by CA1 is stabilized as well. Therefore no new neurons are necessary in CA1. Thus we assume that while dentate gyrus employs new neurons to avoid catastrophic interference, CA1 will simply always adapt such that it optimally inverts the encoding performed by dentate gyrus.

In the model simulation in Section 6 we will see that this argument is not quite as straight forward as it appears here, because the optimal decoding depends not only on the encoding but also on the pattern statistics. But we will also see that it is basically valid in the case of neurogenesis. On the other hand, if one really wanted to employ new neurons in CA1 as well, it might be difficult to coordinate the learning in DG and CA1, for instance the new CA1 might need to 'know' which DG-neurons are old (fixed) and which ones are new (plastic) in order to take advantage of the new neurons in CA1, something that does not seem feasible.

## 5.3 Why are 30% enough?

In a naive view one could argue that new neurons have to be added whenever new stimuli occur and have to be stored. Thus the number of neurons should increase linearly with the number of stored patterns. It is clear that 30% new neurons added over the whole lifetime it too little to support this view.

To understand why 30% might actually be sufficient it is important to make a distinction between *new* and *novel* stimuli. A new stimulus, in our view, is a stimulus that has not occurred before but is made up of features that are all known, much like a new word might occur in text, not seen before but made up of letters that are all known. A novel stimulus, in contrast, is not only new but also contains features not encountered before, much like a Chinese character would be novel to people used to Latin characters. We assume that a new stimulus can be represented without difficulty with the existing code, while a novel stimulus actually requires an extension of the code to represent the novel features. It seems reasonable to assume that new stimuli occur with almost constant rate throughout life while novel stimuli occur primarily early in life or if the environment changes significantly. Thus overall 30% new neurons might be sufficient to account for novel stimuli occuring after the early phase of life.

## 5.4 How is adult neurogenesis regulated?

New neurons are generated in the dentate gyrus all the time. Some of these neurons are integrated in the DG-network and become functional, but most of them are not used but removed through apoptosis (cell death). Both processes, the generation of new neurons and the integration, are regulated and depend on behavioral parameters.

One problem of this regulation are the different time scales. Since the generation and integration of new neurons takes weeks, the animal cannot start generating new neurons when they are needed. It must have some of them in stock all the time; the greater the chance of encountering novel stimuli the more of the new neurons in stock are needed.

This problem of different time scales and delay might explain two observations. Firstly, if we assume that new neurons not integrated in the network cannot be maintained for a long time, it is clear that new neurons must be generated all the time and most of them are removed again just to have some available quickly when needed. Secondly, in order not to waste too many resources on the generation of new neurons, neurogenesis should be regulated depending on what the expectancy of the occurrence of novel stimuli is. However, since the generation of new neurons is such a slow process, the prediction of the occurrence of novel stimuli is fairly unreliable in any case which makes unspecific indicators such as simply physical activity a reasonable predictor.

# 6 A simple computational model of neurogenesis in the dentate gyrus

## 6.1 Methods

### Network architecture

As a first step, we have decided to model the encoding and decoding performed by dentate gyrus and CA1, respectively, with a linear autoencoder network; see Figure 1. Such a network consists of an input layer, a smaller hidden layer, and an output layer that has as many units as the input layer. We identify the input layer with layer II of the entorhinal cortex (EC-II), the hidden layer with the dentate gyrus (DG), and the output layer with layers V/VI of the entorhinal cortex (EC-V/VI). For storage and retrieval experiments we take the hidden layer also as a representation of CA3, implicitly making the simplifying assumption of a one-to-one connectivity between DG and CA3. We do not have a layer representing the CA1 region, but the mapping realized by the connectivity from the hidden to the output layer is interpreted as the decoding performed by CA1. One could extend the model by inserting two additional layers between the hidden and the output layer, one for CA3 and one for CA1 (possibly plus one more for the subiculum). But in this first step we keep the network as simple as possible and consider just one hidden layer representing the dentate gyrus and in some cases also the CA3 region.

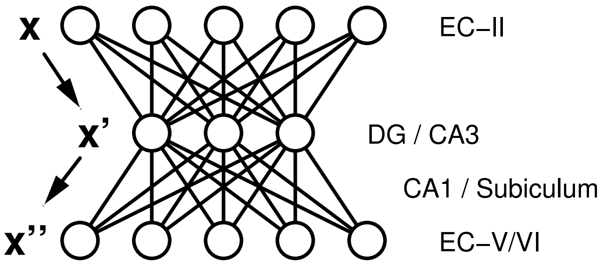



**Figure 1: Architecture of the linear autoencoder network.** The input and output layer represent layer II and layers V/VI of the entorhinal cortex, respectively. The hidden layer represents the dentate gyrus and in the retrieval experiments also CA3. CA1 and subiculum are not represented explicitly but summarized in the hidden-to-output-layer connections. The whole network is linear. In the simulations input and output layer each had 60 units and the hidden layer had 15 units plus 5 units due to neurogenesis.

Since we will consider a number of different test scenarios, we need a concise notation of network architecture and training procedure. The number of units in each layer will be denoted by lowercase letters and little triangles pointing from one layer to the next will denote the direction of connectivity. Thus $n \triangleright m \triangleright n$ indicates a network with $n$ input units, $m$ hidden units, and $n$ output units. As mentioned above, an autoencoder network has always the same number of units in the input and output layer. The hidden layer may undergo neurogenesis in which case the network may change from $n \triangleright m \triangleright n$ to $n \triangleright (m + l) \triangleright n$, where $l$ is the number of added units. This basic notation will be extended below to include information about the training procedure.

### Network computation

The input layer represents the input patterns and does not perform any computation. The hidden layer performs an encoding and the output layer performs a decoding. Both layers are linear, and so is the whole network, since a combination of linear functions is also linear. The name 'autoencoder network' indicates that the task of the network is simply to reproduce its input in the output layer. However, since the hidden layer has fewer units than the input and output layer, the network has a bottleneck, which enforces a dimensionality reduction and therefore a compression from the input to the output layer. Compression is one of the hypothesized functions of the dentate gyrus that is captured by the model. We do not consider sparsification, since this is a nonlinear operation, which is not within the scope of the model. Such a model may seem ridiculously simple, but we will see that it provides some interesting insights into the possible role of adult neurogenesis. The simplicity is actually a great advantage, because it permits an analytical treatment of the network (see Appendix A).

Now consider an input vector $\mathbf{x}$ representing the pattern of activity in EC-II. We have argued above that it will be propagated through dentate gyrus, CA3, CA1, and subiculum back to layers V/VI of EC (EC-V/VI) and will, on its way, undergo several transformations. In our simplified model, we just have two of these stages, the hidden layer corresponds to dentate gyrus and the output layer corresponds to the reconstruction of the pattern in EC-V/VI. The representation of input vector $\mathbf{x}$ in the hidden layer (dentate gyrus) will be referred to as $\mathbf{x}'$; the reconstruction in entorhinal cortex will be referred to as $\mathbf{x}''$.

Figure 2 illustrates optimal en- and decoding for a given data distribution. It is known from principal component analysis that the encoding subspace as well as the decoding subspace should ideally lie in the direction of maximal variance (with identical orientation, of course) (see Appendix A.2).

However, how should the decoding be done if the encoding was not optimal? Figure 3 illustrates that neither the original direction of encoding nor the direction of maximal variance are optimal. A direction in between these two cases is actually optimal including some scaling. If the data distribution is more spherical with similar variance in all directions then the optimal decoding should lie close to the direction of the encoding; if the data distribution is more elongated then the optimal decoding should lie close to the direction of maximal variance (see Rasch, 2003, and Appendix A.4).

Finally, as Figure 4 illustrates, if the encoding is optimal, but the decoding is random, the recoding error is large. It is often actually greater than if one would not decode anything, i.e. set all output vectors to zero. Thus, no decoding is better than random decoding.

### Network adaptation

We assume the animal lives in two different environments, which we denote by uppercase letters $A$ and $B$. Within these environments the animal encounters different stimulus situations (or events), which give rise to different patterns of activity in EC-II representing the stimulus situation. Such patterns will be denoted by vectors $\mathbf{a}$ and $\mathbf{b}$ (instead of $\mathbf{x}$) for the environments $A$ and $B$, respectively. The distribution of stimulus situations depends, of course, on the environment. We model this by drawing the input vectors $\mathbf{a}$ and $\mathbf{b}$ from two different multidimensional Gaussian distributions. The distributions in the different envi-

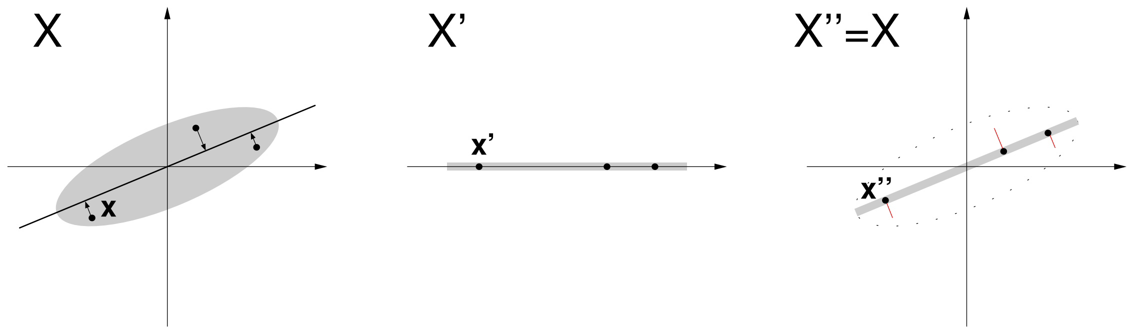



**Figure 2: Optimal recoding.** Illustration of the computation performed by an autoencoder network with two input and output units and just one unit in the hidden layer, i.e. with the architecture $2 \triangleright 1 \triangleright 2$. **Left:** The stimuli are represented by points $\mathbf{x}$ in a two-dimensional space $X$; the gray ellipse indicates the distribution of data points (stimuli) and three points are shown explicitly. During encoding, the two-dimensional distribution of points has to be reduced to a one-dimensional distribution, since there is only one unit in the hidden layer. In the linear case this corresponds to a projection (indicated by the arrows) onto a line (neglecting any scaling for simplicity and without loss of generality). In this example the line lies in the direction of maximal variance of the distribution, which is the optimal choice in order to minimize the reconstruction error. **Middle:** In the hidden layer, the stimuli are now points $\mathbf{x}'$ in a lower-dimensional space $X'$, thus some information is lost. In this example $X'$ is only one-dimensional. **Right:** In order to obtain a reconstruction of the original high-dimensional data points, the data points $\mathbf{x}'$ of the space $X'$ are embedded in a high-dimensional space $X''$ which can be identified with the original space $X$. In this case the optimal direction of embedding is again in the direction of maximal variance of the data distribution. The recoding error made by the transformation from $\mathbf{x}$ to $\mathbf{x}'$ is indicated by the thin solid lines (red).

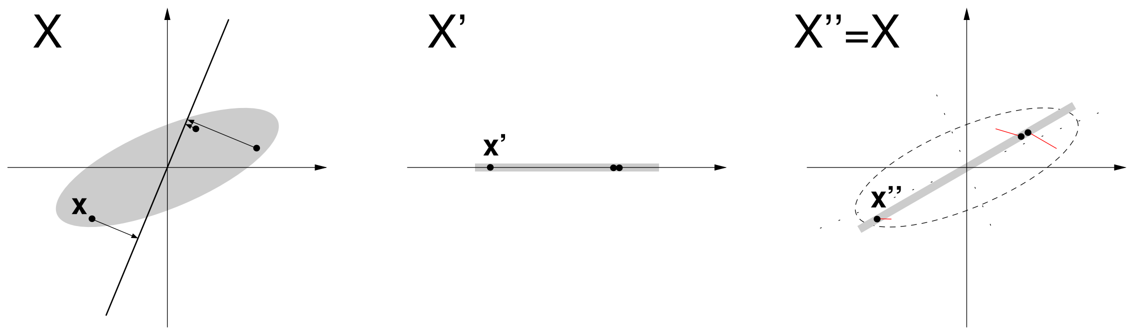



**Figure 3: Optimal decoding given random encoding.** If the encoding is suboptimal, i.e. the subspace to project onto does not lie in the direction of maximal variance, the optimal decoding is neither in the direction of the original encoding nor in the direction of maximal variance but somewhere in between.

ronments differ only in their orientation in space. Thus, the distribution in environment $B$ is a rotated version of the distribution in environment $A$. Mathematically this means that the eigenvalues of the covariance matrices are identical. Figure 5 illustrates data distributions

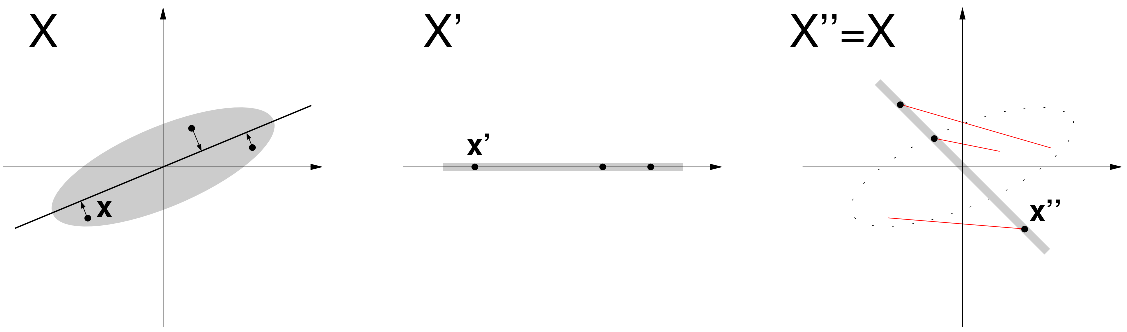



**Figure 4: Random decoding given optimal encoding.** If the encoding is optimal but the decoding is random, the recoding error can be greater than if all output vectors would simply be set to zero.

of the two different environments and shows the eigenvalues of the actual distributions used in the simulation experiments.

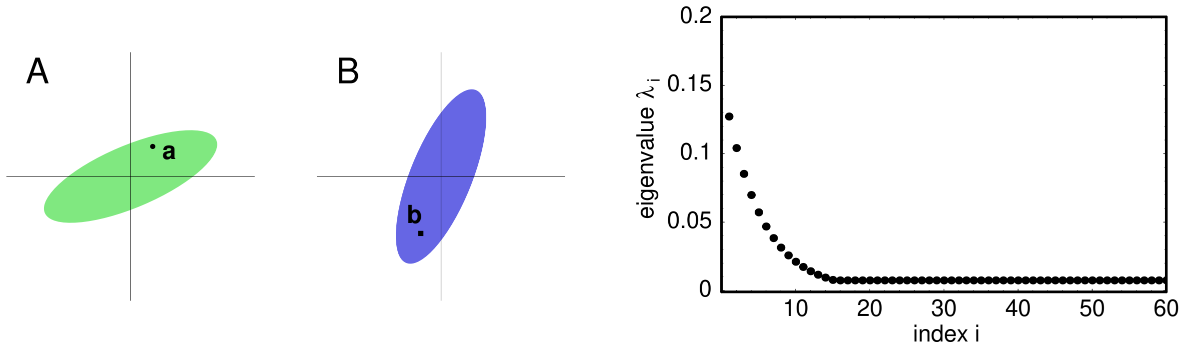



**Figure 5: Distribution of stimulus vectors a and b in environments** $A$ **and** $B$**. Left:** Schematic drawing of the two distributions. They differ only in their orientation in space but not in any other statistical characteristics. **Right:** Eigenvalues of the covariance matrices of the stimulus distributions in environments $A$ and $B$. They correspond to the variances along the principal components of the distributions. The graph illustrates that out of the 60 principal components of the stimulus vectors the first 15 carry most of the variance. The other 45 are interpreted as noise.

Since the input layer only serves as a representation of the input and does not perform any computation, it is only the hidden layer and the output layer that can adapt to a particular environment. This will be denoted by subscripts to the indices of the number of units per layer. $n \triangleright m_A \triangleright n_A$, for instance, indicates that the hidden layer as well as the output layer are optimally adapted to environment $A$. We assume that units in the hidden layer always adapt jointly with the output layer. Units in the output layer, however, can adapt with the connections from the input to the hidden units fixed. Assume the animal with a network adapted to environment $A$, i.e. $n \triangleright m_A \triangleright n_A$, moves to environment $B$. One can imagine that the hidden layer does not adapt but the output layer does. The result would be a network with a hidden layer adapted to environment $A$ and an output layer adapted to $B$ given a hidden layer adapted to environment $A$. Such a network is denoted by $n \triangleright m_A \triangleright n_B$. Note that the adaptation of the output layer to environment $B$ depends on the hidden layer, so that the networks $n \triangleright m_A \triangleright n_B$ and $n \triangleright m_B \triangleright n_B$ actually have different output layers. In our context it is particularly interesting to consider a situation where we add some neurons to the hidden layer when the animal has moved to environment $B$ which would result in the network architecture $n \triangleright (m_A + l_B) \triangleright n_B$. The hidden layer now contains $m$ old units still adapted to $A$ and $l$ new units adapted to $B$. The output layer is fully adapted to environment $B$ given the hidden layer with the $(m_A + l_B)$ units. In the results presented here we always impose the constraint that weight vectors of different hidden units are normalized to one and orthogonal to each other, i.e. maximally different.

### Storage and retrieval

Although our model does not include an explicit CA3 layer, we can still address the question of how well patterns can be stored and retrieved under different scenarios. We assume the role of a CA3 layer would only be to store patterns coming from the dentate gyrus and later recall them. Thus patterns in CA3 would be identical to those coming from the dentate gyrus, which is the reason why we do not need to represent CA3 by an extra layer. Consider the network $n \triangleright m_A \triangleright n_A$. Given an input vector $\mathbf{a}$ this network would produce a hidden vector $\mathbf{a}'$ and an output vector $\mathbf{a}''$. The hidden vector would be stored, which we denote with $\mathbf{a}'_\square$, the square indicating the storage process. Obviously $\mathbf{a}'_\square = \mathbf{a}'$.

In a later stage the network may have changed, let say it has adapted to environment $B$, so that it has changed to $n \triangleright m_B \triangleright n_B$. The question now is how well the original pattern $\mathbf{a}$ is reconstructed when the stored pattern $\mathbf{a}'_\square$ is retrieved from CA3. We probe that by decoding pattern $\mathbf{a}'_\square$ with the new output layer of network $n \triangleright m_B \triangleright n_B$. The decoded stored pattern is denoted by $\mathbf{a}''_\square$. In other words, $\mathbf{a}''_\square$ results from encoding $\mathbf{a}$ with network $n \triangleright m_A \triangleright n_A$ and decoding it with network $n \triangleright m_B \triangleright n_B$.

In case of neurogenesis, the number of units in the hidden layer changes. For instance the network may first be $n \triangleright m_A \triangleright n_A$ in environment $A$ and then change to $n \triangleright (m_A + l_B) \triangleright n_B$ in environment $B$. Since the stored vector $\mathbf{a}'_\square$ has only $m$ components, it does not have the right size to be decoded by the second network. We solve this problem by simply adding the required number of components and setting them to zero. For instance, if $m = 3$ and $l = 2$ the hidden vector stored by the first network might be $\mathbf{a}'_\square = (0.3, -0.2, 0.4)^T$ and before decoding it with the second network it would be extended by two zeros to become $\mathbf{a}'_\square = (0.3, -0.2, 0.4, 0, 0)^T$. The latter can be used for decoding by the second network to produce $\mathbf{a}''_\square$.

### Performance measures

To assess the performance of the autoencoder network we determine two kinds of reconstruction errors. Firstly, we are interested in how well the patterns in EC-II are reconstructed in EC-V/VI and how large the error is due to the recoding. Thus, we compare the input vector $\mathbf{a}$ with the output vector $\mathbf{a}''$ and call the mean squared difference the recoding error. Secondly, we are interested in how well patterns stored in CA3 can be reconstructed when they are retrieved (spontaneously or by a cue). The corresponding error is called the retrieval error. To get reliable measures we average over a great number of input vectors indicated by angle-brackets $\langle \cdot \rangle$.

$$\text{recoding error:} \quad E := \left\langle |\mathbf{a} - \mathbf{a}''|^2 \right\rangle \tag{1}$$

$$\text{retrieval error:} \quad E_\Box := \left\langle |\mathbf{a} - \mathbf{a}''_\Box|^2 \right\rangle \tag{2}$$

We will also consider the recoding error for patterns **b**.

## 6.2 Results

Assume our model animal first lives in environment $A$, then it moves to a new environment $B$, and finally it returns to the old environment $A$. For best performance in this setting the autoencoder network should adapt somehow to the different pattern statistics of the two environments. In this section we consider three different adaptation strategies, which are illustrated in Figure 6. In all three cases the decoding part of the network (hidden-to-output-layer connections) always fully adapts to the current environment to minimize the recoding error. The encoding part (input-to-hidden-layer connections), however, may have one of the following three strategies:

(a) **No DG-adaptation:** In this strategy (Fig. 6 left) we assume there is something like a *critical period* in which the animal adapts to environment $A$. After that no adaptation takes place in the encoder part of the network anymore, so that the encoder does not change when the animal moves to environment $B$. The hidden layer has 20 units.

(b) **Neurogenesis:** In this strategy (Fig. 6 middle) the network starts with a few units in the hidden layer and new units are added as required. Only the most recent hidden units adapt to a new environment to account for those feature dimensions that are not yet well represented. Here we start with 15 hidden units that adapt to environment $A$ and add five new units after the animal has moved to environment $B$.

(c) **Full adaptation:** In this strategy (Fig. 6 right) all neurons always fully adapt to the current environment, first $A$ then $B$. The hidden layer has 20 units.

Note that even though the decoding part of the network always fully adapts to environment $B$ after the animal has moved there, this does not mean that the decoding part is identical for the three different strategies. The reason is that the optimal decoding, of course, depends on the encoding, and since the encoding depends on the strategy, the optimal decoding will do as well. Furthermore, the optimal decoding also depends on the statistics of the patterns, so that it can be different even if the encoding is identical.

To assess the performance of the system we will consider five different errors. The first two are determined for the network adapted to environment $A$ and the other three for the network adapted to environment $B$.

(A) **Network adapted to environment A.**

(i) **Recoding error A:** When adapted to environment $A$ the recoding error of patterns **a**, i.e. $\langle |\mathbf{a} - \mathbf{a}''|^2 \rangle$, should obviously be small. The retrieval error for patterns **a** should also be small, but since the network has not changed yet, the retrieval error is the same as the recoding error and is not considered separately.

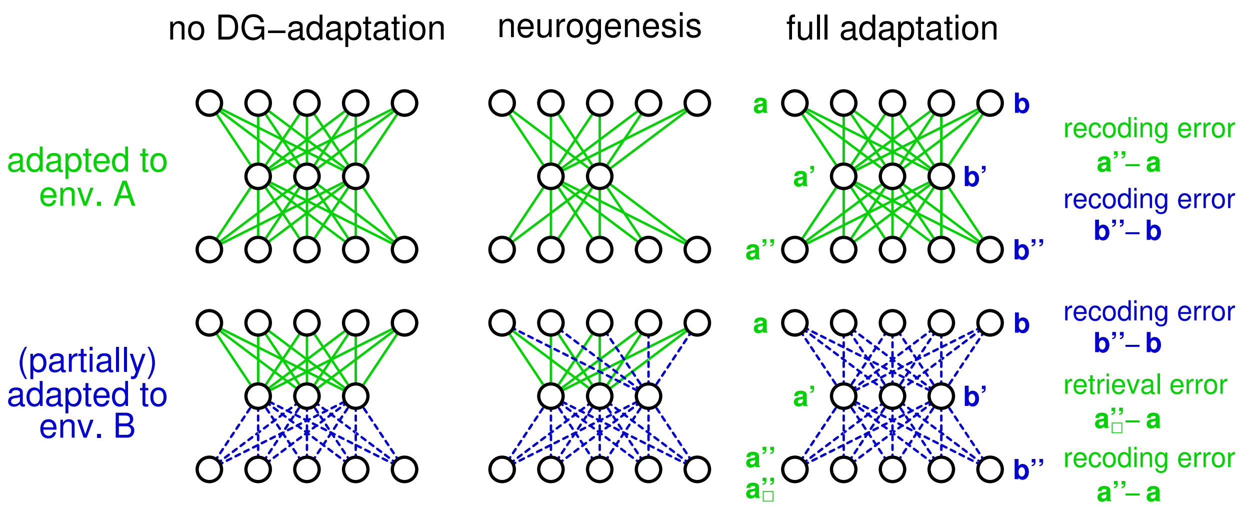



**Figure 6: Three different adaptation strategies.** Assume the animal first lives in environment $A$ (top row) and then moves to environment $B$ (bottom row). In either case we assume that the output layer, i.e. CA1/subiculum/entorhinal cortex, fully adapts to the new environment. For the hidden layer, i.e. the dentate gyrus, consider three different adaptation strategies (connections adapted to environment $A$ and $B$ are drawn as solid (green) and dashed (blue) lines, respectively). **Left (no DG-adaptation):** Once the hidden layer has adapted to an environment ($A$) its weights are fixed and no adaption to a new environment occurs. This corresponds to a transition from $n \triangleright (m+l)_A \triangleright n_A$ to $n \triangleright (m+l)_A \triangleright n_B$. **Middle (neurogenesis):** Those units in the hidden layer that have adapted to an environment ($A$) are fixed, but new units are added that can adapt to the new environment ($B$). This corresponds to a transition from $n \triangleright m_A \triangleright n_A$ to $n \triangleright (m_A + l_B) \triangleright n_B$. **Right (full adaptation):** The hidden layer always fully adapts to the current environment ($A$ or $B$). This corresponds to a transition from $n \triangleright (m+l)_A \triangleright n_A$ to $n \triangleright (m+l)_B \triangleright n_B$. On the very right the five different errors are indicated that are of interest in our scenarios.

(ii) **Recoding error B:** When the animal has just moved to environment $B$ it did not have the time yet to adapt to the new environment but ideally the recoding error for patterns $\mathbf{b}$, i.e. $\langle|\mathbf{b} - \mathbf{b}''|^2\rangle$, should already be small right from the beginning. However, we will find that this is generally not the case in our simulations.

(B) **Network adapted to environment B.**

(iii) **Recoding error B:** Of course, performance improves with adaptation, so we are also interested in the recoding error for patterns $\mathbf{b}$, i.e. $\langle|\mathbf{b} - \mathbf{b}''|^2\rangle$, after the network has adapted to environment $B$.

(iv) **Retrieval error A:** Next we ask how well the animal can retrieve memories that it has stored earlier in environment $A$. This is the retrieval error $\langle|\mathbf{a} - \mathbf{a}''_{\square}|^2\rangle$.

(v) **Recoding error A:** Finally we want to look at the performance of the animal when it returns to environment $A$ but does not have the time to newly adapt to it. This is measured by the recoding error $\langle|\mathbf{a} - \mathbf{a}''|^2\rangle$.

Note that in (i) and (v) as well as in (ii) and (iii) the same recoding errors are computed, but that the values will differ because the network has changed due to adaptation.

Table 1 lists the five different errors for the three different adaptation strategies described above. As one would expect, if a network is fully adapted to environment $A$, recoding error $A$ (i) is small and recoding error $B$ (ii) is large. Furthermore, performance is obviously better if the network has more units in the hidden layer; compare (a, c) with (b). In fact the error would go down to zero if the network had as many hidden units as input or output units, because then no dimensions would be lost in the hidden layer.

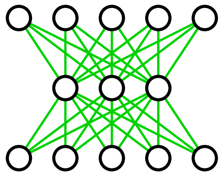
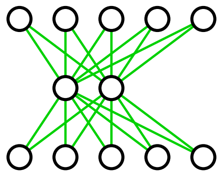
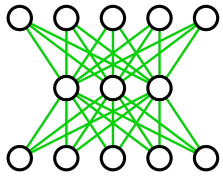
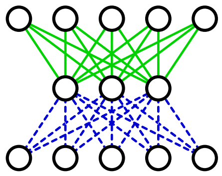
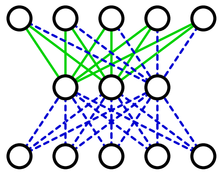
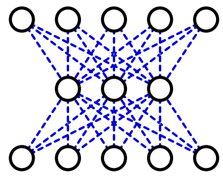
| | | (a) no DG-adaptation | (b) neurogenesis | (c) full adaptation |
|---|---|---|---|---|
| (A) | network adapted to environment $A$ |  $n \triangleright (m+l)_A \triangleright n_A$ |  $n \triangleright m_A \triangleright n_A$ |  $n \triangleright (m+l)_A \triangleright n_A$ |
| (i) | recoding error $A$ $\langle\lvert\mathbf{a}-\mathbf{a}''\rvert^2\rangle$ | 0.30 | 0.33 | 0.30 |
| (ii) | recoding error $B$ $\langle\lvert\mathbf{b}-\mathbf{b}''\rvert^2\rangle$ | 0.67 | 0.75 | 0.67 |
| (B) | network adapted to environment $B$ |  $n \triangleright (m+l)_A \triangleright n_B$ |  $n \triangleright (m_A+l_B) \triangleright n_B$ |  $n \triangleright (m+l)_B \triangleright n_B$ |
| (iii) | recoding error $B$ $\langle\lvert\mathbf{b}-\mathbf{b}''\rvert^2\rangle$ | 0.44 | 0.36 | 0.30 |
| (iv) | retrieval error $A$ $\langle\lvert\mathbf{a}-\mathbf{a}''_{\square}\rvert^2\rangle$ | 0.61 | 0.45 | 1.69 |
| (v) | recoding error $A$ $\langle\lvert\mathbf{a}-\mathbf{a}''\rvert^2\rangle$ | 0.61 | 0.41 | 0.67 |

**Table 1: Five different errors (i–v) for three different adaptation strategies (a–c)**; see text for details. An error value of 0 would indicate no error, i.e. perfect recoding or retrieval, and a value of 1 would be obtained if a network would always return a zero vector, i.e. if it would not do anything. Thus a value greater 1 indicates that the network produces misleading outputs. Each error is an average over 5000 simulation runs with randomly rotated data distributions; standard deviations were between 0 and 0.05. Rows (iv, v) show that neurogenesis (b) indeed avoids the problem of catastrophic interference with still reasonable performance in cases (i, iii). None of the networks performs well in case (ii).

With adaptation to environment $B$, the performance on patters $\mathbf{b}$ improves, of course; compare recoding error $B$ before (ii) and after (iii) adaptation. The more complete the adaptation the better the performance, thus no adaptation (a) is worst and full adaptation (c) is best; neurogenesis (b) lies in between. In the no-adaptation case (a) the encoding learned for environment $A$ is actually random with respect to the distribution of patterns $\mathbf{b}$, because the distributions in the two environments are randomly rotated relative to each other. Thus this case (a-iii) corresponds to the situation of random encoding and optimal decoding illustrated in Figure 3.

So far the results were as one would expect. The interesting question now is how the networks adapted to environment $B$ perform on patterns $\mathbf{a}$; see (iv, v). With no DG-adaptation (a) performance degrades significantly due to catastrophic interference within the decoding part of the network; compare (iv, v) with (i). Recoding and retrieval error are identical, because the encoding part of the network has not changed. Note that in this case decoding changes even though encoding is fixed, because the pattern statistics to optimize for changes. Thus stabilizing the encoding does not necessarily stabilize the optimal decoding as was argued in Section 5. But see below and Figure 7 for why the argument basically holds in the neurogenesis case.

With full adaptation (c), the effect of catastrophic interference is even more severe. The recoding error $A$ (v) increases to the level of the recoding error $B$ (ii) of the network adapted to environment $A$, which is clear for symmetry reasons. The retrieval error $A$ even increases beyond the value of 1, indicating that the reconstructed output is not only not helpful but even misleading. The reason for this is that the representation of the hidden layer has changed completely, so that the stored patterns $\mathbf{a}'_\square$ cannot be interpreted anymore. The network effectively hallucinates random output patterns, which leads to the large retrieval error. This is equivalent to the case of optimal encoding and random decoding discussed in Figure 4.

With neurogenesis (b), the network can largely avoid the effect of catastrophic interference and achieves a performance that is better than that of the two other strategies.

Figure 7 illustrates catastrophic interference and how neurogenesis can avoid it. If the final number of units is limited, the idea is that instead of wasting the units early on to represent also the low-variance directions of distribution $A$, one starts with fewer units and adds new units later to represent high-variance directions of distribution $B$ without the effect of catastrophic interference.

# 7 Discussion

## 7.1 Hypothesis and model

In this paper we have addressed the question of what the functional role of adult neurogenesis in the dentate gyrus is. As a basis for our discussion we have adopted one of the standard models of hippocampal function (Sec. 4) based on concepts from neural network theory (Sec. 3). In this model the dentate gyrus serves as an encoder, the CA3-region as a memory, and the CA1/subiculum-system as a decoder. We assume that the dentate gyrus, like most neural systems, suffers from the problem of catastrophic interference when adapting to new

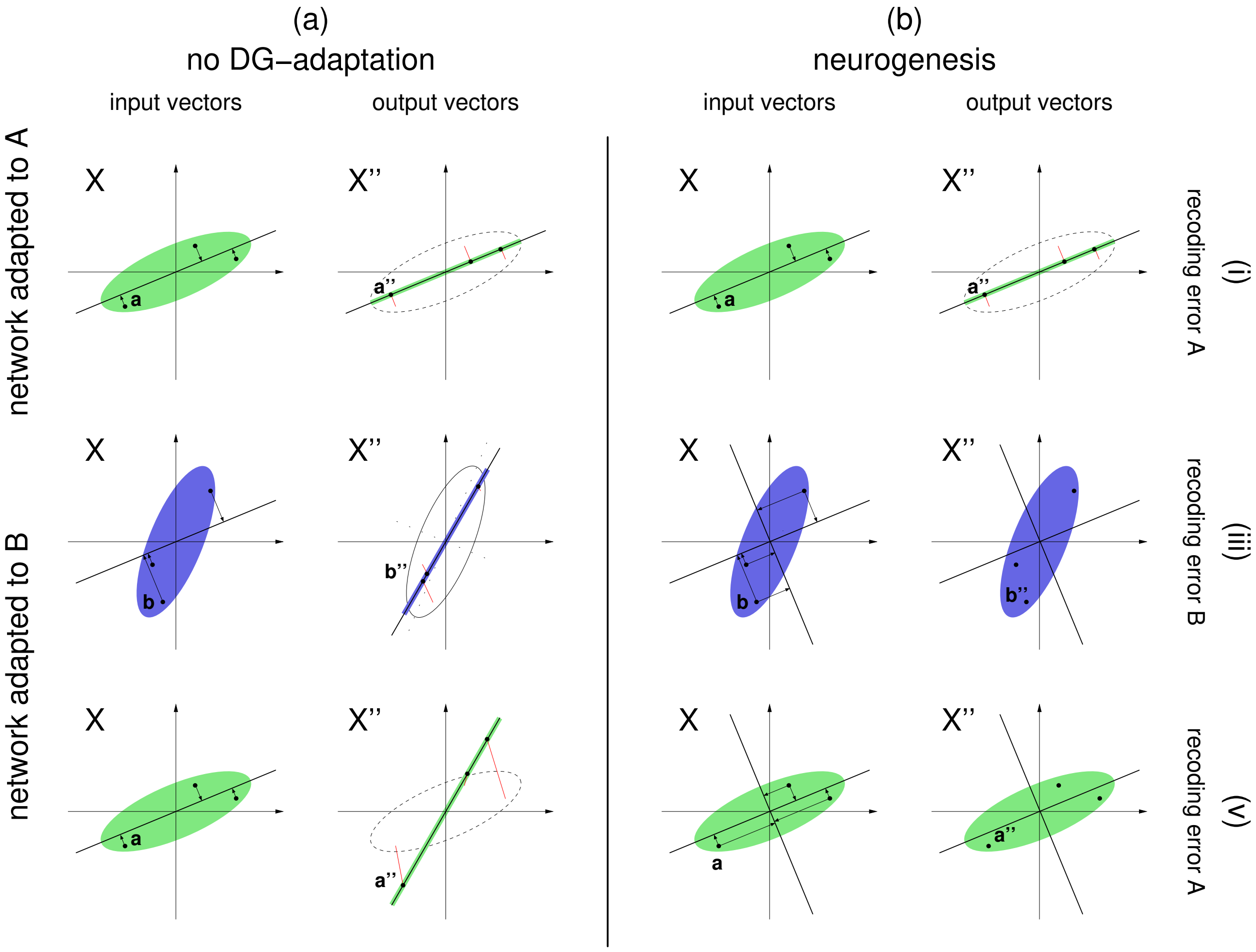



**Figure 7: Effect of catastrophic interference in the no-DG-adaptation scenario and how the problem is avoided by neurogenesis.** See next page for figure caption.

**Figure 7: Effect of catastrophic interference in the no-DG-adaptation scenario and how the problem is avoided by neurogenesis.** In contrast to previous figures, the 2D-plots represent a suitable two-dimensional projection of a high-dimensional space of input (and output) vectors. Otherwise, the figure uses the same conventions as Figure 2. In the top row the two networks are adapted to environment $A$ and recoding error $A$ is considered; in the middle and bottom row they are adapted to $B$ and the recoding errors $A$ and $B$, respectively, are considered. **No DG-adaptation, Panel (a-i):** In the network adapted to environment $A$ the encoding as well as the decoding is optimal for input vectors $\mathbf{a}$ taken from distribution $A$. The recoding error $A$ is correspondingly low (right). **Panel (a-iii):** If the encoding is fixed and no new units can be added (no DG-adaptation), the network can only optimize the decoding to environment $B$ (right). This corresponds to the case of random encoding and optimal decoding already illustrated in Figure 3. The network reduces the recoding error by misusing the encoding dimension (left) optimized for $A$ to represent the high-variance direction of distribution $B$ (right). **Panel (a-v):** Now that the decoding is optimized for $B$ the decoding of vectors $\mathbf{a}$ is poor (right) and the recoding error $A$ correspondingly large. **Neurogenesis, Panel (b-i):** In these 2D-plots the initial network in the neurogenesis scenario looks identical to that in the no-DG-adaptation scenario. However, note that the number of hidden units is smaller, so that there are some dimensions (with low variance in distribution $A$) not shown here that are less well represented than in the no-DG-adaptation scenario. **Panel (b-iii):** The old hidden units are fixed, like in the no-DG-adaptation scenario, but the network can add new hidden units (only one is shown here), which adapt to environment $B$. Thus the encoding gets extended by dimensions that are not yet represented but have high variance in distribution $B$ (left). This reduces (in this figure eliminates) the misuse of encoding dimensions optimized for distribution $A$ (right) and also improves the recoding of vectors $\mathbf{b}$. **Panel (b-v):** The good decoding of vectors $\mathbf{a}$ is largely preserved (right). In this figure the recoding actually looks perfect, since we now have two hidden units representing the two plotted dimensions. Other dimensions (with lower variance) not plotted still suffer from the misuse of encoding dimensions and contribute to the recoding error, but to a much lesser degree than in the no-DG-adaptation scenario.

environments. Our basic hypothesis for the role of adult neurogenesis says that new neurons help the dentate gyrus avoiding the problem of catastrophic interference by keeping the old neurons, which are adapted to earlier environments, fixed and adding new neurons, which are more plastic and can code for those aspects that are qualitatively new in the current environment.

We have used a linear auto-encoder network-model to illustrate and quantify the advantage new neurons might offer according to our hypothesis. This model is very simple and obviously neglects many prominent features of the biological system. Nevertheless, it demonstrates and explains several aspects of our hypothesis and yields some non-trivial insights.

Let us summarize some of the key features of our hypothesis and the network model.

- Our hypothesis requires that old neurons are relatively stable while new neurons are plastic. This is consistent with physiological findings (see Sec. 2.4).
- The results summarized in Table 1 show that adult neurogenesis can indeed reduce the effect of catastrophic interference significantly; see rows (iv) and (v) and compare the neurogenesis strategy (b) with the other two (a, c).
- The results of the no-adaptation strategy show that even if the encoding optimized for environment $A$ is kept fixed does the performance on patterns $\mathbf{a}$ degrade drastically if the decoding is optimized for environment $B$. This is a somewhat surprising result and a consequence of the misuse of high-variance $A$-dimensions for representing $\mathbf{b}$-patterns, as illustrated in Figure 7. This misuse is reduced significantly by adding new neurons.
- It is clear theoretically and can also be observed in the simulation results that in the neurogenesis scenario in order to achieve a small recoding error a large number of new hidden units is necessary initially, but that fewer new units suffice later. Compare entry (b-i) with entry (b-iii) in Table 1; the former required 15 new units, the latter only 5 new units, but both errors are relatively small. Thus the effect of neurogenesis accumulates and with increasing age the need for new units decreases, which is consistent with neuroanatomical observations (see Sec. 2.4).
- According to our hypothesis CA1 always fully adapts to optimize the decoding for the current environment. We have made this assumption mainly because we did not see a feasible way of coordinating learning in dentate gyrus and CA1 in a more specific way. However, the simulations at least partially justify this assumption because in the neurogenesis case performance is good with full CA1 adaptation, see Table 1. The strong plasticity observed in CA1 physiologically also seems to be consistent with our assumption (see Sec. 2.3).
- According to our hypothesis new neurons are only added when the pattern statistics changes, i.e. it depends on the environment. However, the generation of new neurons takes too much time to initiate neurogenesis exactly when needed. Instead there must always be new neurons available that could be used quickly. This is consistent with the fact that most new neurons are not integrated but die after some time.

## 7.2 Comparison with other models

There are a number of artificial neural networks that employ new units (artificial neurons) as part of their learning strategy. A prominent example in the field of vector quantization is the adaptive resonance theory (ART) (Carpenter & Grossberg, 1987; Carpenter et al., 1991). In the ART-network new units are added as new patterns have to be learned that cannot be well represented by the existing units. This is similar in spirit to the model we propose, but the representation is very different. In the ART-network a pattern is represented by a single unit while in our model a pattern is represented with a population code.

Self-organizing maps (SOM) differ from vector quantization networks in that also the topological neighborhood structure of the units is being represented. SOM-networks that employ new units have been extensively studied by Fritzke (1994). In his growing cell structures new units can be added to represent new patterns, much like in the ART-network, or to refine the representation. Again, patterns are represented by single units and not populations of units.

Vector quantization and self-organizing maps are examples of unsupervised learning tasks. For feedforward networks supervised learning algorithms are of particular interest. Also for this class of networks do algorithms exist that successively add new units to improve the performance (see Bishop, 1995, sec. 9.5, for an overview). The model most closely related to ours is probably the cascade-correlation learning architecture (Fahlman & Lebiere, 1990). This architecture is a feedforward network with initially no hidden unit. After having trained this reduced network, a single hidden unit is added and the whole network trained again with only new connections of this hidden unit being modified and all old units and connections kept fixed. Then the first hidden unit is kept fixed, too, and a second hidden unit is added and trained. This way the network is extended by single hidden units and retrained until network performance is satisfactory. The resulting network has an unusual structure in that each hidden unit forms its own hidden layer and most connections skip a number of hidden layers. Like our model this network learns a distributed representation. The main difference is that in the cascade-correlation network new units are added to improve the performance on identical training data, while we have considered the problem of extending a representation to new patterns without catastrophic interference, a problem not addressed at all by Fahlman & Lebiere (1990).

Cecchi et al. (2001) have proposed a biologically motivated model for neurogenesis in the olfactory bulb. They used laterally connected inhibitory units to perform an orthogonalization of odor representations. Neurogenesis was assumed to occur at a constant rate while survival of an inhibitory unit was taken to depend on activity and was highest for units establishing an inhibitory interaction between correlated units. This way correlations between output units were reduced and the network converged to a 1-of-N code. This is an interesting model, although neurogenesis does not seem to be really essential, since the same effect could probably have been achieved with a fixed number of inhibitory units with adapting synaptic weights, like in other models that perform pattern orthogonalization.

Deisseroth et al. (2004) have developed a model for neurogenesis in the hippocampus. Like ours it is a three layer network with the hidden layer showing neurogenesis. However, it is nonlinear and trained to learn hetero-associations (input and output patterns are different). The authors make the assumption that there is a turnover of units in the hidden layer, i.e.

some units die and are replaced by the same number of new units. Such a network tends to forget old memories more quickly but is able to store new memories without the effect of catastrophic interference. Insofar the model offers an interesting explanation for the role of adult neurogenesis. It would be interesting to see, however, whether neurogenesis is really necessary. The same effect might be more efficiently achieved by some kind of weight decay as can be done in Hopfield networks (Howard et al., 1987, cited in Amit, 1989). We also believe that the assumption of a neural turnover in dentate gyrus is not supported by experimental data but that neurogenesis is rather an accumulative effect (see Sec. 2.4).

## 7.3 Future perspectives

The model presented here is obviously only a very first step. It illustrates our hypothesis about the functional role of adult neurogenesis and has the advantage of being simple, which provides insights that might not be so easily accessible with a more complex model. However, it neglects a number of important aspects of the hippocampal formation and needs to be extended in several ways. Let us list only a few important extensions:

- It is entirely unclear what the optimal encoding is the dentate gyrus has to learn. Compression might be an important part of it, but it is definitely not all. A plausible additional objective might be sparsification to achieve orthogonalization. This requires a nonlinear network and will lead to a greater number of units in the dentate-gyrus layer, more consistent with the neuroanatomy (Sec. 2.1).
- CA3 not only receives input via the dentate gyrus but also directly from entorhinal cortex. It is an open question how these two very different pathways complement each other. One hypothesis is that they serve different modes of operation such as write-in and read-out (Treves & Rolls, 1992); another hypothesis states that they help learning to predict events (Borisyuk et al., 1999). We will have to reconsider such hypotheses with respect to neurogenesis and extend the model correspondingly.
- The latter will also require that we model CA3 more explicitly.
- Another interesting direction is to consider how neurogenesis is regulated and how this depends on the functional requirements.

Finally, we would like to emphasize that the simple model presented here is mainly illustrative. Our hypothesis that adult neurogenesis helps the dentate gyrus to avoid the problem of catastrophic interference is much more general than that and could be realized with very different network models.

# Acknowledgments

This work has been supported by grants from the Volkswagen Foundation to LW and GK.

# A Mathematical model

## A.1 Definition of the model

The model considered in this paper is a linear autoencoder network with three layers (Fig. 1). In- and output layer have $n$ units; the hidden layer has $m \leq n$ units. Patterns presented to the network are indicated by $n$-dimensional vectors $\mathbf{x}$. The resulting $m$-dimensional vectors of activity in the hidden layer and the $n$-dimensional vectors in the output-layer are denoted by $\mathbf{x}'$ and $\mathbf{x}''$, respectively. The linear encoding from the input to the hidden layer can be written as an $m \times n$ matrix $\mathbf{E}$ and the decoding can be written as an $n \times m$ matrix $\mathbf{D}$. Thus, the activities in the hidden and the output-layer are

$$\begin{aligned} \mathbf{x}' &:= \mathbf{E}\mathbf{x}\,, &(3)\\ \mathbf{x}'' &:= \mathbf{D}\mathbf{x}' = \mathbf{D}\mathbf{E}\mathbf{x} = \mathbf{R}\mathbf{x}\,, &(4)\\ \text{with}\quad \mathbf{R} &:= \mathbf{D}\mathbf{E}\,. &(5) \end{aligned}$$

The recoding matrix $\mathbf{R}$ is an $n \times n$ matrix and has been defined for brevity. The goal of the network is optimal compression in the least-squares sense, i.e. the output patterns should be as similar to the input patterns as possible as measured by the mean squared Euclidean distance

$$\begin{aligned} E &:= \left\langle |\mathbf{x} - \mathbf{x}''|^2 \right\rangle &(6)\\ &\stackrel{(4)}{=} \left\langle (\mathbf{x} - \mathbf{D}\mathbf{E}\mathbf{x})^T (\mathbf{x} - \mathbf{D}\mathbf{E}\mathbf{x}) \right\rangle\,, &(7) \end{aligned}$$

see equations (1) and (2).

The input patterns $\mathbf{x}$ presented to the network are drawn from an $n$-dimensional multivariate Gaussian distribution (probability density function) with zero mean, completely characterized by its covariance matrix

$$\mathbf{C} := \langle \mathbf{x}\mathbf{x}^T \rangle\,, \tag{8}$$

with $\langle \cdot \rangle$ indicating averaging over all input patterns. The eigenvalues and eigenvectors of the covariance matrix, ordered by decreasing eigenvalue, will be indicated by $\lambda_i$ and $\mathbf{c}_i$, respectively, with $i \in 1, ..., n$. Eigenvectors are normalized to norm 1. In our simulations the eigenvalues are chosen such that the total variance is 1. The first 15 eigenvalues follow an exponential decay, together explaining 2/3 of the total variance, while the remaining ones have a constant value of 1/135, summing up to the missing 1/3. The first 15 dimensions are meant to carry the relevant information while the remaining ones are interpreted as noise (cf. Fig. 5).

## A.2 Optimal autoencoder

From principal component analysis (PCA) it is known that an optimal linear encoder-decoder pair can be obtained by choosing the rows of $\mathbf{E}$ and the columns of $\mathbf{D}$ to be the first $m$

eigenvectors (having largest eigenvalues) of the covariance matrix, i.e.

$$\begin{aligned} & \mathbf{E}^\star &:=& \ (\mathbf{c}_1, ..., \mathbf{c}_m)^T & (9)\\ \text{and} \quad & \mathbf{D}^\star &:=& \ (\mathbf{c}_1, ..., \mathbf{c}_m) & (10)\\ \Longrightarrow \quad & \mathbf{R}^\star &=& \ \mathbf{D}^\star \mathbf{E}^\star & (11)\\ & &=& \ \sum_{i=1}^{m} \mathbf{c}_i \mathbf{c}_i^T \,. & (12) \end{aligned}$$

$\mathbf{R}^\star$ projects the input patterns into the subspace spanned by the first $m$ eigenvectors.

Equations (9, 10) were used to determine the weights of those networks in Table 1 that are fully adapted, i.e. networks (A-a, b, c) and (B-c). Also the encoding weights of network (B-a) and of the old units in network (B-b) are now known, since they are fixed and taken from the networks adapted to environment $A$.

For an optimal autoencoder the preserved variance of the data is just the sum over the first $m$ eigenvalues and the error made is the sum over the remaining $(n - m)$ eigenvalues

$$E(\mathbf{R}^\star) = \sum_{i=m+1}^{n} \lambda_i \,. \tag{13}$$

Equation (13) yields the recoding errors for the networks fully adapted to a particular environment. With the known eigenvalues (Sec. A.1) we find that with 15 hidden units $E = 1 - 2/3 = 1/3 \approx 0.33$ and with 20 hidden units $E = 1 - 2/3 - 5/135 \approx 0.30$, see row (i) and entry (c-iii) of Table 1.

## A.3 Neurogenesis

Assume that $m$ hidden units have been optimized for environment $A$ according to (9, 10). How should new units adapt in the neurogenesis case after the animal has moved to environment $B$?

We did not derive a closed form expression for the optimal encoding-weights of the new units. Instead, in analogy to what we know about the optimal autoencoder, we have chosen the encoding weight-vectors of the new units (i) to be orthogonal to the weight vectors of the old units and (ii) to span the space of greatest remaining variance of the pattern distribution of environment $B$. In mathematical terms, if $\mathbf{E}^\star$ and $\mathbf{D}^\star$ are the optimal en- and decoders of the $m$ old units in environment $A$, then $(\mathbf{I} - \mathbf{D}^\star\mathbf{E}^\star)\mathbf{b}$ is the component of patterns $\mathbf{b}$ from the new environment orthogonal to the space spanned by the weight vectors of the old units. The first $l$ eigenvectors of the corresponding covariance matrix $\langle(\mathbf{I} - \mathbf{D}^\star\mathbf{E}^\star)\mathbf{b}\mathbf{b}^T(\mathbf{I} - \mathbf{D}^\star\mathbf{E}^\star)^T\rangle$ are taken as the weight vectors of the $l$ newly added units. This is the way we have determined the encoding weights of newly added units in network (B-b).

In Table 2 we also considered the case where the encoding weight-vectors where not constrained to be orthogonal to the weight vectors of the old units. In that case they were simply chosen to be the first eigenvectors of the covariance matrix of patterns $\mathbf{b}$. This strategy is indicated with the notation $l_{(B\angle A)}$ and the standard neurogenesis-strategy with the notation $l_{(B\perp A)}$. In Table 1 this distinction is not made, but only the standard strategy is considered.

## A.4 Optimal decoder given an arbitrary encoder

Let us now try to determine an optimal decoder $\mathbf{D}^\star$ given an arbitrary encoder $\mathbf{E}$. Without loss of generality we assume that $\mathbf{E}$ has highest rank, because otherwise some of the hidden units would be redundant and could be discarded without loss of information (although with loss of robustness). After discarding the redundant units $\mathbf{E}$ would have highest rank again and the following considerations would apply. We also assume that the covariance matrix $\mathbf{C}$ has full rank, again without loss of generality, because if it did not have full rank some dimensions would have zero variance in the data and could be discarded without loss of information. The remaining covariance matrix would have full rank. Full rank implies that $\mathbf{C}$ is invertible.

Since $\mathbf{C}$ and $\mathbf{E}$ are fixed, consider the error only as a function of the decoder, i.e. $E = E(\mathbf{D})$. By definition

$$\mathbf{D}^\star := \min_{\mathbf{D}\in\mathbb{R}^{(n,m)}} E(\mathbf{D}) \tag{14}$$

$$\overset{(7,3)}{=} \min_{\mathbf{D}\in\mathbb{R}^{(n,m)}} \left\langle (\mathbf{x}-\mathbf{D}\mathbf{x}')^T(\mathbf{x}-\mathbf{D}\mathbf{x}') \right\rangle . \tag{15}$$

A necessary condition for such a minimum is that the derivative of the error with respect to the decoder matrix vanishes, i.e.

$$\mathbf{0} \overset{!}{=} \left.\frac{\partial E(\mathbf{D})}{\partial \mathbf{D}}\right|_{\mathbf{D}^\star} \tag{16}$$

$$\overset{(7,3)}{=} \left.\frac{\partial}{\partial \mathbf{D}} \left\langle (\mathbf{x}-\mathbf{D}\mathbf{x}')^T(\mathbf{x}-\mathbf{D}\mathbf{x}') \right\rangle\right|_{\mathbf{D}^\star} \tag{17}$$

$$= \left\langle \left.\frac{\partial}{\partial \mathbf{D}} (\mathbf{x}-\mathbf{D}\mathbf{x}')^T(\mathbf{x}-\mathbf{D}\mathbf{x}')\right|_{\mathbf{D}^\star} \right\rangle \tag{18}$$

$$= \left\langle \left.\frac{\partial}{\partial \mathbf{D}} \left(\mathbf{x}^T\mathbf{x} - 2\,\mathbf{x}^T\mathbf{D}\mathbf{x}' - \mathbf{x}'^T\mathbf{D}^T\mathbf{D}\mathbf{x}'\right)\right|_{\mathbf{D}^\star} \right\rangle \tag{19}$$

(since $\mathbf{x}'^T\mathbf{D}^T\mathbf{x}$ is a scalar and thus equal to its transpose $\mathbf{x}^T\mathbf{D}\mathbf{x}'$)

$$= \left\langle -2\,\mathbf{x}\mathbf{x}'^T + 2\,\mathbf{D}^\star\mathbf{x}'\mathbf{x}'^T \right\rangle \tag{20}$$

(according to the rules of matrix derivatives)

$$\overset{(3)}{=} -2\left\langle \mathbf{x}\mathbf{x}^T\right\rangle \mathbf{E}^T + 2\,\mathbf{D}^\star\mathbf{E}\left\langle \mathbf{x}\mathbf{x}^T\right\rangle \mathbf{E}^T \tag{21}$$

$$\overset{(8)}{=} -2\,\mathbf{C}\mathbf{E}^T + 2\,\mathbf{D}^\star\mathbf{E}\mathbf{C}\mathbf{E}^T \tag{22}$$

$$\iff \mathbf{D}^\star\mathbf{E}\mathbf{C}\mathbf{E}^T = \mathbf{C}\mathbf{E}^T \tag{23}$$

$$\iff \mathbf{D}^\star = \mathbf{C}\mathbf{E}^T\left(\mathbf{E}\mathbf{C}\mathbf{E}^T\right)^{-1} \tag{24}$$

($\mathbf{E}\mathbf{C}\mathbf{E}^T$ is invertible because $\mathbf{C}$ and $\mathbf{E}$ have highest rank)

Condition (16) is necessary but not sufficient for a minimum. However, since the error function (7) is quadratic in the coefficients of $\mathbf{D}$, bounded below by zero but not bounded above, and $\mathbf{D}^\star$ is unique, is must yield a minimum (see (Rasch, 2003) for a more detailed proof).

With (24) we have arrived at an expression that gives us the optimal decoder in a closed form given a data distribution with covariance matrix $\mathbf{C}$ and an arbitrary encoder matrix $\mathbf{E}$, both matrices having highest rank. This determines the decoding weights of the networks (B-a) and (B-b).

## A.5 Recoding error

Now consider the recoding error for an arbitrary en- and decoder. From the definition of the recoding error we can derive

$$\begin{aligned}
E(\mathbf{R}) &\overset{(7,5)}{=} \left\langle (\mathbf{x} - \mathbf{R}\mathbf{x})^T (\mathbf{x} - \mathbf{R}\mathbf{x}) \right\rangle && (25)\\
&= \left\langle \mathbf{x}^T (\mathbf{I} - \mathbf{R})^T (\mathbf{I} - \mathbf{R})\mathbf{x} \right\rangle && (26)\\
&= \left\langle \mathrm{tr}\left(\mathbf{x}^T (\mathbf{I} - \mathbf{R})^T (\mathbf{I} - \mathbf{R})\mathbf{x}\right) \right\rangle && (27)\\
&\qquad \text{(since the argument is a scalar)}\\
&= \left\langle \mathrm{tr}\left((\mathbf{I} - \mathbf{R})\mathbf{x}\mathbf{x}^T (\mathbf{I} - \mathbf{R})^T\right) \right\rangle && (28)\\
&\qquad \text{(since } \mathrm{tr}\,(\mathbf{AB}) = \mathrm{tr}\,(\mathbf{BA}) \text{ for valid matrices } \mathbf{A} \text{ and } \mathbf{B})\\
&= \mathrm{tr}\left((\mathbf{I} - \mathbf{R}) \left\langle \mathbf{x}\mathbf{x}^T \right\rangle (\mathbf{I} - \mathbf{R})^T\right) && (29)\\
&\qquad \text{(since } \mathrm{tr}\,(\cdot) \text{ and } \langle\cdot\rangle \text{ commute)}\\
&= \mathrm{tr}\left((\mathbf{I} - \mathbf{R})\mathbf{C}(\mathbf{I} - \mathbf{R})^T\right) . && (30)
\end{aligned}$$

This expression permits us to compute the recoding error directly from the covariance matrix $\mathbf{C}$ and the en- and decoder matrices $\mathbf{E}$ and $\mathbf{D}$ without the need to sample over the input patterns. This simplifies the simulations significantly. The retrieval error can also be computed this way by using the encoder before adaptation and the decoder after adaptation.

Equation (30) was used to compute all the errors in Table 1. If the relation between distribution $A$ and $B$ mattered, like in rows (ii, iii, iv, v) except for entry (c-iii), we averaged over 5000 randomly drawn distributions for $B$ keeping distribution $A$ fixed.

It is of theoretical interest that for the optimal decoder $\mathbf{D}^\star$ equation (30) simplifies further like

$$\begin{aligned}
E(\mathbf{D}^\star, \mathbf{E}) &\overset{(30,5)}{=} \mathrm{tr}\left((\mathbf{I} - \mathbf{D}^\star\mathbf{E})\mathbf{C}(\mathbf{I} - \mathbf{D}^\star\mathbf{E})^T\right) && (31)\\
&= \mathrm{tr}\left((\mathbf{I} - \mathbf{D}^\star\mathbf{E})\mathbf{C}(\mathbf{I} - \mathbf{E}^T\mathbf{D}^{\star T})\right) && (32)\\
&= \mathrm{tr}\left(\mathbf{C} - \mathbf{C}\mathbf{E}^T\mathbf{D}^{\star T} - \mathbf{D}^\star\mathbf{E}\mathbf{C} + \mathbf{D}^\star\mathbf{E}\mathbf{C}\mathbf{E}^T\mathbf{D}^{\star T}\right) && (33)\\
&\overset{(24)}{=} \mathrm{tr}\left(\mathbf{C} - \mathbf{C}\mathbf{E}^T\mathbf{D}^{\star T} - \mathbf{D}^\star\mathbf{E}\mathbf{C} + \mathbf{C}\mathbf{E}^T \underbrace{(\mathbf{E}\mathbf{C}\mathbf{E}^T)^{-1}\mathbf{E}\mathbf{C}\mathbf{E}^T}_{=\mathbf{I}} \mathbf{D}^{\star T}\right) && (34)\\
&= \mathrm{tr}\left((\mathbf{I} - \mathbf{D}^\star\mathbf{E})\mathbf{C}\right) . && (35)
\end{aligned}$$

From this equation it is easy to derive equation (13) for optimal en- and decoder if one takes into account that $\mathrm{tr}\,(\mathbf{C}) = \sum_{i=1}^{n} \lambda_i$ and that the row and column vectors of $\mathbf{E}^\star$ and $\mathbf{D}^\star$, respectively, are the first $m$ eigenvectors of covariance matrix $\mathbf{C}$, resulting in $\mathrm{tr}\left(\mathbf{E}^{\star}\mathbf{C}\mathbf{D}^{\star}\right) = \sum_{i=1}^{m} \lambda_i$.

# B Control experiments

Table 2 shows results from a number of control experiments and includes also the results of Table 1. It is divided into three blocks depending on how many hidden units were used in environment $A$ and $B$. Within each block the results are ordered by the degree to which the network adapts to environment $B$.

The table has a number of regularities and symmetries. Columns (A), (i), and (ii) have identical values for all experiments within the first and the second block, and for all experiments within the third block. The networks adapted to environment $B$ (B) are the same in the third block as in the second block. Therefore, since the recoding errors depend only on the current network, the results of columns (iii) and (v) in the third block are identical to those in the second block. If the roles of $A$ and $B$ are exchanged completely, the results are identical for symmetry reasons; compare, for instance, (11-i) with (14-iii), (11-ii) with (14-v), (31-i) with (34-iii), and (31-ii) with (34-v). If the encoding part of the network does not change between storage and retrieval of patterns $\mathbf{a}$, then retrieval error $A$ is identical to recoding error $A$; compare (11-iv) with (11-v) and (31-iv) with (31-v).

In general performance improves as more units are used in the hidden layer; compare experiment (11) with (21) and (31) and experiment (14) with (24) and (34). One exception to this rule is retrieval error $A$, which increases from (24) to (34). In this case, the representation in the hidden layer changes completely and the retrieved patterns are effectively hallucinated random patterns leading to errors greater than one (cf. Sec. 6.2). The more hidden units the greater the error.

From the order within each block one would expect that recoding error $B$ (iii) decreases and retrieval and recoding error $A$ (iv, v) increase. While the former is true, the latter does actually not hold. Recoding error $A$, for instance, decreases from (11) to (14) even though the network adapts more to environment $B$. This is due to the fact that for (11) the effect of misuse of dimensions illustrated in Figure 7 is particularly strong, since there are only few units in the hidden layer. With more hidden units this effect is less severe; compare the decrease from (11) to (14) with the increase from (31) to (34). The effect of misuse of dimensions in column (v) is greatly reduced by neurogenesis; compare experiments (22, 23) with (21, 24) and experiments (32, 33) with (31, 34). For retrieval error $A$ (iv) the error depends a lot on whether the new units have encoding weight-vectors that are orthogonal to the old ones or not. Neurogenesis with orthogonal weight-vectors performs clearly better; compare experiment (22) with (23) and experiment (32) with (33).

# References

ABRAHAM, W. C. AND WILLIAMS, J. M. (2003). Properties and mechanisms of LTP maintenance. *Neuroscientist*, 9(6):463–474.

| | (A) network adapted to environment $A$ | (i) recoding error $A$ $\langle\lvert\mathbf{a}-\mathbf{a}''\rvert^2\rangle$ | (ii) recoding error $B$ $\langle\lvert\mathbf{b}-\mathbf{b}''\rvert^2\rangle$ | (B) network adapted to environment $B$ | (iii) recoding error $B$ $\langle\lvert\mathbf{b}-\mathbf{b}''\rvert^2\rangle$ | (iv) retrieval error $A$ $\langle\lvert\mathbf{a}-\mathbf{a}''_{\square}\rvert^2\rangle$ | (v) recoding error $A$ $\langle\lvert\mathbf{a}-\mathbf{a}''\rvert^2\rangle$ |
|---|---|---|---|---|---|---|---|
| (11) | $n\triangleright \; m_A \; \triangleright n_A$ | 0.33 0.00 | 0.75 0.01 | $n\triangleright \; m_A \; \triangleright n_B$ | 0.53 0.02 | 0.77 0.05 | 0.77 0.05 |
| (14) | ” | ” | ” | $n\triangleright \; m_B \; \triangleright n_B$ | 0.33 0.00 | 1.65 0.05 | 0.75 0.01 |
| (21) | ” | ” | ” | $n\triangleright \; (m+l)_A \; \triangleright n_B$ | 0.44 0.01 | 0.63 0.04 | 0.61 0.04 |
| (b) (22) | ” | ” | ” | $n\triangleright(m_A + l_{(B\perp A)})\triangleright n_B$ | 0.36 0.01 | 0.45 0.01 | 0.41 0.01 |
| (23) | ” | ” | ” | $n\triangleright(m_A + l_{(B\angle A)})\triangleright n_B$ | 0.36 0.01 | 0.56 0.03 | 0.41 0.01 |
| (24) | ” | ” | ” | $n\triangleright \; (m+l)_B \; \triangleright n_B$ | 0.30 0.00 | 1.65 0.05 | 0.67 0.02 |
| (a) (31) | $n\triangleright(m+l)_A\triangleright n_A$ | 0.30 0.00 | 0.67 0.02 | $n\triangleright \; (m+l)_A \; \triangleright n_B$ | 0.44 0.01 | 0.61 0.04 | 0.61 0.04 |
| (32) | ” | ” | ” | $n\triangleright(m_A + l_{(B\perp A)})\triangleright n_B$ | 0.36 0.01 | 0.49 0.02 | 0.41 0.01 |
| (33) | ” | ” | ” | $n\triangleright(m_A + l_{(B\angle A)})\triangleright n_B$ | 0.36 0.01 | 0.62 0.03 | 0.41 0.01 |
| (c) (34) | ” | ” | ” | $n\triangleright \; (m+l)_B \; \triangleright n_B$ | 0.30 0.00 | 1.69 0.05 | 0.67 0.02 |

**Table 2: Recoding and retrieval errors in different scenarios.** Each number is an average over 5000 simulation runs with randomly oriented distributions for vectors **a** and **b**; standard deviations are given in small letters. Columns (A, B) and (i–v) correspond to the rows and rows (31), (22), and (34) correspond to the columns of Table 1. The latter refer to the three strategies considered in Section 6.2, namely (a) no DG-adaptation, (b) neurogenesis, and (c) full adaptation. Note that column (A) has no relevance for columns (iii) and (v) but for column (iv). The notation characterizing the networks is explained in Sections 6.1 and A.3. Row (22) corresponds to neurogenesis with fixed old neurons and adapting new neurons with weight vectors orthogonal to the old ones. This is the strategy that clearly yields best average performance on vectors **a** and **b**.

Altman, J. and Das, G. D. (1965). Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *The Journal of Comparative Neurology*, 124:319–335.

Amaral, D. G., Ishizuka, N., and Claiborne, B. J. (1990). Neurons, numbers and the hippocampal network. *Progress in Brain Research*, 83:1–11.

Amaral, D. G. and Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience*, 31(3):571–591.

Amit, D. J. (1989). *Modeling Brain Function: The World of Attractor Neural Networks.* Cambridge University Press.

Ans, B. and Rousset, S. (1997). Avoiding catastrophic forgetting by coupling two reverberating neural networks. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie*, 320(12):989–997.

Barnes, C. A., McNaughton, B. L., Mizumori, S. J., Leonard, B. W., and Lin, L. H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Progress in Brain Research*, 83:287–300.

Bentz, H. J., Hagstroem, M., and Palm, G. (1989). Information storage and effective data retrieval in sparse matrices. *Neural Networks*, 2:289–293.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford University Press.

Borisyuk, R., Denham, M., Denham, S., and Hoppensteadt, F. (1999). Computational models of predictive and memory-related functions of the hippocampus. *Reviews in the Neurosciences*, 10(3-4):213–232.

Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, 33:325–340.

Cameron, H. A. and McKay, R. D. (2001). Adult neurogenesis produces a large pool of new granule cells in the dentate gyrus. *The Journal of Comparative Neurology*, 435(4):406–417.

Carleton, A., Rochefort, C., Morante-Oria, J., Desmaisons, D., Vincent, J.-D., Gheusi, G., and Lledo, P.-M. (2002). Making scents of olfactory neurogenesis. *J. Neurophysiology - Paris*, 96:115–122.

Carpenter, G. A. and Grossberg, S. (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 26:4919–4930.

Carpenter, G. A., Grossberg, S., and Rosen, D. B. (1991). ART 2-A; an adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, 4(4):493–504.

Cecchi, G. A., Petreanu, L. T., Alvarez-Buylla, A., and Magnasco, M. O. (2001). Unsupervised learning and adaptation in a model of adult neurogenesis. *The Journal of Computational Neuroscience*, 11(2):175–182.

Chrobak, J. J. and Buzsáki, G. (1998). Operational dynamics in the hippocampal-entorhinal axis. *Neuroscience & Biobehavioral Reviews*, 22(2):303–310.

Chrobak, J. J., Lörincz, A., and Buzsáki, G. (2000). Physiological patterns in the hippocampo-entorhinal cortex system. *Hippocampus*, 10(4):457–465.

Deisseroth, K., Singla, S., Toda, H., Monje, M., Palmer, T. D., and Malenka, R. C. (2004). Excitation-neurogenesis coupling in adult neural stem/progenitor cells. *Neuron*, 42(4):535–552.

D'Hooge, R. and De Deyn, P. P. (2001). Applications of the morris water maze in the study of learning and memory. *Brain Research Reviews*, 36(1):60–90.

Dudek, S. M. and Bear, M. F. (1992). Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proceedings of the National Academy of Sciences of the United States of America*, 89(10):4363–4367.

Eriksson, P. S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A. M., Nordborg, C., Peterson, D. A., and Gage, F. H. (1998). Neurogenesis in the adult human hippocampus. *Nature Medicine*, 4(11):1313–1317.

Fahlman, S. E. and Lebiere, C. (1990). The cascade-correlation learning architecture. In Touretzky, D., editor, *Advances in Neural Information Processing Systems 2 (NIPS'1989)*, pages 524–532. Morgan-Kaufmann.

French, R. M. (1997). Pseudo-recurrent connectionist networks: An approach to the "sensitivity-stability" dilemma. *Connection Science*, 9:353–379.

Freund, T. F. and Buzsáki, G. (1996). Interneurons of the hippocampus. *Hippocampus*, 6(4):347–470.

Fritzke, B. (1994). Growing cell structures - a self-organizing network for unsupervised and supervised learning. *Neural Networks*, 7(9):1441–1460.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58.

Gerstner, W. and Kistler, W. (2002). *Spiking Neuron Models*. Cambridge University Press.

Gluck, M. A. and Myers, C. E. (2001). *Gateway to memory*. MIT Press.

Gould, E., Beylin, A., Tanapat, P., Reeves, A., and Shors, T. J. (1999). Learning enhances adult neurogenesis in the hippocampal formation. *Nature Neuroscience*, 2(3):260–265.

Gould, E. and Gross, C. G. (2002). Neurogenesis in adult mammals: some progress and problems. *The Journal of Neuroscience*, 22(3):619–623.

Gould, E., McEwen, B. S., Tanapat, P., Galea, L. A., and Fuchs, E. (1997). Neurogenesis in the dentate gyrus of the adult tree shrew is regulated by psychosocial stress and NMDA receptor activation. *The Journal of Neuroscience*, 17(7):2492–2498.

Gould, E., Tanapat, P., McEwen, B. S., Flugge, G., and Fuchs, E. (1998). Proliferation of granule cell precursors in the dentate gyrus of adult monkeys is diminished by stress. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6):3168–3171.

Gould, E., Vail, N., Wagers, M., and Gross, C. G. (2001). Adult-generated hippocampal and neocortical neurons in macaques have a transient existence. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10910–10917.

Hammond, C. (2001). *Cellular and molecular neurobiology.* Academic Press, San Diego.

Harding, A. J., Halliday, G. M., and Kril, J. J. (1998). Variation in hippocampal neuron number with age and brain volume. *Cerebral Cortex*, 8(8):710–718.

Henze, D. A., Urban, N. N., and Barrionuevo, G. (2000). The multifarious hippocamal mossy fiber pathway: a review. *Neuroscience*, 98(3):407–427.

Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation.* Addison-Wesley, Redwood City, CA.

Hölscher, C. (1999). Synaptic plasticity and learning and memory: LTP and beyond. *Journal of Neuroscience Research*, 58(1):62–75.

Howard, R. E., Schwartz, D., Denker, J. S., Epworth, R. W., Graf, H. P., Hubbard, W. E., Jackel, L. D., Straughn, B. L., and Tennant, D. M. (1987). An associative memory based on an electronic neural network architecture. *IEEE Trans. ED*, 34:1553.

Jessberger, S. and Kempermann, G. (2003). Adult-born hippocampal neurons mature into activity-dependent responsiveness. *European Journal of Neuroscience*, 18(10):2707–2712.

Jones, S. and Yakel, J. L. (1999). Inhibitory interneurons in hippocampus. *Cell Biochem. Biophys.*, 31(2):207–218.

Jung, M. W. and McNaughton, B. L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*, 3(2):165–182.

Kandel, E. R. (2000). Cellular mechanisms of learning and the biological basis of individuality. In Kandel, E. R., Schwartz, J. H., and Jessel, T. M., editors, *Principles of Neural Science*, chapter 63. McGraw-Hill.

Kempermann, G. and Gage, F. H. (2002). Genetic determinants of adult hippocampal neurogenesis correlate with acquisition, but not probe trial performance, in the water maze task. *European Journal of Neuroscience*, 16(1):129–136.

Kempermann, G., Gast, D., Kronenberg, G., Yamaguchi, M., and Gage, F. H. (2003). Early determination and long-term persistence of adult-generated new neurons in the hippocampus of mice. *Development*, 130(2):391–399.

KEMPERMANN, G. AND KRONENBERG, G. (2003). Depressed new neurons? — Adult hippocampal neurogenesis and a cellular plasticity hypothesis of major depression. *Biological Psychiatry*, 54(5):499–503.

KEMPERMANN, G., KUHN, H. G., AND GAGE, F. H. (1997). Genetic influence on neurogenesis in the dentate gyrus of adult mice. *Proceedings of the National Academy of Sciences of the United States of America*, 94(19):10409–10414.

KEMPERMANN, G., KUHN, H. G., AND GAGE, F. H. (1998). Experience-induced neurogenesis in the senescent dentate gyrus. *The Journal of Neuroscience*, 18(9):3206–3212.

KEMPERMANN, G. AND WISKOTT, L. (2004). What is the functional role of new neurons in the adult dentate gyrus? In GAGE, F. H., BJÖRKLUND, A., PROCHIANTZ, A., AND CHRISTEN, Y., editors, *Proc. Stem Cells in the Nervous System: Functional and Clinical Implications, Paris, January 20, 2003*, Research and Perspectives in Neurosciences (Fondation Ipsen), pages 57–65, Berlin. Springer.

KEMPERMANN, G., WISKOTT, L., AND GAGE, F. H. (2004). Functional significance of adult neurogenesis. *Curr. Opin. Neurobiol.*, 14:186–191.

KRONENBERG, G., REUTER, K., STEINER, B., BRANDT, M. D., JESSBERGER, S., YAMAGUCHI, M., AND KEMPERMANN, G. (2003). Subpopulations of proliferating cells of the adult hippocampus respond differently to physiologic neurogenic stimuli. *The Journal of Comparative Neurology*, 467(4):455–463.

LAVENEX, P. AND AMARAL, D. G. (2000). Hippocampal-neocortical interaction: a hierarchy of associativity. *Hippocampus*, 10(4):420–430.

LEE, J. H., DUAN, W., LONG, J. M., INGRAM, D. K., AND MATTSON, M. P. (2000). Dietary restriction increases the number of newly generated neural cells, and induces BDNF expression, in the dentate gyrus of rats. *Journal of Molecular Neuroscience*, 15(2):99–108.

MARTIN, S. J. AND MORRIS, R. G. (2002). New life in an old idea: the synaptic plasticity and memory hypothesis revisited. *Hippocampus*, 12(5):609–636.

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.

MCCLOSKEY, M. AND COHEN, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In BOWER, G. H., editor, *The psychology of learning and motivation*, volume 24, pages 109–165. Academic Press, New York.

MORRIS, R. G., GARRUD, P., RAWLINS, J. N., AND O'KEEFE, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, 297(5868):681–683.

MOSER, E. I. AND PAULSEN, O. (2001). New excitement in cognitive space: between place cells and spatial memory. *Current Opinion in Neurobiology*, 11(6):745–751.

Nadel, L. and Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, 7(2):217–227.

Nilsson, M., Perfilieva, E., Johansson, U., Orwar, O., and Eriksson, P. S. (1999). Enriched environment increases neurogenesis in the adult rat dentate gyrus and improves spatial memory. *Journal of Neurobiology*, 39(4):569–578.

O'Keefe, J. (1979). A review of the hippocampal place cells. *Progress in Neurobiology*, 13(4):419–439.

O'Reilly, R. C. and McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus*, 4(6):661–682.

Patton, P. E. and McNaughton, B. L. (1995). Connection matrix of the hippocampal formation: I. The dentate gyrus. *Hippocampus*, 5(4):245–286.

Rasch, M. (2003). Modellierung adulter Neurogenese im Hippocampus [Modeling adult neurogenesis in the hippocampus]. Diploma thesis, Institute for Biology, Humboldt University Berlin, D-10115 Berlin, Germany.

Rolls, E. T. (1999). Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, 9(4):467–480.

Santarelli, L., Saxe, M., Gross, C. G., Surget, A., Battaglia, F. P., Dulawa, S., Weisstaub, N., Lee, J. H., Duman, R., Arancio, O., Belzung, C., and Hen, R. (2003). Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science*, 301(5634):805–809.

Schmidt-Hieber, C., Jonas, P., and Bischofberger, J. (2004). Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. *Nature*, 429(6988):184–187.

Seki, T. and Arai, Y. (1995). Age-related production of new granule cells in the adult dentate gyrus. *NeuroReport*, 6(18):2479–2482.

Shors, T. J., Miesegaes, G., Beylin, A., Zhao, M., Rydel, T. A., and Gould, E. (2001). Neurogenesis in the adult is involved in the formation of trace memories. *Nature*, 410:372–376.

Shors, T. J., Townsend, D. A., Zhao, M., Kozorovitskiy, Y., and Gould, E. (2002). Neurogenesis may relate to some but not all types of hippocampal-dependent learning. *Hippocampus*, 12(5):578–584.

Snyder, J. S., Kee, N., and Wojtowicz, J. M. (2001). Effects of adult neurogenesis on synaptic plasticity in the rat dentate gyrus. *Journal of Neurophysiology*, 85(6):2423–2431.

Sutherland, G. R. and McNaughton, B. L. (2000). Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current Opinion in Neurobiology*, 10(2):180–186.

Tanapat, P., Hastings, N. B., Rydel, T. A., Galea, L. A., and Gould, E. (2001). Exposure to fox odor inhibits cell proliferation in the hippocampus of adult rats via an adrenal hormone-dependent mechanism. *The Journal of Comparative Neurology*, 437(4):496–504.

Treves, A. and Rolls, E. T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, 2(2):189–199.

Treves, A. and Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3):374–391.

Urban, N. N., Henze, D. A., and Barrionuevo, G. (2001). Revisiting the role of the hippocampal mossy fiber synapse. *Hippocampus*, 11(4):408–417.

van Praag, H., Christie, B. R., Sejnowski, T. J., and Gage, F. H. (1999a). Running enhances neurogenesis, learning, and long-term potentiation in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 96(23):13427–13431.

van Praag, H., Kempermann, G., and Gage, F. H. (1999b). Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nature Neuroscience*, 2(3):266–270.

van Praag, H., Schinder, A. F., Christie, B. R., Toni, N., Palmer, T. D., and Gage, F. H. (2002). Functional neurogenesis in the adult hippocampus. *Nature*, 415(6875):1030–1034.

Wang, S.-S., Scott, B. W., and Wojtowicz, J. M. (2000). Heterogenous properties of dentate granule neurons in the adult rat. *Journal of Neurobiology*, 42(2):248–257.

West, M. J. and Slomianka, L. (1998). Total number of neurons in the layers of the human entorhinal cortex. (and corrigendum). *Hippocampus*, 8:69–82 (and 426).

Witter, M. P. (1993). Organization of the entorhinal-hippocampal system: a review of current anatomical data. *Hippocampus*, 3(Spec. No.):33–44.

Witter, M. P., Naber, P. A., van Haeften, T., Machielsen, W. C., Rombouts, S. A., Barkhof, F., Scheltens, P., and Lopes da Silva, F. H. (2000). Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways. *Hippocampus*, 10(4):398–410.