From adult dentate gyrus neurogenesis to pattern separation

Thèse N° 9789

Présentée le 19 décembre 2019

à la Faculté des sciences de la vie
Laboratoire de calcul neuromimétique (SV/IC)
Programme doctoral en neurosciences

pour l'obtention du grade de Docteur ès Sciences

par

Olivia GOZEL

Acceptée sur proposition du jury

Prof. R. Schneggenburger, président du jury
Prof. W. Gerstner, directeur de thèse
Prof. J. Bischofberger, rapporteur
Prof. L. Wiskott, rapporteur
Prof. M. Gastpar, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

The future belongs to those who believe
in the beauty of their dreams.
— Eleanor Roosevelt

Acknowledgements

I thank Wulfram Gerstner for giving me the opportunity to pursue my PhD in a stimulating group. The balance of the freedom he gave me to follow my intellectual interests and his rigorous feedback on my work made me grow into a more independent and self-driven researcher. I would also like to thank all LCN lab members, past and present, for contributing to make the lab a lively environment, in particular those that became more than colleagues. A huge thanks to the "SV-IT guys" who were always efficiently and cheerfully solving all computer-related issues. Thanks also to Chantal Mellier and Rosana Turielle for taking care of the administrative work.

I am very grateful to Josef Bischofberger, Laurenz Wiskott, Michael Gastpar and Ralf Schneggenburger to have accepted to be part of my PhD defense committee, for the great discussion during the defense, and the valuable comments they gave on my work.

I would also like to thank the people at the Methods in Computational Neuroscience Course in Woods Hole, in particular Loren Frank, Sara Solla and Ila Fiete for being the trigger to a scary but absolutely worth switch of project. Thanks also to Bard Ermentrout, Mark Goldman and Kathleen Champion to bring scientific excitement with another project.

I am very grateful to the amazing friends who enrich my life outside the lab - in particular "les Rolloises", as well as Marina, Maya and Adi - for all the beautiful moments spent together and for bringing other, enlightening views on life.

A huge thanks goes to Vincent, who always backs up my goals, be they scientific or not, and for embellishing my life no matter the physical distance between us.

Lastly and most importantly, I would like to give heartful thanks to my parents Philippe and Marisa, and my brother Samuel. Without them, I would not be where I am right now. They have always been incredibly supportive and understanding during my studies, and along my PhD journey. They bring joy in my life, and provide me a safe and stable environment. I know that they will always be there for me, and I do not think I will ever thank them enough for that.

Lausanne, October 2019 Olivia Gozel

Abstract

For a few decades, adult dentate gyrus neurogenesis has been widely recognized by the neuroscience community as an intriguing phenomenon. Two observations are particularly puzzling. At the cellular level, the switch from excitation to inhibition of the GABAergic input onto newborn cells has been shown to be crucial for their proper integration into the existing network of dentate gyrus cells. At the behavioral level, adult-born dentate granule cells have been shown to promote pattern separation of similar stimuli in various tasks, while not playing a role in discrimination of distinct stimuli. It is still unclear, however, how these functionalities arise in the network of dentate gyrus cells. Several models of adult dentate gyrus neurogenesis have been designed with various levels of abstraction, and have suggested different roles of newborn cells. Yet, none of these models could explain how newborn cells promote pattern separation of similar stimuli, and not distinct stimuli. Moreover, none of the previous studies modeled the actual integration of adult-born dentate granule cells in the preexisting circuit, but rather initialized their inward connections to random, but fully grown, weights.

In my thesis work, I bridge the gap between biological and theoretical knowledge on adult dentate gyrus neurogenesis. I address the puzzling experimental observations and explain for the first time with a model: (i) how newborn cells integrate into the preexisting dentate gyrus network, and (ii) how they promote pattern separation of similar stimuli.

More specifically, I propose that the early phase of maturation of newborn cells, when GABAergic input has an excitatory effect, drives the synaptic weights towards the subspace of configurations of familiar stimuli through a cooperative effect. In the late phase of maturation, when GABAergic input switches to inhibitory, the synaptic weights move towards novel features of the presented stimuli through a competitive effect. This theory of newborn cells integration also explains why adult-born dentate granule cells promote better pattern separation of similar stimuli, but not distinct stimuli. Indeed, in the late phase of maturation, newborn cells can only learn novel features that are similar enough to familiar features, because the configuration of their synaptic weights makes them sensitive to familiar features at the end of the early phase of maturation.

Keywords: Adult dentate gyrus neurogenesis, Competitive network, Unsupervised learning, Synaptic plasticity, Pattern separation

Résumé

Depuis les années 1990, la neurogenèse adulte dans le gyrus denté est largement reconnue dans la communauté des neurosciences comme un phénomène intrigant. Deux observations sont particulièrement troublantes. Au niveau cellulaire, il a été démontré que le passage d'excitateur à inhibiteur des contributions GABAergiques reçues par les nouvelles cellules granulaires est crucial pour qu'elles s'intègrent correctement dans le réseau existant de cellules du gyrus denté. Sur le plan comportemental, il a été démontré que les cellules granulaires du gyrus denté nées chez l'adulte favorisent la séparation de motifs similaires dans diverses expériences, mais n'ont pas d'effet sur la discrimination de motifs distincts. Toutefois, on ne sait toujours pas comment ces fonctionnalités apparaissent dans le réseau de cellules du gyrus denté. Plusieurs modèles de neurogenèse adulte dans le gyrus denté ont été conçus avec différents niveaux d'abstraction et ont suggéré différents rôles pour les nouvelles cellules granulaires. Pourtant, aucun de ces modèles n'élucide comment les nouvelles cellules granulaires favorisent la séparation de motifs similaires, mais pas celle de motifs distincts. En outre, aucune des études précédentes n'a modélisé l'intégration des nouvelles cellules granulaires dans le circuit adulte préexistant. A la place, elles ont initialisé les connexions des nouvelles cellules granulaires à des valeurs certes aléatoires, mais aussi élevées que celles des cellules matures préexistantes.

Dans mon travail de thèse, je relie les connaissances biologiques et théoriques sur la neurogenèse adulte du gyrus denté. Je me base sur les observations expérimentales déconcertantes et explique pour la première fois avec un modèle : (i) comment les nouvelles cellules granulaires s'intègrent dans le réseau préexistant du gyrus denté adulte, et (ii) comment elles favorisent la séparation de motifs similaires.

Plus spécifiquement, je propose que la phase précoce de la maturation des nouvelles cellules, lorsque les contributions GABAergiques ont un effet excitateur, entraîne les poids des connexions synaptiques vers le sous-espace des configurations de motifs familiers grâce à un effet coopératif. Durant la phase de maturation tardive, lorsque les contributions GABAergiques deviennent inhibitrices, les poids synaptiques évoluent vers de nouvelles caractéristiques des motifs présentés, par le biais d'un effet compétitif. Cette théorie d'intégration des cellules granulaires du gyrus denté nées chez l'adulte explique également pourquoi les nouvelles cellules favorisent une meilleure séparation des motifs similaires, mais pas des motifs distincts. En effet, pendant la phase de maturation tardive, les cellules nées chez l'adulte ne

peuvent apprendre que des nouvelles caractéristiques qui sont assez similaires de caractéristiques familières, car la configuration de leurs poids synaptiques les rend sensibles aux caractéristiques familières à la fin de la première phase de maturation.

Mots-clefs : Neurogenèse adulte du gyrus denté, Réseau compétitif, Apprentissage non supervisé, Plasticité synaptique, Séparation de motifs

Contents

List of Figures

List of Tables

Introduction Part I

1 Classical views of the hippocampus

1.1 Introduction

Animals, including humans, are characterized by their ability to adapt to a wide range of environments and situations. They can learn new tasks, and keep memories of previous experiences in order to avoid unpleasant events and maximize reward.

In 1953, Henry Molaison, known as "patient H.M." and probably the most famous patient in neuroscience, had a bilateral surgical resection of his medial temporal lobe to try to avoid epileptic seizures. It lead to severe anterograde amnesia (Scoville and Milner (1957)). Since then, it is believed that the hippocampus is the substrate of learning and memory (Jarrard (1993)). Subsequently, many experiments, both in vitro and in vivo have supported this view. Hippocampal synapses have been shown to have high plasticity, and different hippocampal areas and synapses have been shown to be necessary for behavioral learning and recall of various tasks (reviewed in Neves et al. (2008)).

It is mostly believed that hippocampus acts as a temporary memory store until memories are consolidated in cortex by a process called "systems consolidation", whose exact timescale is still unclear (reviewed in Preston and Eichenbaum (2013)).

1.2 Hippocampus architecture

The hippocampus is a deep brain area. It is organized in a loop structure, whose main relays are the dentate gyrus (DG), Cornu Ammonis 3 (CA3), and Cornu Ammonis 1 (CA1), see Figure 1.1. The dentate gyrus, the entry area of the hippocampus, receives most of its input from layer II of the entorhinal cortex (EC) through the perforant path (Amaral et al. (2007)). The perforant path also directly projects to CA3, and connections from EC layer III (temporoammonic pathway) also directly connect to CA1. In addition, backprojections from CA3 to DG have been observed (Scharfman (2007)).

The entorhinal cortex provides two independent inputs to the dentate gyrus: one from its

medial part (MEC) and one from its lateral part (LEC) (McNaughton (1980)). Spatial information mainly comes from MEC, because many MEC neurons exhibit finely tuned place fields (grid-like representation of space), while non spatial components mostly come from LEC (Hargreaves et al. (2005)).

The principal cells of the dentate gyrus are called dentate granule cells (DGCs). In the rat, the dentate gyrus contains approximately 10^6 DGCs. As the number of EC input cells is estimated to be about $2 \cdot 10^5$, it makes an approximately 5 folds expansion from EC to DG (Andersen et al. (2007)). The dentate gyrus has sparse activity (Chawla et al. (2005)). This is probably due to the fact that within the dentate gyrus, there are several different classes of inhibitory neurons that are recurrently connected to the DGCs (Amaral et al. (2007)). In addition, the dentate gyrus also contains excitatory Mossy cells, which mostly innervate local inhibitory neurons and DGCs (Amaral et al. (2007)).

DGCs project to CA3 pyramidal cells and interneurons through Mossy fibers (Amaral et al. (2007)). The number of synapses per cell is very low: each CA3 pyramidal cell receives only about 72 Mossy fiber inputs (Andersen et al. (2007)). However, these are highly efficient synapses: input from a single DGC firing at high rate is sufficient to discharge CA3 pyramidal cells and interneurons, hence they are called "detonator synapses" (Henze et al. (2002)).

After the expansion in neuron numbers from EC to DG, there is again a compression from DG to CA3, as there are only about $3 \cdot 10^5$ pyramidal cells in the rat CA3 (Andersen et al. (2007)). Area CA3 contains a lot of plastic excitatory recurrent connections (Debanne et al. (1998)).

CA3 pyramidal cells project to CA1 pyramidal cells through long parallel fibers called the Schaffer collaterals. This handy configuration is the reason why synaptic plasticity has been extensively studied at those synapses (Bear and Malenka (1994)). Using brain slices, the fibers can indeed be extracellularly stimulated far away enough from the synaptic contacts such that there is minimal recording artifacts in CA1.

1.3 Classical models of the hippocampus

David Marr proposed in 1971 a theory of the hippocampus (Marr (1971)). Since then, many papers have been published on the subject, with slightly varying views on the exact computational function of the hippocampus. The classical view, which is presented here, has nevertheless remained very close to the one proposed in the original, groundbreaking, paper.

The hippocampus, called "archicortex" in Marr's paper, is believed to be responsible for memory storage (Marr (1971)). Its particular architecture influenced the putative functionality of its different areas. In addition, the similarity between the cerebellar cortex and the dentate gyrus, in particular the expansion of neuron numbers, contributed to the view that dentate gyrus has a pattern separation function, as was suggested for the cerebellum (Marr (1969)).

Inspired by Marr's original work, Treves and Rolls (1994) proposed an influent and compre-

hensive theory of the computational function of the hippocampus. According to their view, hippocampus is a temporary storage buffer for episodic memories, before their consolidation in the cortex. This stands in agreement with the observation that bilateral hippocampal lesions induce loss of recent memories, but do not affect remote memories (Scoville and Milner (1957)). It is not clear, however, for how long memories are stored in the hippocampus, or in other words for how long hippocampus is required for memory retrieval. Treves and Rolls (1994) suggested that it is not a matter of absolute time per se, but rather dependent on how many new memories have to be stored. Accordingly, a lively environment with lots of relevant events to remember would lead to faster memory consolidation into the cortex. Another competing view is that instead of being the storage location of the memories themselves, hippocampus would only store pointers to those memories (Káli and Dayan (2004)). As hippocampus would also be required for memory retrieval in this case, it is difficult to discard one of these views. Another influent theory of hippocampus suggests that it is not simply a blind memory store, but also performs spatial computation (O'Keefe and Nadel (1978); McNaughton et al. (1991)).

As mentioned above, due to the high expansion in neuron numbers between EC and DG, as well as the sparse activity in DG, it is mainly believed that DG acts as a pattern separator by decorrelating input activity (Treves and Rolls (1994)). In this way, similar (overlapping) inputs can be represented by distinct neuronal ensembles in DG, therefore promoting more accurate memory storage in the hippocampus.

Because CA3 contains numerous excitatory recurrent connections which follow a Hebbian plasticity rule (Debanne et al. (1998)), it is believed to be the center of associative memories (Treves and Rolls (1994)). According to this view, each memory is stored as an ensemble of neurons with strong excitatory recurrent connections between them, as in a Hopfield network (Hopfield (1982)). Therefore, CA3 is believed to perform pattern completion: if a degraded input is presented to the network, CA3 will recover the memory through the recurrent excitatory connections, in agreement with what has been observed experimentally (Neunuebel and Knierim (2014)).

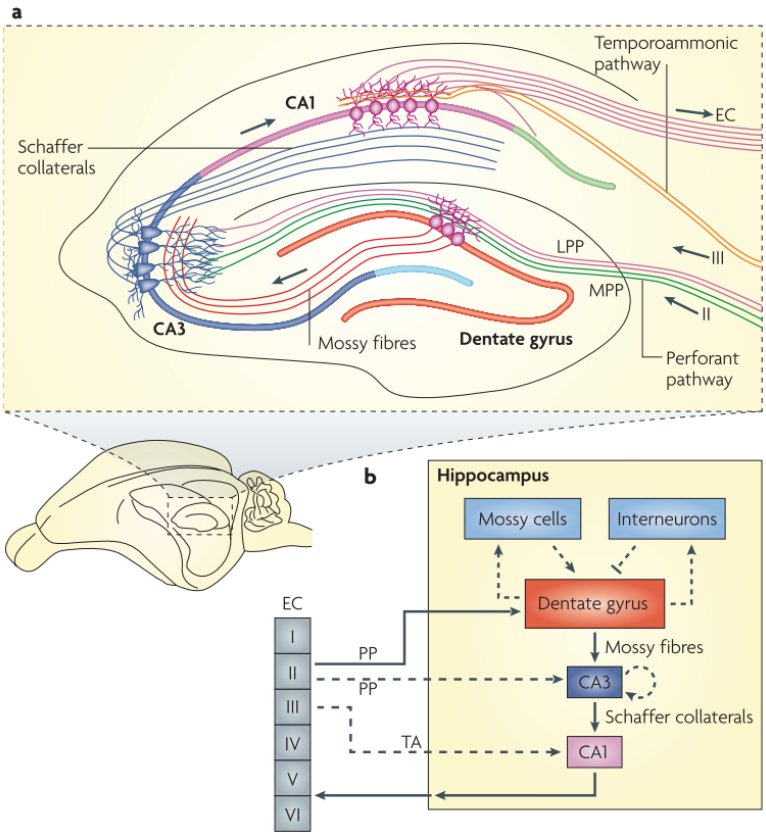

Figure 1.1 – Architecture of the rodent hippocampus. (**a**) The hippocampus consists of the dentate gyrus (DG), CA3 and CA1. Dentate gyrus receives most of its input from EC layer II through the perforant pathway (PP), which also directly project to CA3. The temporoammonic pathway (TA) connects EC layer III to CA1. (**b**) The dentate gyrus shows lots of recurrent connections between its principal cells, the dentate granule cells, and excitatory Mossy cells as well as inhibitory interneurons. The dentate granule cells provide detonator synapses to CA3 through the Mossy fibre pathway. CA3 connects to CA1 through the Schaffer collaterals. CA1 connects back to EC, projecting to layers V/VI. (Reproduced with permission from Figure 1 of Deng et al. (2010).)

2 Adult dentate gyrus neurogenesis

2.1 Introduction

For a long time, it was believed that all neurons were produced during development, and no new neurons were generated at the adult stage.

The first account of neurogenesis in adult brains dates from 1962, when Joseph Altman lesioned rat brains and injected tritiated thymidine (thymidine-H^3) in the lesioned area. Thymidine-H^3 is an analog of the DNA building block thymidine, hence it can be incorporated into the DNA of dividing cells. The animals were sacrificed between 1 day and 2 months after the lesion. Using radiography, labeled cells were detected indicating that cells were proliferating (Altman (1962)). A few years later, Altman and Das did autoradiography of brains from healthy rats after injection of tritiated thymidine. They discovered a high number of labeled granule cells in the dentate gyrus, which decreased with age (Altman and Das (1965)). This indicated that adult neurogenesis is a natural brain process which does not need injury to be triggered. Dentate granule cells have been subsequently observed to be generated at the adult stage in other species, such as guinea-pigs (Altman and Das (1967)), mice (Caviness Jr (1973)), and rabbits (Gueneau et al. (1982)). However, these findings were mostly ignored.

Thanks to the development of novel labeling and imaging techniques, the presence of adult neurogenesis was confirmed by Gage and colleagues in the nineties (Kuhn et al. (1996)). Bromodeoxyuridine (BrdU), another analog of the DNA building block thymidine, was injected intraperitoneally in adult rats of different ages, which were then sacrificed at different time points after injection. Immunohistochemistry was performed to label proliferating and migrating dentate granule cells precursors. A significant number of BrdU-labeled cells was observed at all ages studied, up to 27 month old (Kuhn et al. (1996)). Using the same technique, newborn cells were observed in the dentate gyrus of adult humans (Eriksson et al. (1998)), further corroborating the presence of adult neurogenesis.

These results finally got accepted by the neuroscience community, and adult neurogenesis became extensively examined in different brain areas of various species. Most studies were

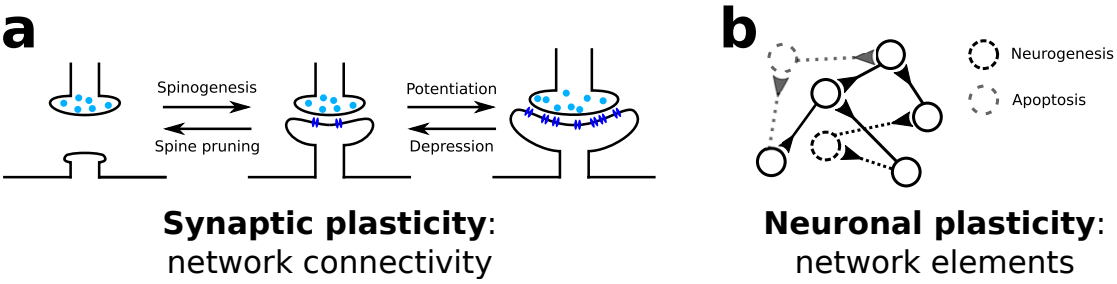

Figure 2.1 – Different forms of plasticity. (**a**) Synaptic plasticity is the modification of the connections between neurons. New connections, or spines, are created through spinogenesis. Inversely, connections can be eliminated by spine pruning. The strength of existing connections can vary as well: it either increases through potentiation, or decreases as a consequence of depression. (**b**) Neuronal plasticity provides a more profound change of the network structure. It involves the generation of new neurons through neurogenesis or the elimination of neurons by apoptosis. It also includes shifts of firing threshold and excitability (not considered here).

performed in mammals, more specifically rodents, but adult neurogenesis was also observed in non-mammalian vertebrates (see Chapouton et al. (2007) for a review). Mammalian adult neurogenesis seems to be restricted to two brain areas under normal conditions: the olfactory bulb and the dentate gyrus (thoroughly reviewed in Ming and Song (2011)). In the first case, newborn cells are produced in the subventricular zone of the lateral ventricles and migrate through the rostral migratory stream to become interneurons in the olfactory bulb. In the second case, newborn cells are created on the spot, in the subgranular zone of the dentate gyrus, and migrate only slightly radially.

The acceptance of adult neurogenesis has lead to a paradigm shift, as it challenges the previously supposed stability of the "neuronal content" of the brain. Indeed, the number of neurons is not simply reduced through aging and disease, but can be replenished. This adds a second dimensionality and a longer timescale to brain plasticity, see Figure 2.1. Indeed, before the work of Gage and colleagues the only established form of plasticity was the modification of the connection strength between neurons, namely synaptic plasticity, spanning timescales from milliseconds to days. On one side of the spectrum, it involves the strengthening (potentiation), or weakening (depression) of the synaptic contacts between neurons, which is induced on a timescale of tens of milliseconds (Levy and Stewart (1983)) and can persist from several hundreds of milliseconds in the case of short-term plasticity (Varela et al. (1997)), to minutes or even hours in the case of long-term plasticity (Sajikumar and Frey (2004)). Long-term potentiation (LTP) denotes an increase in the synaptic strength of a connection, whereas long-term depression (LTD) implies a decrease in the strength of the connection. On the other side of the spectrum, new dendritic spines can be created and existing ones can disappear through pruning, on a timescale of days (Trachtenberg et al. (2002); De Paola et al. (2006)). Adult neurogenesis broadens the timescale of plastic changes to an order of weeks, as newborn cells integrate in the preexisting circuit in about 8 weeks (Ming and Song (2005)). This process

is counteracted by the death, or apoptosis, of neurons.

Even though rodent adult dentate gyrus neurogenesis is nowadays well established, it is still controversial if it is important in humans. Recently, a paper claiming that hippocampal neurogenesis is negligible in adult humans (Sorrells et al. (2018)) triggered lively discussions about its significance. One month later, another paper on the contrary asserted the persistence of neurogenesis in the hippocampus of adult humans (Boldrini et al. (2018)), adding fuel to the debate. Several arguments have been put forward for the discrepancy of these results, in particular the postmortem interval before analysis, the tissue handling and labeling techniques, as well as the clinical status of the patients (Tartt et al. (2018)). It has also been suggested that seemingly contradictory results could be reconciled by taking the neurodevelopmental timing into account (Snyder (2019)). Overall, these studies emphasize that adult neurogenesis is still a hot topic in the field of neuroscience, and that it contains many mysterious aspects which deserve to be investigated.

2.2 Adult-born dentate granule cells

We focus here on adult hippocampal neurogenesis, where most findings come from rodent studies. Newborn neurons mostly develop into dentate granule cells (DGCs), the main excitatory cell type of the dentate gyrus, while the remaining 15 to 25% become astrocytes or present another phenotype (Van Praag et al. (1999)). It has been estimated that the amount of newly generated DGCs per month represents about 6% of the total granule cell population in the rat (Cameron and McKay (2001)). Despite this relatively small percentage, it corresponds to several thousands of cells due to the high number of cells in the dentate gyrus: about 10^6 in one dentate gyrus of the rat (West et al. (1991); Rapp and Gallagher (1996)).

The number of proliferating cells in the dentate gyrus, as well as the percentage of surviving newborn DGCs after one month, can be modulated by experimental conditions (Van Praag et al. (1999)). Yet, interestingly, the dentate gyrus volume remained similar among all experimental groups (Van Praag et al. (1999)). This is in agreement with the observation that in rats the number of DGCs is about constant as a function of age (Boss et al. (1985); Rapp and Gallagher (1996)) and the number of apoptotic cells seems to counterbalance newly generated cells (Biebl et al. (2000)), favoring the view of adult dentate gyrus neurogenesis as a replacement process over an addition process. We now review these and related results in more details.

2.2.1 Experience-dependent survival and integration

Proliferation of adult-born DGCs can be increased by an enriched environment (Kempermann et al. (1997); Tashiro et al. (2007)), voluntary running (Zhao et al. (2006)), and hippocampus-dependent learning (Gould et al. (1999)). Many of the newborn DGCs die in the early stages of their maturation, both in rats (Biebl et al. (2000); Dayer et al. (2003)) and in mice (Kempermann et al. (2003); Sierra et al. (2010)). The first few weeks after birth are critical for their long-term

survival (Dayer et al. (2003); Kempermann et al. (2003)). Importantly, survival has been shown to be experience-dependent: it depends on the input on newborn DGCs in a restricted time window early in their maturation. It is promoted by high-frequency stimulation of perforant path fibers targeting the dentate gyrus (Kitamura et al. (2010)), and by the learning of tasks that require the hippocampus (Gould et al. (1999); Kee et al. (2007)). Furthermore, voluntary running has been shown to slightly advance their spine growth (Zhao et al. (2006)), and accelerate their neuronal maturation (Piatti et al. (2011)).

Van Praag et al. (1999) compared several groups of mice: (1) standard housing (control), (2) enriched housing, (3) cage with a wheel (voluntary runners), (4) mice which were placed in a water-maze devoid of any platform for a given amount of time (forced swimmers), and (5) mice which were put in a water-maze with a platform (learners). Voluntary runners were the only mice with a significantly higher proliferation of cells in the dentate gyrus than all other conditions. Similarly, voluntary runners and mice housed in an enriched cage were the only groups whose numbers of surviving adult-born DGCs after one month was higher than control. This suggests that voluntary exercise is beneficial for neuronal turnover in the adult dentate gyrus, and that arousal by running or rich environments promotes the survival of newborn DGCs.

Rats which were trained on a trace eyeblink conditioning task or a spatial water-maze task, both hippocampus-dependent associative tasks, had a higher number of newborn DGCs than animals trained on hippocampus-independent tasks, such as delay-eyeblink conditioning or cue-maze training (Gould et al. (1999)). Using the water-maze task, analogous increases of newborn DGCs (Van Praag et al. (1999)), and integration of adult-born DGCs in the spatial memory network (Kee et al. (2007)) have been observed in mice.

NMDA receptor activation has been shown to be crucial for proper integration, and hence survival, of adult-born DGCs. Pharmacological inhibition of NMDA receptors prevents enhanced survival of newborn DGCs upon high-frequency stimulation in rats (Kitamura et al. (2010)). Similarly, a local knockout of the NMDA receptor NR1 subunit, which is necessary for its function, significantly reduces the survival rate of newborn DGCs which are in their third week after birth (Tashiro et al. (2006)). Moreover adult-born neurons are preferentially reactivated by stimuli similar to the ones they experienced during their second to third week after birth (Tashiro et al. (2007)).

Spatial learning has been shown to increase the complexity of the dendritic arborescence, as well as the number of spines of adult-born DGCs (Tronel et al. (2010)). These morphological changes are long-lasting: they are still present at least 3 months after learning. They correlate with the cognitive load of the learning task, and depend on activation of NMDA receptors (Tronel et al. (2010)). However, these structural changes are restricted to adult-born DGCs, suggesting that mature DGCs have lost their plasticity.

Taken together, these findings suggest that newborn DGCs encode features of stimuli that they experienced in a restricted and early phase of their maturation. Because LTP induction is

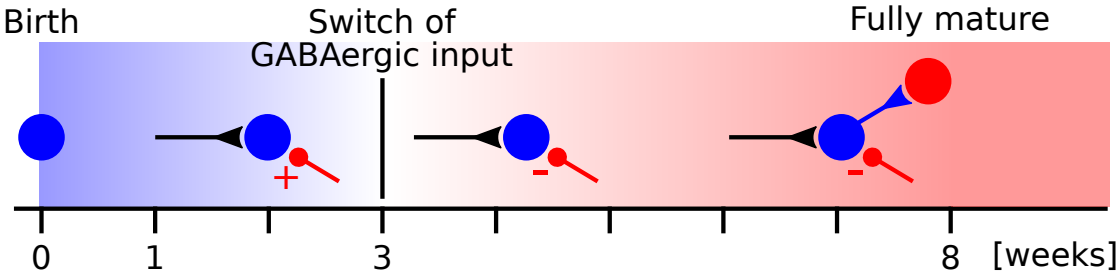

Figure 2.2 – Maturation timeline of the newborn DGCs. Integration of an adult-born DGC (blue) into the preexisting network as a function of time. Around 1 week after birth, the cell receives GABAergic synaptic input (red) (Espósito et al. (2005); Ge et al. (2006); Deshpande et al. (2013)), and a few days later also glutamatergic synaptic input (black) (Ge et al. (2006); Vivar et al. (2012)). At about 3 weeks, the GABAergic input switches from excitatory to inhibitory (Ge et al. (2006)). At 7 weeks, the newborn DGC connects to local interneurons (red) (Temprana et al. (2015)).

necessary, this encoding relies upon an associative-type of learning, which has morphological correlates in the form of dendritic branching and spinogenesis.

2.2.2 Time course of connectivity

The cellular and connectivity properties of adult-born DGCs change as a function of their maturation stage, until they become indistinguishable from any other mature DGCs at approximately 8 weeks (Deng et al. (2010); Johnston et al. (2016)). At birth, newborn DGCs are completely disconnected from the preexisting dentate gyrus circuit. If they mature successfully, they connect to pre- and postsynaptic partners in a sequence of steps, see Figure 2.2. Many experiments using retroviral labeling of adult-born DGCs of specific ages have been performed to determine the precise time course of their connectivity. On one hand, morphological aspects of connectivity have been investigated using retrograde tracing and microscopy. On the other hand, electrophysiology and optogenetics have been used to assess the functionality of the detected connections. All these studies were mostly performed in mice.

Initially, newborn DGCs do not receive synaptic contacts and they do not project to postsynaptic partners. Indeed, between 1 and 7 days after birth, no synaptic partners have been detected (Espósito et al. (2005)). Ambient GABA is thus their first source of tonic activation, detected as early as 3 days after birth (Ge et al. (2006)). Even though newborn DGCs lack synaptic connections, they can already sense dentate gyrus activation, because stimulation of local interneurons induces an increase in the tonic activation of newborn DGCs (Ge et al. (2006)). This result is further supported by the observation that focal application of GABA under voltage or current clamp induces a strong response in newborn DGCs, indicating that neurotransmitter receptors are already present in their plasma membrane before the arrival of presynaptic terminals (Espósito et al. (2005)).

Newborn DGCs are first synaptically innervated by local GABAergic interneurons. GABAergic afferents were identified in 8 days old, but not 6 days old, newborn DGCs (Espósito et al. (2005)). They induce GABAergic synaptic inputs with slow rise and decay phases (Espósito et al. (2005); Overstreet-Wadiche et al. (2005)). Similarly, GABAergic postsynaptic currents (PSCs) were observed in some newborn DGCs that were 7 days old. These GABAergic PSCs were sensitive to bicuculline, an antagonist of $GABA_A$ receptors, highlighting the synaptic nature of the connections (Ge et al. (2006)). Innervation from both local and distant interneurons increases with maturation (Vivar et al. (2012)). Back projections from CA3 interneurons were also observed in 21 days old newborn DGCs (earliest time point investigated) and they remained stable thereafter (Vivar et al. (2012)).

After about one to two weeks, newborn DGCs start receiving glutamatergic synapses. They come from different sources, such as local Mossy cells, entorhinal cortex and back projections from CA3 pyramidal cells. Glutamatergic afferents have been detected in newborn DGCs that were 18 days old (Espósito et al. (2005)), and even as young as 14 days old (Ge et al. (2006)). The glutamatergic PSCs were sensitive to CNQX, an antagonist of AMPA receptors (Ge et al. (2006)). Innervation from Mossy cells was observed in 10 days old newborn DGCs (Deshpande et al. (2013)), and it increases over time (Vivar et al. (2012)). Innervation from entorhinal cortex was still sparse in newborn DGCs that were 21 days old (Vivar et al. (2012); Deshpande et al. (2013)), but the number of presynaptic partners increased fivefold in the following 2 weeks (Deshpande et al. (2013)). Patch-clamp recordings support innervation by the lateral entorhinal cortex (LEC), cortical layers II/III, rather than from the medial entorhinal cortex (MEC) (Vivar et al. (2012)). A sparse (10% of the presynaptic partners) but stable innervation from CA3 pyramidal cells in newborn DGCs from 21 to 90 days of age (earliest and last time point investigated respectively) has also been observed (Vivar et al. (2012)).

One study has suggested a major transient synaptic input from mature DGCs to newborn DGCs that are between 21 and 30 days of age (Vivar et al. (2012)). However, subsequent morphological and functional studies failed to support this result, and it is mostly believed that the observed excitatory feedback loop was an artifact due to pseudo-transduction (Deshpande et al. (2013); Temprana et al. (2015); Alvarez et al. (2016)).

Adult-born DGCs initiate Mossy fiber projections to CA3 relatively early in their maturation, but several weeks are needed for information to be stably transmitted to CA3. At 10 days of age newborn DGCs already establish axonal projections to the CA3 area (Zhao et al. (2006); Faulkner et al. (2008)), and at 17 days of age synapses with postsynaptic targets have been observed (Toni et al. (2008)). Optical stimulation of 2 weeks old adult-born DGCs produces weak glutamatergic excitatory postsynaptic currents (EPSCs) in CA3 pyramidal neurons, indicating that functional synapses were already established (Gu et al. (2012)). However, maturation is needed for newborn DGCs to reliably recruit CA3 pyramidal cells. Mossy fibers boutons grow in size between 2 and 4 weeks of age (Faulkner et al. (2008)), and induction of stable maximal EPSCs has been observed from newborn DGCs that were at least 4 weeks of age (Gu et al. (2012); Temprana et al. (2015)). Ultrastructural analysis has revealed that mossy fiber boutons

reach morphologic maturity by 8 weeks and remain stable thereafter (Faulkner et al. (2008)). It is interesting to note that newborn DGCs do not only project to CA3 pyramidal neurons, but also to CA3 interneurons, thus inducing indirect inhibition in addition to direct excitation to CA3 principal cells (Toni et al. (2008)). Furthermore, it has been observed that newborn DGCs initially make synaptic contacts with spines that are already receiving connections from mature DGCs, suggesting that newborn DGCs may compete for postsynaptic targets (Toni et al. (2008)).

Even though four-week old newborn DGCs efficiently drive distal CA3 pyramidal cells and interneurons, they only weakly activate local dentate gyrus interneurons, as revealed by small inhibitory postsynaptic currents (IPSCs) recorded in mature DGCs upon newborn DGCs stimulation (Temprana et al. (2015)). Newborn DGCs thus recruit local interneurons relatively late in their maturation, at around 7 weeks of age (no time point investigated between 4 and 7 weeks) (Temprana et al. (2015)).

To conclude, all these experiments suggest that maturation and integration of adult-born DGCs is analogous to what is observed during development, though at a slower pace. Tonic ambient GABA excitation is the first source of activation, followed by GABAergic synaptic inputs, and finally glutamatergic synaptic inputs. Dendritic growth is only slightly delayed in adult-born DGCs (Zhao et al. (2006)). The simultaneous maturation of dendritic spines and Mossy fiber axons suggests that the functional input and output of newborn DGCs are synchronized and emerge at a well-controlled time point during maturation.

2.2.3 Enhanced plasticity

During their maturation, adult-born DGCs have been shown to have enhanced plasticity relative to mature DGCs.

Using rat hippocampal slices, Wang et al. (2000) have shown that tetanic stimulation (4 trains of 100 Hz) of the medial perforant pathway induces LTP in newborn DGCs, but not in mature DGCs. The absence is due to compensatory inhibitory mechanisms: if an antagonist of $GABA_A$ receptors was present, mature DGCs also underwent LTP. Furthermore, pairing weak afferent stimulation (2 Hz) with postsynaptic depolarization under voltage clamp was sufficient to induce LTP in newborn DGCs, but had no effect on mature DGCs.

Schmidt-Hieber et al. (2004) have observed in slice that young (1 to 3 weeks) newborn DGCs have a lower threshold for spike-timing dependent plasticity induction than mature DGCs. Brief presynaptic bursts (ten stimuli at 100 Hz) paired with a single postsynaptic action potential are highly efficient for LTP induction in young adult-born DGCs, while a burst of postsynaptic action potentials is required in mature DGCs.

More precisely, the enhanced plasticity of newborn DGCs has been shown to occur in a restricted time window, when they are about 1 to 1.5 month old. During this period, they have a lower threshold for LTP induction and higher LTP amplitude (Ge et al. (2007)). Noteworthy,

LTP induction at any DGC age is dependent on the NMDA receptor subunit NR2B (Ge et al. (2007)).

Likewise, efferent connections from newborn DGCs to CA3 also show temporarily enhanced plasticity (Gu et al. (2012)). Upon theta-burst optical stimulation of newborn DGCs, LTP of the field excitatory postsynaptic potentials was reliably observed when the DGCs were 3- and 4-week old, but was more sporadic with 8-week old newborn DGCs. Furthermore, LTP amplitude was highest at 4 weeks, indicating that newborn DGCs have enhanced plasticity which peaks at around 4 weeks of age.

2.2.4 Enhanced excitability

During their maturation, adult-born DGCs were also shown to have enhanced excitability relative to mature DGCs. Schmidt-Hieber et al. (2004) observed that young newborn DGCs have a longer membrane time constant. However, they fire very few action potentials upon stimulation, whereas mature DGCs fire trains of action potentials.

Importantly, however, the enhanced excitability of newborn DGCs does not imply that they are hyperactive (Dieni et al. (2013); and thoroughly reviewed in Lodge and Bischofberger (2019)). Indeed, the enhanced excitability of newborn DGCs is counterbalanced by reduced excitatory innervation and functional shunting inhibition upon high GABAergic input. In addition, adult-born DGCs exhibit a slower membrane time constant, a reduced slope of action potential and lower firing rate than mature DGCs. All these compensating mechanisms are believed to preserve the sparse and orthogonal population activity of DGCs observed in vivo.

2.2.5 GABA switch

During maturation, at about 3 weeks of age, the γ-aminobutyric acid (GABAergic) input from interneurons to adult-born DGCs switches from excitatory to inhibitory, see Figure 2.2. This switch stems from the fact that in the early phase of maturation, newborn DGCs have a high intracellular chloride concentration due to a high expression of the Na^+-K^+-$2Cl^-$ cotransporter NKCC1 (a Cl^- importer). The GABA reversal potential is thus higher than the membrane resting potential (Overstreet-Wadiche et al. (2005); Ge et al. (2006); Heigele et al. (2016)). Hence, upon GABAergic input stimulation, $GABA_A$ ionic receptors let Cl^- flow out of the newborn DGCs, which results in depolarization. In the late phase of maturation, expression of the K^+-Cl^- -coupled cotransporter KCC2 (a Cl^- exporter) kicks in, which causes a reduction of the intracellular chloride concentration of newborn DGCs to a level similar to the one of mature DGCs. Hence GABAergic inputs drive Cl^- to flow in through $GABA_A$ ionic receptors, which generates neuronal hyperpolarization (reviewed in Ben-Ari (2002); Owens and Kriegstein (2002)).

This transition of GABAergic input effect is analogous to the so-called “GABA switch” observed

during embryonic and early postnatal stages. The early excitatory phase is thought to contribute to the proper development of the brain by supporting activity-dependent growth of dendrites and axons. The timing of the switch is crucial for healthy development of the brain. If it occurs too early, sensorimotor gating deficits are induced due to a short GABA-mediated depolarization (Wang and Kriegstein (2010)). If it happens too late, seizures are facilitated (Wang and Kriegstein (2010)) and lasting cognitive and physical deficits are promoted (Furukawa et al. (2017)) due to excessive excitation.

The switch is sharp: it occurs in about a day, both during adult dentate gyrus neurogenesis (Heigele et al. (2016)) and during development (Khazipov et al. (2004); Tyzio et al. (2007); Leonzino et al. (2016)). An early study had suggested a rather continuous switch, because the recorded GABA reversal potential decreased gradually during maturation of adult-born DGCs (Ge et al. (2006)). However, only few time points were investigated, possibly smoothing the apparent GABA reversal potential curve as a function of maturation. Indeed, a subsequent study has revealed a rather step-like change of GABA reversal potential (Heigele et al. (2016)).

The timing of the switch can be bidirectionally modulated. In cell cultures of hippocampal neurons harvested during rat early postnatal development, an increase of GABAergic activity upregulates KCC2 expression, hence advancing the timing of the switch (Ganguly et al. (2001)). On the contrary, a reduction of GABAergic input through blockade of $GABA_A$ receptors downregulates expression of KCC2, leading to a delayed switch. GABAergic activity itself seems to be crucial, and not simply the overall activity of the neurons, because neuronal spiking and glutamatergic activity alone do not alter the timing of the switch (Ganguly et al. (2001)).

2.2.6 GABAergic input importance for integration

The GABAergic input onto adult-born DGCs has been shown to be crucial for their integration into the preexisting dentate gyrus circuit.

In newborn DGCs that are between 1.5 and 3 weeks of age (when the GABA reversal potential is higher than the membrane resting potential), moderate activation of GABAergic synaptic inputs onto newborn DGCs has been shown to be sufficient to induce action potential firing, in particular when they were paired with glutamatergic synaptic inputs which would have stayed subtreshold otherwise (Heigele et al. (2016)). However, if GABAergic activation is too strong, it induced shunting at the reversal potential of the GABA channels, hence newborn DGCs are effectively inhibited (Heigele et al. (2016)).

As specified above, newborn DGCs are first activated by ambient GABA, then receive GABAergic synaptic inputs, and only later glutamatergic synaptic inputs (Figure 2.2). Therefore, their first inputs are GABAergic, which have an excitatory effect due to their depolarized GABA reversal potential. In addition, it has been shown that mature DGCs indirectly activate newborn DGCs that are early in their maturation through activation of local interneurons (Alvarez et al. (2016)). As mature DGCs are responding to environmental stimulation, the indirect activation

of newborn DGCs might provide a way for them to sense the dentate gyrus network activity. In hippocampal slices from mice, mature DGCs have been optically stimulated while 10-days old DGCs were intracellularly recorded (Alvarez et al. (2016)). A high depolarizing GABAergic current (but no glutamatergic current) has been reported in the newborn DGCs, suggesting that mature DGCs indirectly activate newborn DGCs through interneuron activation (Alvarez et al. (2016)). In addition, at the morphological level, newborn DGCs which were indirectly activated by stimulation of mature DGCs showed much more dendritic processes than control newborn DGCs, suggesting that they were better integrated into the preexisting circuit through synaptic contacts (Alvarez et al. (2016)).

This interpretation has been confirmed by the observation that early GABA-mediated depolarization is required for functional glutamatergic synaptic innervation. Early in maturation, even before the appearance of presynaptic terminals, newborn DGCs already have GABAergic and glutamatergic receptors in their membrane (Espósito et al. (2005)). More specifically, adult-born DGC glutamatergic synapses initially only contain NMDA receptors, but no AMPA receptors. They are "silent" synapses, because of the voltage-dependent magnesium block of NMDA receptors. Therefore, to establish functional glutamatergic synapses, the magnesium block has to be alleviated. GABA receptor mediated depolarization has been shown to be necessary to "unsilence" the synapses and trigger the incorporation of AMPA receptors, both in vitro and in vivo (Chancey et al. (2013)). Furthermore, unsilencing of synapses through GABA depolarization has been shown to be experience-dependent, as only 2 hours in an enriched environment is sufficient to promote this process (Chancey et al. (2013)).

The importance of the switch from early excitation to late inhibition of the GABAergic input onto newborn DGCs has been highlighted by studies that have either abolished or prolonged the early excitatory phase.

On one hand, GABA-mediated depolarization is necessary for proper development of functional GABAergic and glutamatergic synapses onto adult-born DGCs. Early excitation can be suppressed by knocking down the expression of NKCC1 using short hairpin RNA (shRNA). Hence shRNA-NKCC1$^+$ DGCs have lower intracellular chloride concentration and a lower GABA reversal potential (Ge et al. (2006)). At the morphological level, shRNA-NKCC1$^+$ newborn DGCs show an impaired dendritic development. Their dendritic arborisation, in terms of dendritic length and branch number, is significantly reduced (Ge et al. (2006)). This effect can be counteracted by the injection of an agonist of $GABA_A$ receptors, which promotes dendritic growth of shRNA-NKCC1$^+$ newborn DGCs in vivo (Ge et al. (2006)). At the functional level, shRNA-NKCC1$^+$ DGCs that are 7 days old do not receive postsynaptic currents (PSCs) through their GABAergic synapses, and when they are 14 and 28 days old, recorded PSPs has lower amplitude than control (Ge et al. (2006)). Furthermore, shRNA-NKCC1$^+$ DGCs that are 14 days old did not get PSCs through their glutamatergic synapses, and when they are 28 days old a lower percentage of cells receive PSCs than control, with a smaller amplitude (Ge et al. (2006)).

On the other hand, if GABA-mediated depolarization of adult-born DGCs is too long-lasting,

aberrant behavior is induced. For example, following an epileptogenic injury, expression of KCC2 is decreased in the DGCs for several weeks, leading to a higher GABA reversal potential (Pathak et al. (2007)). The resulting enhanced excitability of DGCs drives alteration of synaptic integration, as well as facilitate seizure initiation and propagation (Pathak et al. (2007)). Similarly, high GABA reversal potential due to a decreased KCC2-to-NKCC1 ratio has been observed in hippocampal slices from status epilepticus rats. It promotes hyperexcitability of the hippocampus following status epilepticus (Barmashenko et al. (2011)).

Taken together, these results suggest that early excitation, late inhibition, and proper timing of the switch of the GABAergic input onto adult-born DGCs is crucial for their proper integration into the preexisting dentate gyrus network.

2.3 Functional role of adult dentate gyrus neurogenesis

Even though the number of cells which are generated at the adult stage is rather low with respect to the whole DGC population, adult dentate gyrus neurogenesis is involved in hippocampus-dependent memory, and its modulation has a high impact on behavior, as summarized in Table 2.1.

2.3.1 Involvement in hippocampus-dependent memories

Adult-born DGCs have been shown to be implicated in hippocampus-dependent memory traces. Shors et al. (2001) used a systemic approach to reduce neurogenesis, through subcutaneous injection of a toxin for proliferating cells, the DNA methylating agent methylazoxymethanol acetate (MAM). Rats with decreased neurogenesis were compared to control in their acquisition of two eyeblink conditioning tasks, where the conditioned stimulus (CS) was white noise, and the unconditioned stimulus (US) a periorbital eyelid stimulation. The first task was trace conditioning, a hippocampus-dependent task where there is a temporal gap between the CS and the US. The second task was delay conditioning, a hippocampus-independent task where CS and US overlap. A reduction of neurogenesis was shown to impair acquisition of trace conditioning, but not of delay conditioning. More specifically, it was observed that newborn DGCs that were about 2 weeks old during training were involved in the trace memory, but not 1 week old newborn DGCs. Furthermore, a recovery period to replenish the number of newborn DGCs was sufficient for normal acquisition of trace conditioning memory.

In a follow-up study, Shors et al. (2002) observed that dentate gyrus neurogenesis is not involved in all hippocampus-dependent memories. Indeed, reduction of neurogenesis using the same systemic approach also impaired acquisition of a cued fear conditioning task, where rats were cued with a tone and received a foot shock after some temporal interval. However, contextual fear conditioning, exploratory behavior in an elevated plus maze, and spatial navigation learning in a Morris water maze, all hippocampus-dependent tasks, were not

impaired by a reduction in neurogenesis.

Analogously, decreased neurogenesis does not impair acquisition of the Morris water maze task when whole brain irradiation is used to reduce neurogenesis (Snyder et al. (2005)). However, newborn DGCs that are 4 to 28 days old at the time of training are necessary for long term memory retrieval at 2 and 4 weeks, but not 1 week (Snyder et al. (2005)).

2.3.2 Incorporation in memory networks

Newborn DGCs are incorporated into memory circuits and activated upon memory recall. Kee et al. (2007) labeled dividing cells, and trained mice in a Morris water maze task at different time points after the labeling procedure (1, 2, 4, 6 and 8 weeks). Ten weeks after labeling, they tested the mice in the same Morris water maze without the platform before sacrificing them. By investigating the overlap between newborn cells labeling and an activity-dependent marker, they observed that 1 to 2 weeks old newborn DGCs are not functionally integrated into spatial memory networks. On the other hand, newborn DGCs that are 4 weeks or older are incorporated. More specifically, 4-week to 8-week old cells have a 2 to 3 higher probability than mature DGCs to be recruited during memory recall, suggesting that they play an important role in dentate gyrus memory through their encoding abilities.

In addition, newborn DGCs which are recruited into memory circuits are necessary for recall. Similarly to Kee et al. (2007), Gu et al. (2012) have labeled newborn DGCs and trained mice in a Morris water maze task at different time points after labeling (2, 4 and 8 weeks). They observed that if newborn DGCs that were 4 weeks old during training were optogenetically silenced during testing, mice were spending less time in the quadrant where the hidden platform was located during training. This effect was not present upon silencing 2 or 8 weeks old newborn DGCs. Silencing of 4 weeks old newborn DGCs during training did not affect memory acquisition though, suggesting that other DGCs became responsible for the encoding of the spatial memory. Analogous results were found using contextual fear conditioning, another hippocampus-dependent learning task. During testing in the same context, optogenetic silencing of newborn DGCs that were 4 weeks old during training significantly reduced freezing.

The two above studies suggest that newborn DGCs have a critical period when they are preferentially incorporated into memory ensembles, and that they are required for memory retrieval due to their encoding of a novel experience. In agreement with these results, using a lentiviral approach to reduce adult dentate gyrus neurogenesis in rats by inhibiting Wnt signaling (which is involved in newborn DGCs generation), Jessberger et al. (2009) have observed that animals with a reduced number of newborn DGCs were impaired in the novel object recognition task, an hippocampus-dependent task. Indeed, retention at 3h and 4 weeks (but not 1min) was lower in the group with a large decrease of newborn DGCs than the control group. The group with a smaller reduction of neurogenesis was only impaired in retention at 3h with respect to control, and significantly less than the group with a high reduction of

newborn DGCs. In the Morris water maze task, in agreement with the observations of Gu et al. (2012), Jessberger et al. (2009) did not monitor any impairment in acquisition, but found an affected retention (from 2 to 8 weeks) in the group with high reduction of newborn DGCs.

2.3.3 Importance for pattern separation

Clelland et al. (2009) were the first to show the requirement of adult dentate gyrus neurogenesis for discrimination of inputs with low spatial separation, but not for discrimination of inputs with high spatial separation. The performance of mice with focal X-ray irradiation to ablate dentate gyrus neurogenesis were compared with the one of control mice on different hippocampus-dependent tasks. First, mice were trained in a delayed non-matching to place radial 8-arm maze task. Each trial consisted of a sample phase and a choice phase. In the sample phase, a mouse would be placed at the end of one arm (the "start arm"), and a single other arm (the "sample arm") was open, with a food reward at its end. In the choice phase, the mouse would be placed again in the same start arm, and this time two other arms were open: the sample arm, as well as a previously closed "choice arm". Mice would get a food reward only at the end of the choice arm, denoted as a correct trial. Different spatial separations were used between the sample and choice arms. When the spatial separation was high (two or three arms between the choice and sample arms), mice with ablated neurogenesis performed as well as controls. However, for trials with low spatial separation (only one arm between the choice and sample arms) mice with ablated neurogenesis were significantly impaired. Similar results were obtained when dentate gyrus neurogenesis was reduced using a lentiviral vector to knock-down Wnt signaling. Second, the authors confirmed the importance of adult dentate gyrus neurogenesis for spatial pattern separation using a spatial discrimination paradigm: mice were required to choose the correct spatial location between two illuminated boxes in two out of five possible locations on a touch screen. The two illuminated locations had either a high spatial separation (3 dark locations in between) or a low spatial separation (a single dark location in between). Again, mice with ablated adult dentate gyrus neurogenesis were impaired in trials with low spatial separation, but not in trials with high spatial separation. Yet, these same mice were not impaired in another spatial touch screen task which tests the ability to associate objects with spatial locations, further supporting the role of adult dentate gyrus neurogenesis in tasks with a pattern separation component. Furthermore, using a similar touch screen task to test spatial pattern separation abilities, Vivar et al. (2012) revealed the importance of a functional LEC input. Indeed, if an agonist of NMDA receptors was bilaterally injected in the LEC, mice were impaired in low spatial separation trials, but not in high spatial separation trials.

Using a contextual fear conditioning task, Sahay et al. (2011a) have confirmed the implication of adult dentate gyrus neurogenesis in behavioral pattern separation. First, they have shown that adult dentate gyrus neurogenesis is necessary for discrimination of two similar contexts. On day 0, mice were placed in a context A where they got a foot shock (Context A: "stainless steel grids were exposed; house fan and lights were switched on; and a mild lemon scent

was used as an olfactory cue"). Then, to test fear generalization, over several days mice were subsequently placed in context A where they got a foot shock, and in a safe similar context B ("exposed stainless steel grid floor and roof (a salient feature of the context); house fan and lights were turned off; two plastic inserts were used to cover the walls; chamber door was left ajar during testing; and a mild mint scent was used as an olfactory cue"). Control mice did freeze when placed in the safe context B the first few days of training, highlighting the similarity between the two contexts, but over time they learned to discriminate the two contexts, as emphasized by decreased freezing. On the other hand, mice with ablation of neurogenesis through X-ray irradiation in dentate gyrus kept freezing in the safe context B over all 5 training days, indicating that they were not able to discriminate the two contexts (Sahay et al. (2011a)). (Similarly, Tronel et al. (2012) observed that mice with reduced dentate gyrus neurogenesis through overexpression of Bax, a proapoptotic protein that induces programmed cell death, were impaired in the discrimination of two contexts with overlapping features using a contextual fear conditioning task.) Second, adopting the opposite procedure, Sahay et al. (2011a) have enhanced adult dentate gyrus neurogenesis by increasing the survival of newborn DGCs through conditional knockout of Bax in dentate gyrus neural stem cells. As in the previous experiment, mice were placed on day 0 in context A where they received a foot shock. Then, for several days mice were consecutively placed in context A where they always got a foot shock, and in a safe similar context B (in random order every day) to assess fear generalization. On day 1, control mice and mice with enhanced dentate gyrus neurogenesis froze equally in both contexts. However, on days 2 to 9, mice with enhanced dentate gyrus neurogenesis showed higher discrimination abilities between the two similar contexts, reflected by lower levels of freezing in the safe context B than in context A. Finally, Sahay et al. (2011a) have emphasized the fact that adult dentate gyrus neurogenesis promotes pattern separation when contexts are similar, but not if contexts are distinct. On day 0, mice were placed in context A where they got a foot shock. To measure freezing, on day 1 they were placed again in context A without getting a foot shock, and on day 2 they were placed in a safe distinct context C ("stainless steel grid floor was covered with a plastic panel and cage bedding; house fan and lights were turned off; chamber walls were covered using plastic inserts; chamber door was left ajar during testing; and a mild anise scent was used as an olfactory cue"). During testing, both control mice and mice with enhanced dentate gyrus neurogenesis showed equivalent freezing in context A, and lack of freezing in distinct context C.

2.3.4 Links to cognition

The link between adult dentate gyrus neurogenesis and cognition has been thoroughly reviewed (Ming and Song (2005); Anacker and Hen (2017)). Briefly, brain injuries and pathological conditions, such as epileptic seizures and degenerative neurological diseases, frequently lead to increased proliferation in neurogenic regions, and sometimes even in regions where neurogenesis is negligible in normal conditions. Furthermore, adult neurogenesis has been extensively linked to chronic stress with various origins. Rodents and nun-human primates, but also patients with depression, exhibit reduced levels of neurogenesis. Accordingly, an-

Table 2.1 – Summary of experimental results addressing the functional role of adult dentate gyrus neurogenesis

	Hippocampus-dependent task	Manipulation of neurogenesis	Main results
Shors et al. (2001)	Trace eyeblink conditioning	Reduction by subcutaneous injection of MAM	Impaired acquisition vs. controls; Incorporation of 2-week (not 1-week) old DGCs in memory
Shors et al. (2002)	Cued fear conditioning	"	Impaired acquisition vs. controls
	Contextual fear conditioning; Exploration in elevated plus maze; Morris water maze	"	Same acquisition as controls
Snyder et al. (2005)	Morris water maze	Reduction by whole brain irradiation	Same acquisition as controls; 4- to 28-day old DGCs required for recall at 2 and 4 weeks (but not 1 week)
Kee et al. (2007)	Morris water maze	Control	1- and 2-week old DGCs not integrated; 4-, 6- and 8-week old DGCs integrated and recruited during recall
Gu et al. (2012)	Morris water maze; Contextual fear conditioning	Optogenetic silencing	4-week (but not 2- or 8-week) old DGCs necessary for recall; 4-week old DGCs not necessary for acquisition
Jessberger et al. (2009)	Morris water maze	Reduction by WNT knockdown	Impaired recall vs. controls; Same acquisition as controls
	Novel object recognition	"	Impaired recall at 3h and 4 weeks (but not 1min) vs. controls
Clelland et al. (2009)	Delayed non-matching to place radial 8-arm maze; Spatial discrimination in the touch screen	Reduction by WNT knockdown or focal X-ray irradiation	Impaired for low (but not high) spatial separation vs. controls
	Paired associates in the touch screen	"	Same performance as controls
Vivar et al. (2012)	Spatial discrimination in the touch screen	Injection of agonist of NMDA receptors in LEC	Impaired for low (but not high) spatial separation vs. controls
Sahay et al. (2011a)	Contextual fear conditioning	Reduction by focal X-ray irradiation	Impaired for similar contexts vs. controls
	Contextual fear conditioning	Increase by Bax knockout	Better for similar (but not distinct) contexts than controls
Tronel et al. (2012)	Contextual fear conditioning	Reduction by Bax overexpression	Impaired for similar contexts vs. controls

tidepressant drugs as well as non-pharmacological antidepressants such as running increase neurogenesis. It has been proposed that enhanced neurogenesis might decrease depressive symptoms by promoting a better separation between an experienced stressful event and other similar innocuous events, thus decreasing fear generalization.

2.4 Modeling of adult dentate gyrus neurogenesis

Long before adult neurogenesis was accepted by the neuroscience community, investigations in artificial neural networks (ANN) had hinted to a beneficial effect of the introduction of new nodes for encoding of novel information. For example, similar to earlier work of Carpenter and Grossberg (1988) (also reviewed in Hertz et al. (1991)), Platt (1991) had proposed a network that incorporates a new node when an input pattern is badly represented by the network because it is too distinct from already stored input patterns. The new node is allocated directly with a selectivity for the novel input pattern. It is conceptually similar to a model of additive neurogenesis. A few years later, Yingwei et al. (1997) proposed a similar approach, where the incorporation of new nodes is counterbalanced by the pruning of nodes that do not contribute much to network computation, hence conceptually similar to replacement neurogenesis. Simultaneously, building on the neural gas theory (Martinetz and Schulten (1991)), Fritzke (1995) implemented an incremental network ("growing" neural gas) which learns topological relations between input patterns.

Since adult dentate gyrus neurogenesis is now well established, its putative role has been investigated by models. In particular, ANN representing the hippocampal network with various levels of abstraction have been used. I summarize here some modeling papers, which present the main views on the functional role of adult-born dentate granule cells.

2.4.1 Clearance of old memories

Chambers et al. (2004) presented one of the first modeling studies, suggesting that adult-born dentate granule cells solve the stability-plasticity dilemma by promoting learning of new representations while still mostly preserving old memories. They used a 3-layer feedforward network with 35 input units (representing EC), 10 hidden units (representing dentate gyrus) and 26 output units, see Figure 2.3. The input units have binary activity corresponding to the 2-dimensional visualizations of letters. The hidden and output units have a nonlinear gain function with a bias. All feedforward weights are initialized to small random (positive and negative) values. First, the 26 letters of the Roman alphabet are presented to the network, and the feedforward weights and biases are learned through gradient descent using the backpropagation algorithm. It is a supervised learning algorithm: the activity of the output units is compared to the one-hot representation of the presented letter (for ex. if letter 'B' is presented, the second unit of the teaching signal has value 1 while all others have value 0, see Figure 2.3), and the error is backpropagated. Hence the feedforward weights are updated in such a way that the difference between output and teaching signal is minimized. To model

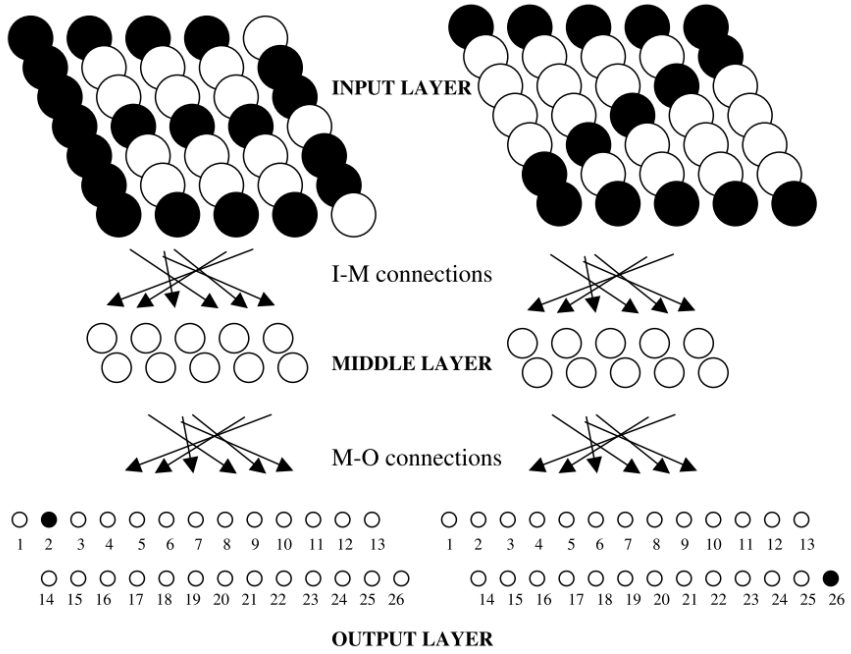

Figure 2.3 – Network for Chambers et al. (2004). Letters are presented to the network by setting the binary activity of the 35 input layer units accordingly. The activity is propagated to 10 middle layer (hidden) units, which in turn propagate the activity to 26 output layer units. Turnover neurogenesis is modeled by replacing some middle layer units, as well as their incoming and outgoing connections, by as many naive units with random connections. (Adapted from Figure 1 of Chambers et al. (2004) with permission.)

turnover neurogenesis, at the end of the learning of the Roman alphabet, between 0 and 10 randomly selected hidden units are replaced by naive hidden units (initialized to small random values). As expected, the authors found that the more hidden units are replaced by naive units, the larger the degradation of the memory for the previously learned Roman alphabet. Then the 24 letters of the Greek alphabet were presented to the network, and all weights (including the ones from units that were not replaced) were learned in the same way as before. Consistent with the previous result, the more hidden units were replaced by naive units, the better the novel Greek alphabet was learned. However, surprisingly, the authors found that recall for the Roman alphabet remained at a good performance level for the networks with high turnover (between 5 and 10 hidden units), but was significantly worse for the networks with low turnover (0 and 2 hidden units). They hypothesized that it was due to a predominant effect of catastrophic interference at low turnover rate due to the high similarity between the two data sets. In addition, it was shown that if the hidden units with larger input and output weight change (during Roman alphabet learning) were selected to be replaced by naive units, learning of the novel Greek alphabet was improved.

Simultaneously, Deisseroth et al. (2004) reached similar conclusions. Their 3-layer network contained 500 units in each layer, and excitatory all-to-all connections, see Figure 2.4a. Units

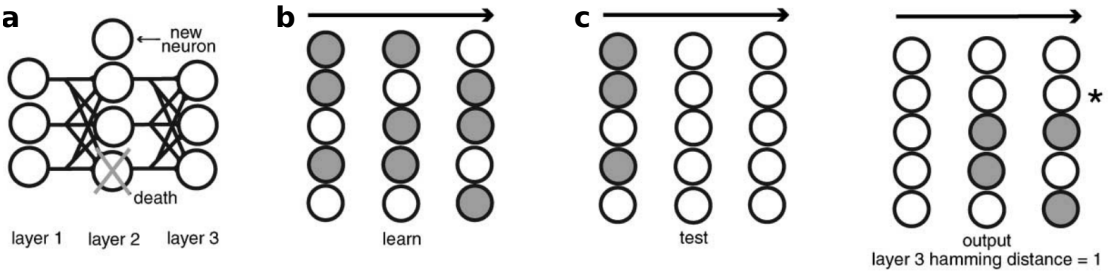

Figure 2.4 – Network for Deisseroth et al. (2004). (**a**) The 3-layer network contains 500 units in each layer, which are fully feedforwardly connected. Turnover neurogenesis is modeled by replacing some hidden layer units with as many new naive units with random connections. (**b**) During learning, the activity of all units of the network are clamped to binary values, and all connections are learned through a batch Hebbian plasticity rule. (**c**) During testing, only the activity of the input layer units is set, and it is propagated through the feedforward connections to the hidden and output units. Recall performance is assessed by computing the Hamming distance between the output layer activity and its corresponding activity during learning. (Adapted from Figure 8 of Deisseroth et al. (2004) with permission.)

had binary activity, and their activity threshold was set up in such a way that a desired sparsity per layer was achieved, effectively corresponding to k-WTA. During learning, the activity of all units of the network were clamped, see Figure 2.4b. Connections were learned according to a batch Hebbian plasticity rule (taking into account all input patterns). Then, the units of the first layer only were set to the activity of a given pattern, and activity was propagated to the output layer through the hidden layer, see Figure 2.4c. Recall performance was assessed by computing the Hamming distance between the obtained output pattern and the correct activity in the output layer (the one that was used during learning, compare output activity in panels b and c of Figure 2.4). Turnover neurogenesis was implemented in the hidden layer by deleting some units and replacing them with new naive units that had to learn their connections from scratch. As expected, turnover neurogenesis induces more rapid clearance of old memories than a control network without neurogenesis, as connections that were learned during previous memory storage are lost along their corresponding hidden units. On the other hand, turnover neurogenesis improves recall of newer memories, and this beneficial effect correlates with the amount of new memories that have to be stored in the network. Interestingly, the authors found an equivalent increase in neurogenesis as a function of activity in a hippocampal cell culture (inferred from extracellular Ca^{2+} concentration).

The paper of Chambers et al. (2004) triggered several follow-up studies with increased biological plausibility. Crick and Miranker (2006) implicitly modeled interneurons through full lateral inhibitory connections within the hidden and the output layers. They also used a more biologically plausible learning rule than backpropagation: unsupervised Hebbian learning. Feedforward (resp. lateral) weights were forced to be positive (resp. negative), and bounded to avoid runaway dynamics of the weights. Chambers and Conroy (2007) used the same network with gradient descent backpropagation for learning of the connections and investigated the

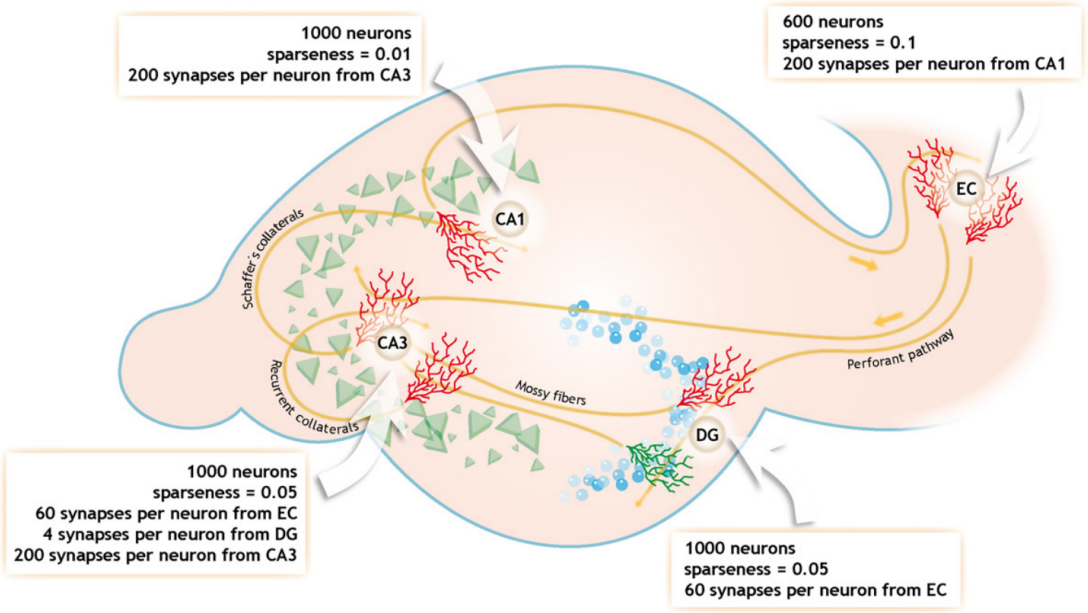

Figure 2.5 – Network for Weisz and Argibay (2009) and Weisz and Argibay (2012). The network models the full hippocampal loop: EC, DG, CA3 and CA1. For each population, the number of neurons, the enforced sparsity, and the number of incoming synapses per neuron is written on the figure. Additive neurogenesis is modeled by adding new neurons to the DG population. (Reproduced with permission from Figure 1 of Weisz and Argibay (2012).)

beneficial effect of turnover neurogenesis to learn a second alphabet after having learned the Roman alphabet. The new alphabet had varying levels of similarity with the previously learned alphabet: "Romalt" (Roman alphabet with slightly different writing style) > Russian > Hebrew. The higher the degree of the novelty of the new alphabet, the more beneficial was a large extent of turnover neurogenesis for the network to accurately learn it.

Using a more complex network representing the whole hippocampal loop and additive neurogenesis in dentate gyrus (DG), Weisz and Argibay (Weisz and Argibay (2009, 2012)) reached the same conclusion that adult-born DGCs promote learning of novel input while mostly preserving recall of old memories. The network consists of four populations of binary neurons: EC, DG, CA3 and CA1, see Figure 2.5. EC projects to DG and CA3, DG projects to CA3, CA3 projects to CA1 and to itself through recurrent connections, and CA1 projects back to EC. The number of synapses at each stage is fixed, and k-WTA is implemented in each population by enforcing a fixed number of neurons to be active. All connections weights are randomly initialized. The connections between EC and DG, and between CA1 and EC, follow a Hebbian learning rule with an heterosynaptic term to ensure weight normalization. It is different from Oja's rule because the heterosynaptic term depends on the postsynaptic activity, and not the postsynaptic activity squared. The recurrent connections within CA3, and the connections between EC and CA3, follow a covariance rule. All other synapses are fixed to a given value. Interestingly, activity of CA3 units depends only on input from DG during learning of inputs,

and only on direct input from EC and recurrent input from CA3 during retrieval (the recurrent collateral loop is repeated a fixed number of times). After learning of a few inputs, the (control) network was compared with the same network undergoing additive neurogenesis in DG in their ability to learn new inputs and recall old inputs. The neurogenesis network kept the same level of sparsity in DG, and the connections to and from adult-born DGCs were randomly initialized. The learning rules were kept identical to those of the control network, except that the synapses from newborn DGCs to CA3 have a larger learning rate than the ones from mature DGCs. As expected due to the larger number of DGCs, the network with neurogenesis had a bigger storage capacity than the control network. It also had a better recall accuracy for recent inputs, due to the fact that additional newborn DGCs preferentially learn new patterns. This is emphasized by their higher activity for recent patterns than for remotely stored patterns. On the other hand, the network with neurogenesis had a slightly higher tendency to forget old memories than the control network, probably because plasticity of mature DGCs was not completely abolished. These results, obtained with a full hippocampal model, are in agreement with previous studies adopting simpler models (Chambers et al. (2004); Deisseroth et al. (2004)). However, remarkably, the authors observed that adult-born DGCs did not promote pattern separation of EC input patterns. They rather increase the overlap between representations of recent memories, and provide temporal separation of remote and recent memories. It stems from the fact that newborn DGCs mainly encode recent memories (the ones that were presented to the network during their enhanced plasticity phase). Hence, they rather combine all representations of new inputs. This last observation has been further studied in Aimone et al. (2006, 2009), see Section 2.4.5.

2.4.2 Encoding distinct memories of highly similar inputs

Becker (2005) proposed that adult-born DGCs encode distinct memories for highly similar inputs. The network models the full hippocampal loop, set up as an autoencoder: the EC input layer connects to DG, which connects to CA3 (modeled with recurrent connections), which connects to CA1, which connects to the EC output layer, see Figure 2.6. The EC input layer provides a direct teaching signal to the EC output layer, as well as a direct projection to CA3 and CA1. The main assumption of the model is that all hippocampal areas are optimizing a common objective function, namely the faithful linear reconstruction of the input. The biological constraint of sparse activity in DG is enforced by an artificial k-WTA network. Interneurons are not modeled, so the k DG units with largest input are active while the activity of the others was set to zero. Encoding and retrieval rely differently on the hippocampal areas. Most notably, DG units are active only during encoding, and do not engage significantly in retrieval: they do not drive CA3 activation, rather it is the direct connection from EC which does. With all these assumptions and constraints on the circuit, Becker obtains Hebbian-style learning rules. However, to avoid runaway dynamics of the weights, normalization of all weight vectors is imposed by weight clipping. Neurogenesis is modeled in the DG layer, and recall of random binary input patterns is assessed. Two data sets are used: one where all patterns are independent (hence distinct), and another where patterns are highly overlapping

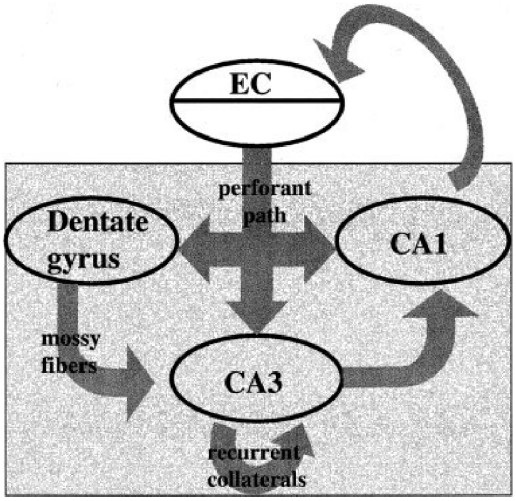

Figure 2.6 – Network for Becker (2005). The network models the full hippocampal loop: EC, DG, CA3 and CA1. It is set up as an autoencoder: the EC population is separated into an input and an output population of same size. Neurogenesis is modeled by either adding new units to the DG population, or replacing some of the existing units. (Reproduced with permission from Figure 1 of Becker (2005).)

(hence highly similar). Additive neurogenesis is shown to promote recall of both sets of input patterns, with better recall for more units added in the DG. On the other hand, replacement, or turnover, neurogenesis only promotes recall for the data set with highly similar patterns.

2.4.3 Avoidance of catastrophic interference

Wiskott et al. (2006) have proposed that adult neurogenesis could be a solution to avoid catastrophic interference in the dentate gyrus. They used an abstract model consisting of an input layer (representing layer II of EC), which is connected to an hidden layer (corresponding to DG/CA3), which is in turn connected to an output layer (layer V/VI of EC), see Figure 2.7. It is thus a reduced and simplified version of the circuit in Becker (2005). The network is fully feed-forward with linear units, hence it has no particular dynamics and can be solved analytically (as described in Wiskott et al. (2004)). The network is set as an autoencoder: the output layer contains as many neurons N as the input layer and aims to reconstruct the input as best as possible. The hidden layer contains M units with $M < N$, so the input has to be compressed in an optimal way to keep most of the information. The weight matrices for encoding (from input to hidden) and decoding (from hidden to output) can be analytically computed such that they minimize the reconstruction error of the input. The authors considered inputs coming from two distributions, A and B. Distribution B is a rotated version of distribution A (only the orientation in space is different, the eigenvalues of the covariance matrix are identical). No synaptic plasticity rule is involved: the decoding weights are always set to the analytically computed optimal values which correspond to the environment to which the network is

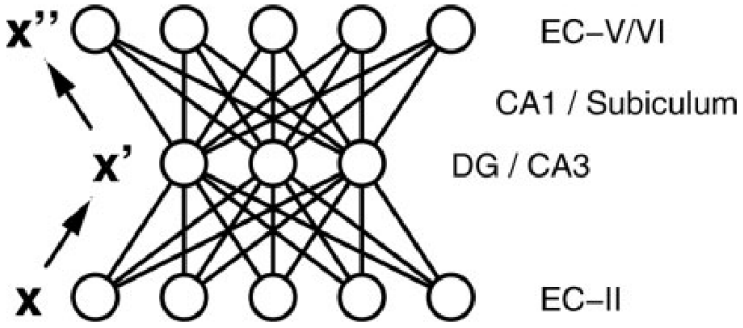

Figure 2.7 – Network for Wiskott et al. (2006) and Appleby and Wiskott (2009). The network is an autoencoder with an input layer (representing EC layer II) fully feedforwardly connected to a hidden layer (corresponding to DG/CA3), which is fully feedforwardly connected to the output layer (representing EC layer V/VI). Neurogenesis is modeled by adding units to the hidden layer or replacing some existing units by new units. (Reproduced with permission from Figure 1 of Wiskott et al. (2006).)

"adapted" (meaning from which the current input patterns are drawn). The encoding weights are either set to the analytically computed optimal values from the same, or the other environment. The authors first show that an input from distribution A is not well reconstructed when B-decoding weights are used, no matter if encoding weights are set to A- or B-weights. They then show that if an additive model of neurogenesis is used, the reconstruction error is lower. In that case, they start with a network containing $M-L$ hidden units, and whose encoding weights are fixed to A-weights. Then L units are introduced in the hidden layer, and their encoding weights are set to B-weights (which are in addition set orthogonal to the encoding weights of the first $M-L$ hidden units). The network with neurogenesis reconstructs better patterns drawn from any of the two distributions than the control network. Hence, this model suggests that adult neurogenesis provides a way to represent novel features of the input space, which are orthogonal to already stored features.

To make the model more biologically plausible, an extended model was presented in a follow-up paper (Appleby and Wiskott (2009)). The main differences with the initial model are that: (1) there is an expansion from EC to DG, meaning that $M > N$, in agreement with experimental findings (Andersen et al. (2007)), (2) DG units are nonlinear, the DG layer is modeled as an artificial single WTA network analogous to Becker (2005), and (3) besides additive neurogenesis, turnover neurogenesis is modeled as well. In both cases, the total number of DG units at the end of learning is bounded to M. In the turnover case, the DG units that are replaced are randomly chosen. Similarly to the earlier paper, all weights are analytically computed such that they minimize input reconstruction error. Additive neurogenesis appears to be superior to turnover neurogenesis, as it permits learning of the input statistics of the second, novel, environment while preserving accurate retrieval of inputs stemming from the first environment. These results were confirmed in another study with realistic inputs, spatially organized like grid cells (Appleby et al. (2011)).

The view of Wiskott et al. (2006) is similar to the one of Becker (2005), as both suggest that adult-born DGCs encode memory traces of novel inputs, thus promoting recall. However, Becker (2005) hypothesizes that DG does not play any role during recall, which conflicts experimental findings demonstrating that adult-born DGCs are required for recall (Kee et al. (2007); Gu et al. (2012)). In addition, both Becker (2005) and Wiskott et al. (2006) propose a beneficial effect of adult-born DGCs in recall no matter the similarity between inputs. This stands against behavioral studies that have manipulated DG neurogenesis and shown a beneficial effect of newborn DGCs for discriminating similar patterns, but not distinct patterns (Clelland et al. (2009); Sahay et al. (2011a)).

2.4.4 Input discrimination

Consistent with the idea that adult-born DGCs are involved in encoding of novel inputs, Temprana et al. (2015) further proposed that inhibition from local interneurons is crucial for shaping small non-overlapping receptive fields spanning the (novel) input space. They supported their idea with a simple model. EC input units are connected to DG units through feedforward weights that are randomly initialized to nonzero values. Inputs are presented to the network by setting the activity of the EC units accordingly. Activity is propagated to the DG units, whose nonlinear activity depends on a value of inhibition which is set based on the overall DG activity. Hence a soft WTA is implemented even though interneurons are not explicitly modeled. For each input presentation, the feedforward weights are learned according to a Hebbian rule, and the weight vectors are normalized at each time step to avoid runaway dynamics of the weights. The authors showed that the level of inhibition during learning lead to different outcomes concerning the receptive fields of the DG units. With low inhibition, DG units develop wide receptive fields which are overlapping between units. With high inhibition, most DG units are always silent (nonexistent receptive fields) and the whole input space is represented by a few units with large, but non-overlapping, fields. The best outcome occurs when inhibition is gradually enhanced from a low to a high value, thus giving rise to DG units with small non-overlapping fields. Hence different inputs can be discriminated.

2.4.5 Temporal separation

Building on the observation that adult-born DGCs have enhanced excitability early in their maturation, Aimone et al. (2006) postulated that they rather deteriorate the pattern separation function of the dentate gyrus, because they are active for all temporally proximal events, irrespective of the extent of their contextual similarity. Hence, they suggested that mature DGCs are responsible for pattern separation, while newborn DGCs promote temporal pattern integration by associating temporally proximal events (in a timescale of hours/days). They implemented a complex model to support this idea (Aimone et al. (2009)). It consists in a spiking network made of 6 neuronal populations of conductance-based model neurons. Lateral EC and medial EC populations represent the contextual and spatial input, respectively.

They are feedforwardly connected to DGCs and inhibitory Basket cells, which are recurrently connected with excitatory Mossy cells and inhibitory Hilar cells. The only plastic connections are the excitatory connections to DGCs, namely from lateral EC, medial EC and Mossy cells. The rate of weight change is determined by the age of the synapses. Synapses are formed in an age- and experience-dependent manner, and initialized at random strengths. They are learned according to a plasticity rule which has a Hebbian component, but which also depends on a preset rate of synapse maturation. In addition, connection weights are hard bounded to minimal and maximal values, which depend on the maturation stage of the corresponding synapses. A network with additive neurogenesis in the DGC population was compared to a control network without neurogenesis. For any similarity between EC input patterns (=events), when events where temporally proximal, the DGC population of the control network had better pattern separation abilities than the DGC population of the network with neurogenesis. Conversely, when events were temporally distant, the higher the rate of neurogenesis, the better the pattern separation in the DGC population. Furthermore, for temporally proximal events, the extent of temporal pattern integration negatively correlates with the similarity between EC firing patterns, indicating that adult-born DGCs particularly promote pattern integration for very distinct input patterns.

2.5 Open questions

In this work, I address some of the above puzzling observations and investigate the following two questions: First, what is the functional role of the switch from lateral excitation to lateral inhibition in adult DG neurogenesis? And second, why are newborn DGCs only relevant for novel stimuli having a high degree of similarity with old ones?

While experimental results suggest that the switch from excitation to inhibition of the GABAergic input onto adult-born DGCs is crucial for their integration into the preexisting circuit (Ge et al. (2006)), and for proper behavior (Furukawa et al. (2017)), it remains unclear why such a link between channel properties and behavior arises. To my knowledge, none of the previous modeling studies ever took this mechanism into account to examine if it could have an effect on integration of newborn DGCs into a preexisting network. Temprana et al. (2015) did show with a model that a gradual change of inhibition from low to high levels enhances the fine-grain discrimination of novel inputs, but they used a simplified model where newborn DGCs were born already connected with random weights to the preexisting circuit. The model in Temprana et al. (2015), as all the other models I reviewed, did not address the question of how newborn DGCs grow connections to neighboring neurons. They bypassed this aspect by simply setting random weights to those connections.

Besides, experimental observations reveal that adult dentate gyrus neurogenesis improves behavioral pattern separation performance only if the stimuli that have to be discriminated are similar, but not if they are distinct (Clelland et al. (2009); Sahay et al. (2011a)). However, it is still unclear how adult-born DGCs impact pattern separation, if it is through a cell-autonomous

function as individual encoding units, or through a modulatory role (Sahay et al. (2011b); Aimone et al. (2011)). Becker (2005) already showed with a model that newborn DGCs could encode distinct memory traces for highly similar inputs, hence promoting pattern discrimination of similar inputs. However, the same model promoted a comparable improvement of pattern separation of distinct inputs with more newborn DGCs. Therefore, it cannot explain the differential behavioral outcomes for similar versus distinct stimuli. To my knowledge, all previous models showed a beneficial effect of dentate gyrus neurogenesis, be it through an additive or a turnover process, by promoting input discrimination and avoidance of catastrophic interference. Yet, none of them elucidates the different outcomes that are observed behaviorally when inputs that have to be discriminated are similar or distinct, and none of them connects the GABAergic switch from excitation to inhibition with the network function of adult-born DGCs.

3 Classification using competitive networks

3.1 Introduction

Unsupervised competitive learning is widely used in artificial intelligence to perform clustering of input patterns into a few categories, thus providing generalization or compression of the input (reviewed in Grossberg (1987b); Hertz et al. (1991); Du (2010)). In these types of networks, an input layer is fully feedforwardly connected to an output layer through plastic excitatory connections (Figure 3.1). The "competitive" aspect stems from the recurrent inhibitory connections between output units. They ensure that when an output unit is highly activated, it silences activity of other output units. In some instances of competitive networks, output units self-excite themselves.

The basic idea is that upon presentation of several patterns, the competitive network unsupervisedly learn to cluster input into different categories. After learning of the feedforward connections, input patterns that are similar to each others activate the same output unit, whereas input patterns that are different from each other activate different output units. Therefore each output unit is said to be selective for a category, or cluster, of input patterns, and the input patterns can ultimately be classified into several categories.

The simplest form of a competitive network is the single winner-take-all (WTA) network. As its name suggests, for any input pattern presentation, a single output unit is active, and it silences all the other output units through the recurrent inhibitory connections. Other forms of competitive networks exist, where ensembles of output units get activated for a given input pattern. Hence classification is distributed over the output units, as different ensemble of output units are selective for different categories. If the output units have a step gain function (e.g. activity is either 0 or 1) and the number of active output units is set to k for any input pattern, the network is called k-WTA. On the other hand, if the output gain function is continuous (e.g. hyperbolic tangent, which is nonlinear and bounded but differentiable), the number of output units in each ensemble cannot be easily determined, so it is called a soft WTA network.

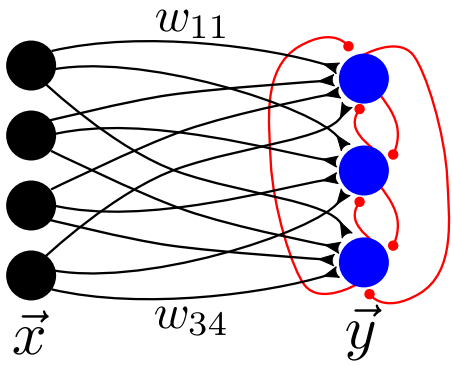

Figure 3.1 – Simple competitive network. An input layer, with firing rate vector $\vec{x}$ of input units (black, firing rate of single unit is x_j) has all-to-all feedforward connections with an output layer, with firing rate vector $\vec{y}$ of output units (blue, firing rate of single unit is y_i) which have all-to-all recurrent inhibitory connections (red). Connections with a triangular end (black) are excitatory, connections with a round end (red) are inhibitory. w_{ij} denotes the connection weight from unit j to unit i.

In single WTA settings, the maximum number of categories that can be classified is given by the number of output units in the network. Therefore it is good to know beforehand how many categories are present in the input data to build the network architecture accordingly. In some instances, due to random initialization of the feedforward weight vectors onto the output units, some of the output units might never win the competition (remain dead units), so the network clusters the input patterns in fewer categories. Instead, in soft WTA settings, there is no need to know how many categories are present in the input data to design the network, because an ensemble of units wins for each input pattern. It is thus more modular.

3.2 Importance of normalization

During presentation of a pattern μ, the activity of the input layer is given by: $\vec{x}^\mu = \left(x_1^\mu, x_2^\mu, ..., x_N^\mu\right)$, with $x_j^\mu \geqslant 0 \; \forall \, j$, where N is the number of input units. For the patterns to be comparable, they should be normalized to the same L2-norm:

$$||\vec{x}^\mu|| = \sqrt{\sum_{j=1}^{N} \left(x_j^\mu\right)^2} = \sqrt{L} \quad \forall \mu \tag{3.1}$$

with $L > 0$.

We note here that in the particular case where the input patterns are binary (elements are 0 or 1), L1-normalization ensures L2-normalization. Indeed, the L1-norm is given by:

$$||\vec{x}^\mu||_1 = \sum_{j=1}^{N} |x_j^\mu| = L \quad \forall \mu \tag{3.2}$$

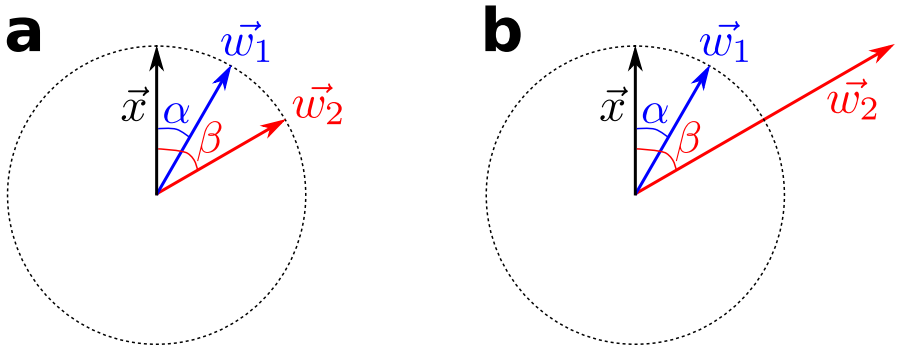

Figure 3.2 – Importance of weight vector normalization. The input pattern ($\vec{x}$, black), the feedforward weight vector onto a first output unit ($\vec{w}_1$, blue), and the feedforward weight vector onto a second output unit ($\vec{w}_2$, red) are represented, as well as the angle α (resp. β) between $\vec{x}$ and $\vec{w}_1$ (resp. $\vec{w}_2$). Vectors with arrow end on the dotted circle have the same norm. (**a**) Case where the weight vectors are normalized. (**b**) Case where the weight vectors are not normalized: the red arrow is much longer than the blue arrow. Note that the input pattern does not have to have the same norm as the weight vectors for the argument.

with L an integer positive value which denotes the number of active input units, then the L2-norm of the input patterns is also identical ($||\vec{x}^{\mu}|| = \sqrt{L}\,\forall\mu$).

The activity of any output unit i upon presentation of pattern $\vec{x}$ is defined as:

$$y_i = f\left(\sum_{j=1}^{N} w_{ij}x_j\right) = f\left(\vec{w}_i \cdot \vec{x}\right) \tag{3.3}$$

with f a gain function which is monotonously increasing, w_{ij} the connection weight between input unit j and output unit i, and $\vec{w}_i$ the feedforward weight vector onto output unit i (the i^{th} row of the connectivity matrix between input and output layers).

In a single WTA setting the output unit which wins the competition is, according to equation (3.3), the one which has the largest dot product between its feedforward weight vector $\vec{w}_i$ and the input pattern $\vec{x}$. Therefore, it is crucial that all feedforward weight vectors are bounded to the same norm to ensure fair competition between output units. Indeed, if an input pattern is presented to the network, we would like that the output unit whose feedforward weight vector aligns best with the input pattern wins the competition. The norm of the feedforward weight vectors does not need to be equal to the one of the input patterns, though.

Figure 3.2 provides an illustrative example. We have $\alpha < \beta$, hence the output unit receiving the blue $\vec{w}_1$ feedforward weight vector should intuitively win the competition because it is closer to $\vec{x}$ than the red $\vec{w}_2$ feedforward weight vector. However, according to equation (3.3) the unit

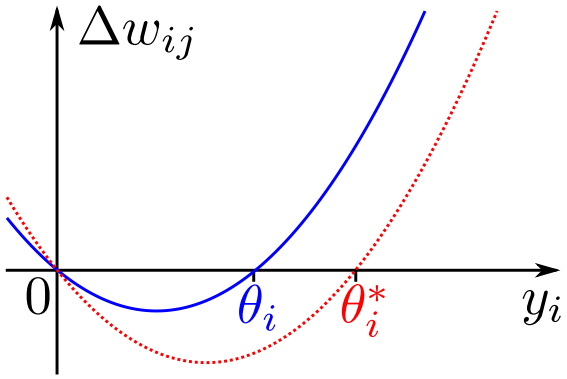

Figure 3.3 – BCM learning rule. The weight update Δw_{ij} as a function of the postsynaptic activity y_i is plotted for a given value of the presynaptic activity x_j and two different neuronal activity thresholds θ_i and θ_i^*. When the postsynaptic activity is smaller than the activity threshold ($y_i < \theta_i$) the weights undergo LTD, otherwise ($y_i > \theta_i$) they undergo LTP. The maximum LTD value is $-\frac{\theta_i}{4}$, hence the blue curve shows smaller LTD updates than the red dotted curve because $\theta_i < \theta_i^*$.

with $\vec{w}_1$ wins only if $\vec{w}_1 \cdot \vec{x} > \vec{w}_2 \cdot \vec{x}$. Using the definition of the dot product, we have:

$$\vec{w}_1 \cdot \vec{x} = ||\vec{w}_1|| \, ||\vec{x}|| \cos(\alpha) \tag{3.4}$$

$$\vec{w}_2 \cdot \vec{x} = ||\vec{w}_2|| \, ||\vec{x}|| \cos(\beta) \tag{3.5}$$

Without loss of generality, let's assume that $||\vec{x}|| = ||\vec{w}_1|| = 1$. Therefore, we obtain:

$$\vec{w}_1 \cdot \vec{x} = \cos(\alpha) \tag{3.6}$$

$$\vec{w}_2 \cdot \vec{x} = ||\vec{w}_2|| \cos(\beta) \tag{3.7}$$

Here, it is clear that if $||\vec{w}_2|| = ||\vec{w}_1||$ (normalized case) the first output unit wins the competition, as it aligns best with the input pattern ($\cos(\alpha) > \cos(\beta)$ because $\alpha < \beta$), see Figure 3.2a. However, if $||\vec{w}_2|| > ||\vec{w}_1||$ (non-normalized case), it is possible that the second output unit wins the competition, even though its feedforward weight vector is farther away from the input pattern than the one from the first output unit, see Figure 3.2b. For example if $\beta = 2\alpha = \frac{\pi}{3}$, it is sufficient that $||\vec{w}_2|| > \sqrt{3}$ for the second output unit to win the competition (because then $\vec{w}_1 \cdot \vec{x} = \cos(\alpha) = \frac{\sqrt{3}}{2} < \vec{w}_2 \cdot \vec{x} = ||\vec{w}_2|| \cos(\beta) = \frac{1}{2}||\vec{w}_2||$).

3.2.1 Example: Runaway dynamics of the weights with the BCM learning rule

Here we illustrate the importance of proper normalization of the feedforward weight vectors with a learning rule which is frequently used for biologically plausible networks: the BCM learning rule (Bienenstock et al. (1982)), see Figure 3.3:

$$\Delta w_{ij} = y_i (y_i - \theta_i) x_j - \epsilon w_{ij} \tag{3.8}$$

with y_i the activity of the output unit i, x_j the activity of the input unit j, ϵ a scalar that determines the amplitude of weight decay, and θ_i a neuronal activity threshold. The BCM rule is biologically plausible, because it is local (e.g. it uses information available at the synapse: the presynaptic and postsynaptic activity) and can be implemented online. In addition, it contains a "triplet" term ($y_i^2 x_j$) which has been shown to be needed to fit experimental data (Pfister and Gerstner (2006)).

Let us focus on a neuron i with a single input j, and a monotonically increasing function f in equation (3.3). The derivative of equation (3.8) with respect to the postsynaptic activity is given by:

$$\frac{\mathrm{d}\left(\Delta w_{ij}(y_i)\right)}{\mathrm{d}y_i} = x_j\left(2y_i - \theta_i\right) \tag{3.9}$$

If ϵ is small enough, the last term in equation (3.8) is negligible, hence the fixed points and their stabilities are (keeping in mind that $x_j \geqslant 0\ \forall j$, and $\theta_i \geqslant 0\ \forall i$):

- $y_i = 0$: Stable fixed point because

$$\left.\frac{\mathrm{d}\left(\Delta w_{ij}(y_i)\right)}{\mathrm{d}y_i}\right|_{y_i=0} = -x_j\theta_i < 0 \tag{3.10}$$

- $y_i = \theta_i$: Unstable fixed point because

$$\left.\frac{\mathrm{d}\left(\Delta w_{ij}(y_i)\right)}{\mathrm{d}y_i}\right|_{y_i=\theta_i} = x_j\theta_i > 0 \tag{3.11}$$

so that w_{ij} increases and y_i increases as well because of equation (3.3). The limitations of these qualitative arguments have been discussed in Zenke et al. (2015) (Turrigiano and Nelson (2000); Vitureira et al. (2012)).

The neuronal activity threshold θ_i is frequently defined as a sliding threshold, which depends on the average postsynaptic firing activity $\langle y_i \rangle$. Because θ_i depends on the activity history of the postsynaptic neuron, it provides a solution to the stability-plasticity dilemma in neuronal networks. Indeed, if a postsynaptic neuron remains silent for some time, its activity threshold decays to 0. As the activity threshold gets lower, the postsynaptic neuron has higher chance to have an activity $y_i > \theta_i$, hence increase its weights ($\Delta w_{ij} > 0$ for $y_i > \theta_i$). Conversely, if a postsynaptic neuron is highly active ($y_i > \theta_i$) for some time, its sliding threshold increases, making it harder for it to be active in the future (see Figure 3.3).

Despite its nice properties, the standard BCM learning rule is not appropriate in unsupervised competitive settings. Indeed, if all output units start with an identical L2-norm of their feedforward weight vectors, they initially compete fairly for activation. However, as long as one output unit wins the competition, its feedforward weight vector grows ($\Delta w_{ij} > 0$) to a larger L2-norm. Hence the same output unit has higher chance to win again the competition

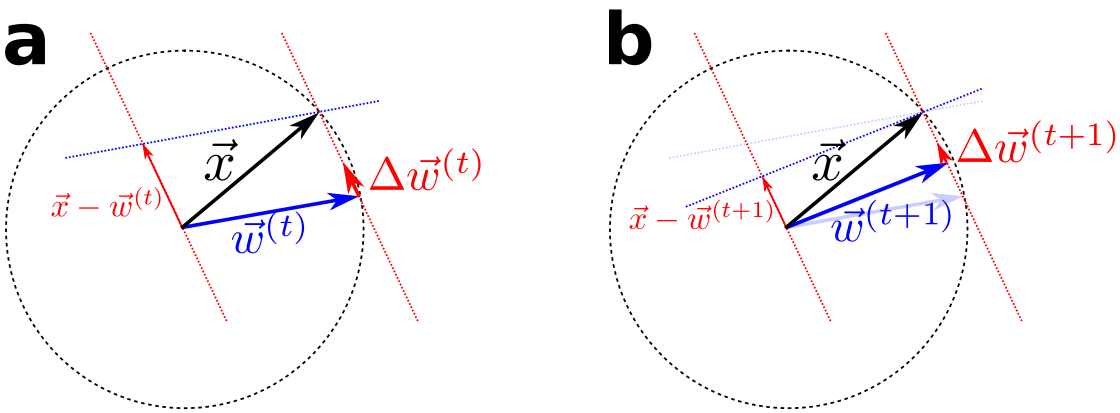

Figure 3.4 – Standard competitive learning. When a standard competitive learning rule is used, the feedforward weight vector $\vec{w}$ (blue) onto the winning output unit is updated in the direction of the input pattern ($\vec{x}$, black). (**a**) Let's assume that at time t, $||\vec{w}^{(t)}|| = ||\vec{x}|| = \sqrt{L}$. The weight update $\Delta\vec{w}^{(t)}$ moves the feedforward weight vector onto the winning output unit in the direction of the input pattern $\vec{x}$. (**b**) At time $t+1$, the feedforward weight vector, $\vec{w}^{(t+1)}$ of the winning output unit has a smaller norm, and aligns better with the input pattern. The new weight update, $\Delta\vec{w}^{(t+1)}$, is thus smaller.

for the next input pattern, therefore increasing even more its feedforward weight vector. This positive feedback loop yields to an unbounded increase of the feedforward weight vector onto one output unit. Therefore that output unit always wins the competition (Figure 3.2b), so the input patterns cannot be categorized.

3.3 Unsupervised competitive learning

At each input pattern presentation, the feedforward weight connections are learned. The output units whose feedforward weight vectors align best with the input pattern win the competition and silence all other output units through the inhibitory recurrent connections. Therefore if a Hebbian learning rule such as the BCM rule (Bienenstock et al. (1982)) is used, the winning output units are the only ones which update their weights. Here, using a single WTA setting for simplicity, we show that the feedforward weight vector onto the winning output unit moves in the direction of the presented input pattern.

3.3.1 Standard competitive learning rule

The standard competitive learning rule reads (Hertz et al. (1991)):

$$\Delta w_{ij} = \eta y_i \left(x_j - w_{ij}\right) \tag{3.12}$$

The weight update of the connection between input unit j and output unit i depends on a learning rate η, the presynaptic activity x_j, the postsynaptic activity y_i, and the current weight

of the connection w_{ij}. Only the winning output unit i^* updates its weights, as $y_i = 0 \; \forall i \neq i^*$. This rule contains a heterosynaptic term, $-\eta y_i w_{ij}$, which depends on the postsynaptic activity only and prevents runaway dynamics of the weights. This rule can also be written in a vector form:

$$\Delta \vec{w}_i = \eta y_i (\vec{x} - \vec{w}_i) \tag{3.13}$$

where $\vec{w}_i$ is the feedforward weight vector onto output neuron i.

It can be seen mathematically that this learning rule updates the feedforward weight vector onto each output unit in the direction of the input patterns for which it is the winner. Indeed, at convergence, the expectation of the weight change over all input patterns is zero: $\langle \Delta \vec{w}_i \rangle = 0 \; \forall i$. Hence, each output unit becomes selective for the center of mass of all input patterns for which it is the winner: $\vec{w}_{i^k} = \vec{x}^k_{\text{CM}}$ for the winning output unit i^k in response to patterns from cluster k, $\vec{x}^k$.

Figure 3.4 illustrates how the feedforward weight vector onto the winning output unit converges to the presented input pattern. The feedforward weight vector onto each output unit converges similarly to the input pattern it is closest to. If η is set to $\frac{1}{y_{i^*}}$ (if output units have binary activity, $y_{i^*} = c > 0 \; \forall i^*$, so $\eta = \frac{1}{c}$), the weight vector $\vec{w}$ aligns with the input pattern $\vec{x}$ in one learning step. Therefore the weight vector would jump for each pattern presentation for which it wins the competition, so the network learning would be temporally unstable (it would never reach convergence). Hence to have stable learning, η should be set small enough. Convergence of the feedforward weight vector onto a particular input pattern thus requires a long input presentation.

Noteworthy, as observable in Figure 3.4, even if all feedforward weight vectors are initialized to the same norm, during the course of learning, their norm is varying. If the environment is not well-behaved, it is possible that it poses some problem, as competition between output units is not fair anymore (Section 3.2). Other learning rules ensure that the feedforward weight vectors are always normalized to the same norm.

3.3.2 Oja's learning rule

Oja's learning rule (Oja (1982)) is a variant of unsupervised learning rules that ensures that all output units get a feedforward weight vector whose L2-norm is equal to 1 under some conditions, as will be explained below. It reads (Figure 3.5):

$$\Delta w_{ij} = \eta y_i \left(x_j - y_i w_{ij} \right) \tag{3.14}$$

The only difference with the standard competitive rule is that the heterosynaptic term, $-\eta y_i^2 w_{ij}$, depends on the square of the postsynaptic activity. Oja's learning rule is equivalent to the standard competitive learning rule when the activity of the output units is binary (0 or 1, step gain function).

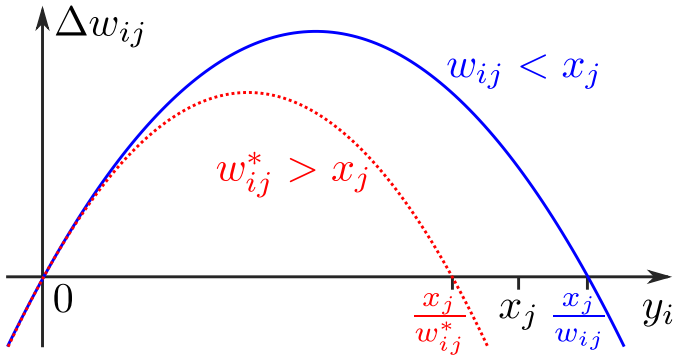

Figure 3.5 – Oja's learning rule. The weight update Δw_{ij} as a function of the postsynaptic activity y_i is plotted for a given value of the presynaptic activity x_j and two different weight values w_{ij} and w^*_{ij}. When the weight is smaller than the presynaptic activity ($w_{ij} < x_j$, blue curve) we have an LTP-dominated regime. On the contrary, when the weight is bigger than the presynaptic activity ($w^*_{ij} > x_j$, red dotted curve) we have an LTD-dominated regime. The maximum LTP value is $\frac{x_j^2}{4w_{ij}}$, hence the blue curve shows bigger LTP updates than the red dotted curve because $w_{ij} < w^*_{ij}$.

If the output units have a linear gain function, it can be shown that Oja's learning rule enforces all feedforward weight vectors to have an L2-norm of 1. Indeed, at convergence, each weight has reached a stable value, hence the expectation of the weight update over the presentations of input patterns is $\langle \Delta w_{ij} \rangle = 0 \; \forall i, j$. Let's consider a single output unit for simplicity, and thus denote by w_i the weight from input unit i to that single output unit. Then, at equilibrium:

$$\begin{aligned} 0 = \langle \Delta w_i \rangle = \langle y\left(x_i - y w_i\right) \rangle &= \langle y x_i - y^2 w_i \rangle \\ &= \langle \left(\sum_{j=1}^{N} w_j x_j \right) x_i - \left(\sum_{j=1}^{N} w_j x_j \right) \left(\sum_{k=1}^{N} w_k x_k \right) w_i \rangle \\ &= \sum_{j=1}^{N} \langle x_i x_j \rangle w_j - \left(\sum_{j=1,k=1}^{N} w_j \langle x_j x_k \rangle w_k \right) w_i \\ &= \sum_{j=1}^{N} C_{ij} w_j - \left(\sum_{j=1,k=1}^{N} w_j C_{jk} w_k \right) w_i \end{aligned} \tag{3.15}$$

where we have defined the matrix $C = \langle \vec{x}\vec{x}^T \rangle$. Its elements are given as an expectation over the input patterns: $C_{ij} = \langle x_i x_j \rangle$. Therefore it is a symmetric matrix ($C_{ij} = C_{ji}$). We now look at the expectation of the whole feedforward weight vector update by stacking the expected weight update of each of its elements ($\langle \Delta w_i \rangle$):

$$\vec{0} = \langle \Delta \vec{w} \rangle = C\vec{w} - \left(\vec{w}^T C \vec{w} \right) \vec{w} \tag{3.16}$$

It implies that:

$$\begin{aligned} C\vec{w} &= \left(\vec{w}^T C\vec{w}\right)\vec{w} \\ C\vec{w} &= \lambda\vec{w} \end{aligned} \tag{3.17}$$

with $\lambda = \vec{w}^T C\vec{w}$ a scalar. Therefore, $\vec{w}$ is an eigenvector of C, with eigenvalue λ. So we have:

$$\lambda = \vec{w}^T (C\vec{w}) = \vec{w}^T (\lambda\vec{w}) = \lambda\left(\vec{w}^T\vec{w}\right) = \lambda\left(||\vec{w}||\right)^2 \tag{3.18}$$

where the second equality comes from equation (3.17), and the third equality from the fact that λ is a scalar. According to equation (3.18), we have $\left(||\vec{w}||\right)^2 = 1$. Therefore Oja's learning rule yields feedforward weight vectors whose L2-norm is equal to 1 when the gain function of the output units is linear.

For arbitrary gain functions of the output units, we can still show that Oja's learning rule provides feedforward weight vectors whose L2-norm is bounded under some constraints. Analogously to Section 3.2.1, let us focus on a neuron i with a single input j, whose activity is $x_j \geqslant 0$. In this particular case, the fixed points of the learning rule ($\Delta w_{ij} = 0$) and their stability are determined. The derivative of equation (3.14) with respect to the postsynaptic activity is given by:

$$\frac{\mathrm{d}\left(\Delta w_{ij}(y_i)\right)}{\mathrm{d}y_i} = \eta\left(x_j - 2w_{ij}y_i\right) \tag{3.19}$$

The stability of the fixed points can thus be determined:

- $y_i = 0$: Unstable fixed point because:

$$\left.\frac{\mathrm{d}\left(\Delta w_{ij}(y_i)\right)}{\mathrm{d}y_i}\right|_{y_i=0} = \eta x_j > 0$$

- $y_i = \frac{x_j}{w_{ij}}$: Stable fixed point, because:

$$\left.\frac{\mathrm{d}\left(\Delta w_{ij}(y_i)\right)}{\mathrm{d}y_i}\right|_{\frac{x_j}{w_{ij}}} = -\eta x_j < 0$$

Therefore, during presentation of a single input pattern, the weight w_{ij} moves towards its stable fixed point $w_{ij} = \frac{x_j}{y_i}$ over the course of learning.

We will now compute the expectation of the L2-norm of the feedforward weight vectors at convergence of learning, based on a generalized version of Oja's equation (3.14) with power p and amplitude α of the heterosynaptic term:

$$\Delta w_{ij} = \eta\left(y_i x_j - \alpha y_i^p w_{ij}\right) \tag{3.20}$$

Oja's learning rule is a special case with $p = 2$ and $\alpha = 1$. The expectation of the weight update over all input pattern presentations reads:

$$\langle \Delta w_{ij} \rangle = \eta \left(\langle y_i x_j \rangle - \alpha \langle y_i^p w_{ij} \rangle \right) \tag{3.21}$$

At convergence of learning, the expectation of the weight update is zero: $\langle \Delta w_{ij} \rangle = 0$, hence:

$$\begin{aligned} 0 &= \eta \left(\langle y_i x_j \rangle - \alpha \langle y_i^p \rangle w_{ij} \right) \\ \Rightarrow w_{ij} &= \frac{\langle y_i x_j \rangle}{\alpha \langle y_i^p \rangle} \end{aligned} \tag{3.22}$$

We keep in mind that the activity y_i of output unit i is a function of its input a_i (equation (3.3)): $y_i = f(a_i) = f\left(\sum_{j=1}^N w_{ij} x_j\right)$. Therefore, if the gain function f is invertible:

$$\begin{aligned} a_i &= f^{-1}\left(y_i\right) \\ \sum_{j=1}^N w_{ij} x_j &= f^{-1}\left(y_i\right) \end{aligned} \tag{3.23}$$

The gain function f is invertible everywhere except at its zero values. For example, if a piecewise linear gain function is chosen, it is invertible at all its linear sections, but not at the breakpoints between linear sections.

The expected L2-norm of the feedforward weight vector is then given by:

$$\begin{aligned} \sqrt{\sum_{j=1}^N w_{ij}^2} &= \sqrt{\sum_{j=1}^N w_{ij} \frac{\langle y_i x_j \rangle}{\alpha \langle y_i^p \rangle}} \\ &= \sqrt{\frac{1}{\alpha \langle y_i^p \rangle} \langle y_i \left(\sum_{j=1}^N w_{ij} x_j \right) \rangle} \\ &= \sqrt{\frac{1}{\alpha \langle y_i^p \rangle} \langle y_i f^{-1}(y_i) \rangle} \end{aligned} \tag{3.24}$$

For a special case with Oja's learning rule ($p = 2$, $\alpha = 1$) and a linear gain function f: $y_i = \sum_{j=1}^N w_{ij} x_j$ and $f^{-1}(y_i) = y_i$, we recover (as above):

$$\sqrt{\sum_{j=1}^N w_{ij}^2} = \sqrt{\frac{1}{1 \langle y_i^2 \rangle} \langle y_i y_i \rangle} = 1 \tag{3.25}$$

For an arbitrary gain function and an arbitrary value of the power p, however, convergence of the L2-norm should be investigated case by case. For example, in a single WTA settings with step gain function for the output units (hence $y_i = 1$ when output unit i is a winner, $y_i = 0$ otherwise), Oja's learning rule ensures that the L2-norm of all feedforward weight vectors is

identical. However, in more complicated settings, such as soft WTA, a classical Oja's learning rule cannot be used because different winners for the same input pattern can have different activity values y_i, hence the feedforward weight vectors are not all normalized to the same norm.

3.3.3 Temporally unstable learning

During unsupervised competitive learning with step gain function, the vector of feedforward weights onto a given output unit converges to the direction of the center of mass of the cluster of input patterns for which it is a winner. Therefore, with constant learning rate, the feedforward weight vector will always move around the center of mass of the patterns it is selective for, as presentation of each pattern of the cluster will pull the weight vector towards it. Furthermore, if some clusters are overlapping or very close to each others, it is possible that some output units switch their selectivity from one cluster to the other over the course of learning. Hence, for arbitrary input patterns, unsupervised competitive learning is temporally unstable. Grossberg has demonstrated that stability and convergence of incremental updating can only be proven if the input space is very sparse, meaning that the minimal overlap between two patterns of the same cluster is bigger than the maximal overlap between patterns of different clusters (Grossberg (1987b)).

To overcome the issue of temporally unstable learning, the learning rate can be decreased over time. It reduces the update of the feedforward weight vectors upon each input pattern presentation, consequently also avoiding a change of the winning output unit upon presentation of the same input pattern at different time points during learning.

3.4 Initialization of the weights

In standard unsupervised competitive learning, the feedforward weight vectors onto each output unit are randomly initialized. They cannot start with zero elements, otherwise none of the output units would ever be activated, because their sole source of activation comes from the input layer. In addition, they should be initialized to the same L2-norm to ensure fair competition between output units (see Section 3.2).

Depending on the random initialization, different cases arise during learning, see Figure 3.6. Ideally, the feedforward weight vectors are each initialized relatively close to different clusters of inputs, therefore upon learning they each become selective for a different cluster (Figure 3.6a). However, some instances of random initialization might not be optimal for perfect clustering of the input patterns, hence subclustering or dead units issues may happen.

Subclustering occurs if two (or more) feedforward weight vectors are initialized close to the same input cluster (Figure 3.6b). In this case, they will each become selective for a subset of that cluster, and represent different prototypes of the same category. As it provides a more

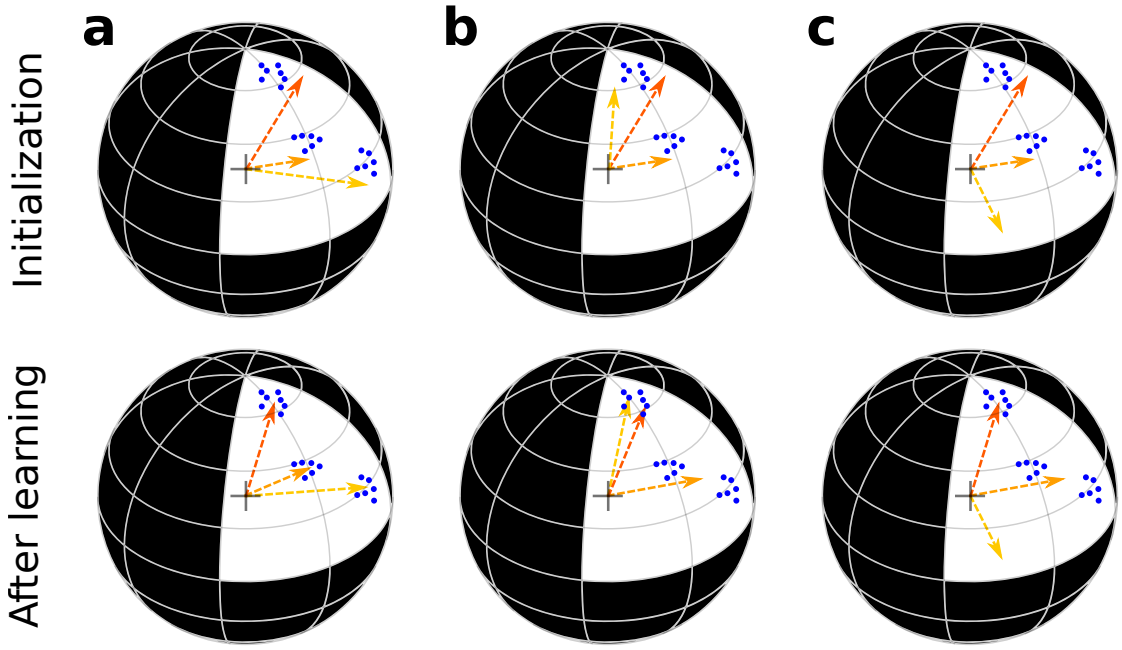

Figure 3.6 – Outcome for different initializations of the feedforward weight vectors. Normalized input patterns are represented by blue dots on the surface of a hypersphere. The dimension of the hypersphere corresponds to the dimension of the input layer. We restrict ourselves to positive input activities, hence the input patterns are all located on the positive quadrant of the hypersphere. Arrows represent the feedforward weight vectors onto three output units. For simplicity, they all have a norm equal to the one of the input patterns (hence the arrow ends are on the surface of the hypersphere as well). Upon learning, they move towards the center of mass of the input patterns for which they are the winner. A case where one output unit has to win the competition for each input pattern is considered. (**a**) Ideal initialization: Each weight vector becomes selective for a different cluster, more specifically the one they were initialized the closest to (and hence for which they always won the competition). (**b**) Subclustering: Two weight vectors (yellow and red) were initialized close to the top cluster, therefore they alternatively won the competition for some patterns of the top cluster, and became selective for two different subsets of the same cluster. The orange weight vector is initialized further away, hence it becomes selective for the center of mass of the two remaining clusters. (**c**) Dead unit: One weight vector (yellow) is initialized very far from the subspace where the input patterns lie, hence it is never updated because the corresponding output unit never wins the competition. Instead, the output unit with the orange weight vector wins the competition for all input patterns of the two bottom clusters, therefore the orange weight vector becomes selective for the center of mass of the two bottom clusters.

fine-grained representation of that cluster, it is not a problem in itself. However, if the network had been designed as a single WTA network with as many output units as the number of input categories to classify, it implies that at least one category cannot be properly represented by the network. More specifically, if the network was set up in a way that enforces a winning output unit for each pattern presentation, at least one output unit becomes selective for a mixture of several clusters (case in Figure 3.6b). On the other hand, if a winner is not enforced, some clusters of input patterns may never succeed in activating any output unit, so they are not represented by the network.

Dead units are output units that never get activated by presentation of any input pattern. As they never win the competition, their weights are never updated in the direction of the input patterns, hence they always stay silent. Dead units arise if some feedforward weight vectors are initialized in a direction that is too far from the subspace of input patterns (Figure 3.6c). Dead units do not help network computation, therefore they should be avoided.

3.4.1 Algorithmic solutions to the problem of dead units

Dead units are a common issue in unsupervised competitive learning, and several algorithmic strategies have been developed to avoid this problem (reviewed in Hertz et al. (1991)).

The first and most straightforward solution is to use smart initialization of the feedforward weight vectors. By smart, it is meant that each weight vector is initialized close to a cluster of input patterns (ideal case in Figure 3.6a). They can be each set to one input pattern of each cluster, or even directly to the center of mass of the clusters. However, this strategy is only applicable if there is knowledge about the input patterns, which is not the case in unsupervised settings.

Second, in the leaky learning strategy every output unit updates its feedforward weights (Rumelhart and Zipser (1985); Grossberg (1987b)). The winner of the competition uses a large (fast) learning rate, while the losers use a smaller (slower) learning rate. In this way, all output units move towards the presented input pattern, to different extents. It is not a Hebbian rule anymore, as silent postsynaptic output units still update their feedforward weights.

Third, in the “conscience” strategy, every output unit is assigned an activity threshold. It is increased for the winning output unit, to make it harder for that same output unit to win again the competition in the future. The activity thresholds of the losing output units are decreased, therefore making them more likely to win the competition for subsequent input patterns, even if they were initialized very far from the subspace of input patterns. This algorithm can even be set up in such a way that every output unit wins for the same proportion of input patterns on average (Grossberg (1976); Bienenstock et al. (1982); Rumelhart and Zipser (1985); DeSieno (1988)).

3.5 Open questions

In this work, I investigate whether adult dentate gyrus neurogenesis might provide a biological solution to the problem of dead units in unsupervised competitive learning.

The dentate gyrus is frequently modeled as a WTA network because of its numerous inhibitory recurrent connections (Acsády et al. (1998); Amaral et al. (2007)) and its sparse activity (Chawla et al. (2005)). In addition, it has been shown to be necessary for discrimination of similar (but not distinct) stimuli (Gilbert et al. (2001); Hunsaker and Kesner (2008)), and it receives most of its input from entorhinal cortex (Amaral et al. (2007)) through perforant pathway connections which exhibit Hebbian plasticity (Schmidt-Hieber et al. (2004); Ge et al. (2007); McHugh et al. (2007)). Therefore, it is natural to model the whole circuit from entorhinal cortex to dentate gyrus using unsupervised competitive learning.

Several algorithmic approaches have been developed to avoid that some units of the WTA network never win the competition and remain dead units in unsupervised learning settings (Hertz et al. (1991)). However, most of these strategies lack a biological interpretation. The adult-born dentate granule cells have been shown, however, to promote discrimination of similar (but not distinct) stimuli (Clelland et al. (2009); Sahay et al. (2011a)), even though they represent only a small proportion of the whole dentate granule cells population (Van Praag et al. (1999); Cameron and McKay (2001)). Therefore, they must be integrated in the preexisting circuit in a way that ensure that they become functional units, and help computation of the network. To my knowledge, none of the previous modeling studies has taken into account the particular scheme of integration of adult-born dentate granule cells to investigate if it facilitates that newborn cells do not remain dead units when behavioral demands are present.

Besides, normalization of the feedforward weight vectors onto the output units is critical in unsupervised competitive learning to ensure fair competition between output units (Section 3.2). Some biologically plausible learning rules already make sure that all weight vector keep the same norm (Oja (1982)). However, they fulfill this requirement only in particular conditions of the network and gain function of its units. In addition, most of the learning rules that are used for modeling neurogenesis bypass this issue by simply artificially normalizing the weight vectors at every learning step (Section 2.4).

Here, to stay closer to biology, we would like to model the entorhinal cortex to dentate gyrus circuit as a soft WTA network with assemblies of units active for each input pattern presentation. Therefore, we need to implement an unsupervised learning rule for the feedforward connections which is biologically plausible and ensures normalization of all feedforward weight vectors to the same norm despite the different activity levels of output units for different input pattern presentations.

Results Part II

4 Integration of adult-born dentate granule cells

4.1 Introduction

In the adult mammalian brain, neurogenesis, the creation of new neurons, is restricted to a few brain areas, such as the olfactory bulb and the dentate gyrus (Deng et al. (2010)). The dentate gyrus is the entry point of input from cortex, primarily entorhinal cortex (EC), to the hippocampus (Amaral et al. (2007)), which is believed to be a substrate of learning and memory (Jarrard (1993)). Adult-born cells in dentate gyrus mostly develop into dentate granule cells (DGCs), the main excitatory cell type (Deng et al. (2010)).

The properties of rodent adult-born DGCs change as a function of their maturation stage, until they become indistinguishable from other mature DGCs at approximately 8 weeks (Deng et al. (2010); Johnston et al. (2016)) (Figure 2.2). Many of them die before they fully mature (Dayer et al. (2003)). Their survival is experience-dependent, and depends on NMDA receptor activation (Tashiro et al. (2006)). Initially, newborn DGCs have enhanced excitability (Schmidt-Hieber et al. (2004); Li et al. (2017)) and stronger synaptic plasticity than mature DGCs, reflected by a larger LTP amplitude and a lower threshold for induction of LTP (Wang et al. (2000); Schmidt-Hieber et al. (2004); Ge et al. (2007)). Furthermore, after 4 weeks of maturation adult-born DGCs have only weak connections to interneurons, while at 7 weeks of age their activity causes strong feedback inhibition of mature DGCs (Temprana et al. (2015)).

Newborn DGCs receive no direct connections from mature DGCs (Deshpande et al. (2013); Alvarez et al. (2016)), (yet see Vivar et al. (2012)), but are indirectly activated via interneurons (Alvarez et al. (2016); Heigele et al. (2016)). During maturation, the γ-aminobutyric acid (GABAergic) input from interneurons to adult-born DGCs switches from excitatory in the early phase to inhibitory in the late phase of maturation (Deng et al. (2010)) (Figure 2.2). Analogous to a similar transition during embryonic and early postnatal stages (Wang and Kriegstein (2010)), this transition is caused by a change in the expression profile of chloride cotransporters, from NKCC1 in the early phase to KCC2 in the late phase (Ben-Ari (2002); Owens and Kriegstein (2002); Ge et al. (2006)). Importantly, it has been shown that GABAergic inputs are crucial for the integration of newborn DGCs into the preexisting circuit (Ge et al.

(2006); Chancey et al. (2013); Alvarez et al. (2016); Heigele et al. (2016)).

Adult-born DGCs are preferentially reactivated by stimuli similar to the ones they experienced during their early phase of maturation, up to 3 weeks after cell birth (Tashiro et al. (2007)). Even though the amount of newly generated cells per month is rather low, 3 to 6% of the total DGCs population (Van Praag et al. (1999); Cameron and McKay (2001)), adult-born DGCs are critical for behavioral pattern separation (Clelland et al. (2009); Sahay et al. (2011a); Jessberger et al. (2009)), in particular in tasks where similar stimuli or contexts have to be discriminated (Clelland et al. (2009); Sahay et al. (2011a)). However, the functional role of adult-born DGCs is controversial (Sahay et al. (2011b); Aimone et al. (2011)). One view is that newborn DGCs contribute to pattern separation through a modulatory role (Sahay et al. (2011b)), via a combination of enhanced excitability (Schmidt-Hieber et al. (2004); Li et al. (2017)), and absence of inhibition (Alvarez et al. (2016); Heigele et al. (2016)). Another view suggests that newborn DGCs act as encoding units. This view is supported by the fact that newborn DGCs have a critical window of maturation when they encode features of their environment (Kee et al. (2007); Tashiro et al. (2007)). Yet, the mechanism by which they promote pattern separation is still unknown. Some authors have even challenged the role of newborn DGCs in pattern separation in the classical sense and have proposed a pattern integration effect instead (Aimone et al. (2011)).

To address this controversy, we present a model of how newborn DGCs integrate into the pre-existing circuit. Our results suggest that the switch from lateral excitation to lateral inhibition during the maturation of newborn DGCs may play a crucial role for their proper integration into the existing network of DGCs. Our model shows how the particular integration scheme of newborn DGCs makes them particularly suitable to encode novel stimuli.

4.2 Methods

4.2.1 Network architecture and rate neurons dynamics

DGCs are the principal cells of the dentate gyrus. They mainly receive excitatory projections from the entorhinal cortex through the perforant path and GABAergic inputs from local interneurons, as well as excitatory input from Mossy cells. They project to CA3 pyramidal cells, inhibitory neurons, and Mossy cells (Acsády et al. (1998); Henze et al. (2002); Amaral et al. (2007); Temprana et al. (2015)). In our model, we omit Mossy cells and describe the dentate gyrus as a competitive circuit consisting of N_{DGC} dentate granule cells and N_I GABAergic interneurons (Figure 4.1). The activity of N_{EC} neurons in EC represents an input pattern $\vec{x} = (x_1, x_2, ..., x_{N_{EC}})$, with $x_i \geqslant 0 \forall\, i$ because firing rates are positive quantities. As the perforant path also induces strong feedforward inhibition in the dentate gyrus (Li et al. (2013)), we assume that the effective EC activity is normalized, such that $||\vec{x}|| = 1$ for any input pattern $\vec{x}$. We use P different input patterns $\vec{x}^{\mu}$, $1 \leqslant \mu \leqslant P$ in the simulations of the model.

The EC neurons have excitatory all-to-all connections to the DGCs. In rodent hippocampus,

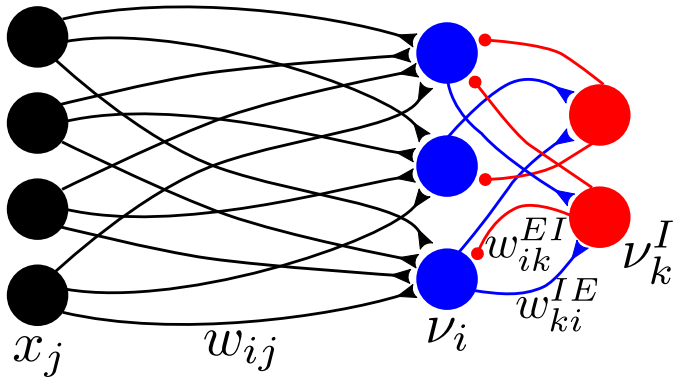

Figure 4.1 – Architecture of the biologically plausible network. EC neurons (black, rate x_j) are fully connected with weights w_{ij} to DGCs (blue, rate ν_i). DGCs and interneurons (red, rate ν_k^I) are mutually connected with probability p_{IE} and p_{EI} and weights w_{ki}^{IE} and w_{ik}^{EI}, respectively. Connections with a triangular end (black and blue) are glutamatergic, the others (red) are GABAergic.

spiking of mature DGCs activate interneurons in DG, which in turn inhibit other mature DGCs (Leutgeb et al. (2007); Temprana et al. (2015); Alvarez et al. (2016)). In our model, the DGCs are thus recurrently connected with inhibitory neurons (Figure 4.1). Connections from DGCs to interneurons exist with probability p_{IE} and have a weight w_{IE}. Similarly, connections from interneurons to DGCs occur with probability p_{EI} and have a weight w_{EI}. All parameters are reported in Table 4.1.

Before an input pattern is presented, all rates of model DGCs are initialized to zero. Upon stimulation with input pattern $\vec{x}$, the firing rate ν_i of DGCs i evolves according to (Miller and Fumarola (2012)):

$$\tau_m \frac{\mathrm{d}\nu_i}{\mathrm{d}t} = -\nu_i + \tanh\left(\frac{[I_i - b_i]_+}{L}\right) \tag{4.1}$$

where $[.]_+$ denotes rectification: $[a] = a$ for $a > 0$ and zero otherwise. L indicates steepness of the tanh, b_i is a firing threshold and I_i the total input to cell i:

$$I_i = \sum_{j=1}^{N_{EC}} w_{ij} x_j + \sum_{k=1}^{N_I} w_{ik}^{EI} \nu_k^I \tag{4.2}$$

with x_j the activity of EC input neuron j, $w_{ij} \geqslant 0$ the feedforward weight from EC input neuron j to DGC i, and w_{ik}^{EI} the weight from inhibitory neuron k to DGC i. The sum runs over all inhibitory neurons, but the weights are set to $w_{ik}^{EI} = 0$ if the connection is absent. The firing rate ν_i is unit-free and normalized to a maximum of 1, which we interpret as a firing rate of 10 Hz. We take the synaptic weights as unit-less parameters such that I_i is also unit-free.

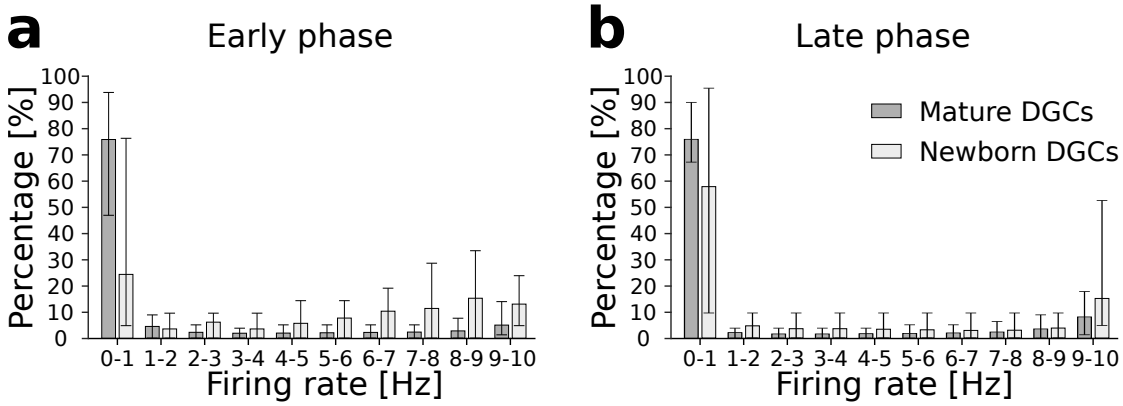

Figure 4.2 – Firing rate distribution. Distribution of the percentage of DGCs (mean with 10th and 90th percentiles) in each bin of firing rate upon presentation of MNIST patterns: (**a**) at the end of the early phase of maturation, and (**b**) at the end of the late phase of maturation. Percentages are obtained by normalizing by the number of neurons in each subpopulation (79 mature neurons in dark grey, 21 newborn neurons in light grey).

The firing rate v_k^I of inhibitory neuron k, is defined as:

$$\tau_{\text{inh}} \frac{\mathrm{d}v_k^I}{\mathrm{d}t} = -v_k^I + [I_k^I - p^* N_{DGC}]_+ \tag{4.3}$$

with p^* a parameter which relates to the desired ensemble sparsity, and I_k^I the total input towards interneuron k, given as:

$$I_k^I = \sum_{i=1}^{N_{DGC}} w_{ki}^{IE} v_i \tag{4.4}$$

with w_{ki}^{IE} the weight from DGC i to inhibitory neuron k. (We set $w_{ki}^{IE} = 0$ if the connection is absent.) The feedback from inhibitory neurons ensures a sparse activity of model DGCs for each pattern. With $p^* = 0.1$ we find that more than 70 % of model DGCs are silent (firing rate < 1 Hz (Senzai and Buzsáki (2017))) for each pattern presentation, and less than 10% are highly active (firing rate > 9 Hz) (Figure 4.2), consistent with the experimentally observed activity sparsity in dentate gyrus (Chawla et al. (2005)).

4.2.2 Plasticity rule

Projections from EC onto newborn DGCs exhibit Hebbian plasticity (Schmidt-Hieber et al. (2004); Ge et al. (2007); McHugh et al. (2007)), and NMDA receptors in DGCs are necessary for discrimination of similar contexts in a fear conditioning task in mice (McHugh et al. (2007)). Therefore, in our model the connections from EC neurons to DGCs are plastic, following a Hebbian learning rule which exhibits long-term depression (LTD) or long-term potentiation (LTP) depending on the firing rate v_i of the postsynaptic cell (Bienenstock et al.

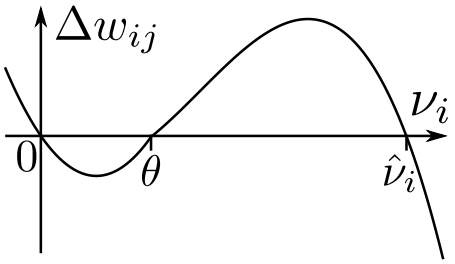

Figure 4.3 – Weight update as a function of postsynaptic activity. Weight update Δw_{ij} as a function of the firing rate ν_i of the postsynaptic DGC showing LTD for $\nu_i < \theta$ and LTP for $\theta < \nu_i < \hat{\nu}_i$.

(1982); Artola et al. (1990); Pfister and Gerstner (2006)) (Figure 4.3). Input patterns $\vec{x}^\mu$, $1 \leqslant \mu \leqslant P$, are presented in random order. Once the firing rate of the DGCs in response to pattern $\vec{x}^\mu$ have converged, the weight between any EC input neuron j and a DGC i is updated ($w_{ij}^{(\mu)} = w_{ij}^{(\mu-1)} + \eta \Delta w_{ij}$), according to the following plasticity rule:

$$\Delta w_{ij} = -\alpha x_j \nu_i [\theta - \nu_i]_+ + \gamma x_j \nu_i [\nu_i - \theta]_+ - \beta w_{ij} [\nu_i - \theta]_+ \nu_i^3 \tag{4.5}$$

with $\alpha = \frac{\alpha_0}{\theta^3}$, $\gamma = \gamma_0 - \theta$, x_j the firing rate of presynaptic EC input neuron j, and ν_i the firing rate of postsynaptic DGC i. The values of the parameters α_0, γ_0, β, and θ are given in Table 4.1. The weights are hard-bounded from below at 0, i.e. if equation (4.5) leads to a new weight smaller than zero, w_{ij} is set to zero. The first two terms of expression (4.5) are a variation of the BCM rule (Bienenstock et al. (1982)). The third term implements heterosynaptic plasticity (Chistiakova et al. (2014); Zenke and Gerstner (2017)). Because the first two terms of the plasticity rule are Hebbian and proportional to the presynaptic activity x_j, the active DGCs ($\nu_i > \theta$) update their feedforward weights in direction of the input pattern $\vec{x}$. Moreover, all weights onto neuron i are downregulated heterosynaptically by an amount that increases supra-linearly with the postsynaptic rate ν_i. Similar to learning in a competitive network, the vector of feedforward weights onto active DGCs will move towards the center of mass of the cluster of patterns they are selective for (Kohonen (1989); Hertz et al. (1991)).

As visible on Figure 4.3, for a given input pattern $\vec{x}^\mu$, there are three fixed points for the postsynaptic firing rate (values for which $\Delta w_{ij}(\nu_i) = 0$): $\nu_i = 0$, $\nu_i = \theta$, and $\nu_i = \sqrt{\frac{\gamma x_j^\mu}{\beta w_{ij}}} =: \hat{\nu}_i$ (the negative root is omitted, because $\nu_i \geqslant 0$ by construction). For $\nu_i < \theta$, there is LTD, so the weights move toward zero: $w_{ij} \to 0$, while for $\nu_i > \theta$, there is LTP, so the weights move toward: $w_{ij} \to \frac{\gamma x_j^\mu}{\beta \nu_i^2}$. If a pattern $\vec{x}^\mu$ is presented only for a short time these fixed points are not reached during a single pattern presentation.

Table 4.1 – Parameters for the simulations of the biologically plausible network

Network	$N_{EC} = 144$	$N_{DGC} = 100$
	$N_I = 25$	
Connectivity	$w_{IE} = 1$	$w_{EI} = -\frac{1}{p_{EI} * N_I}$
	$p_{IE} = 0.9$	$p_{EI} = 0.9$
Dynamics	$\tau_m = 20$ ms	$\tau_{\text{inh}} = 2$ ms
	$L = 0.5$	$p^* = 0.1$
Plasticity	$\alpha_0 = 0.05$	$\beta = 1$
	$\gamma_0 = 10$	$\theta = 0.15$
	$\nu_0 = 0.2$	
Numerical simulations	$\Delta t = 0.1$ ms	$\eta = 0.01$
	$\eta_b = 0.01$	

Fixed points and their stabilities for a single input pattern

The fixed points of the dynamics for a single input pattern are given by the zero crossings of the $\Delta w_{ij}(\nu_i)$ curve. If a slight increase of the activity results in an increase of the weight, this would further increase the activity and make the fixed point unstable. Therefore, the stability of the fixed points is assessed by computing the derivative of $\Delta w_{ij}(\nu_i)$ with respect to ν_i and determining its sign when evaluated on each fixed point. The derivative is given by:

$$\frac{\mathrm{d}\left(\Delta w_{ij}(\nu_i)\right)}{\mathrm{d}\nu_i} = \begin{cases} \alpha x_j(2\nu_i - \theta), \text{ if } \nu_i < \theta \\ -4\beta w_{ij}\nu_i^3 + 3\beta\theta w_{ij}\nu_i^2 + 2\gamma x_j \nu_i - \gamma\theta x_j, \text{ if } \nu_i > \theta \end{cases} \tag{4.6}$$

Hence the stability of the fixed points is as follows:

- $\nu_i = 0$: stable fixed point, because:

$$\left.\frac{\mathrm{d}\left(\Delta w_{ij}(\nu_i)\right)}{\mathrm{d}\nu_i}\right|_{\nu_i=0} = -\alpha\theta x_j < 0$$

- $\nu_i = \theta$: unstable fixed point, because:
  - starting from below ($\nu_i^* < \theta$):

$$\lim_{\nu_i^* \to \theta} \left.\frac{\mathrm{d}\left(\Delta w_{ij}(\nu_i)\right)}{\mathrm{d}\nu_i}\right|_{\nu_i=\nu_i^*} = \alpha\theta x_j > 0$$

– starting from above ($\nu_i^* > \theta$):

$$\lim_{\nu_i^* \to \theta} \left.\frac{\mathrm{d}\left(\Delta w_{ij}(\nu_i)\right)}{\mathrm{d}\nu_i}\right|_{\nu_i=\nu_i^*} = -4\beta w_{ij}\theta^3 + 3\beta\theta w_{ij}\theta^2 + 2\gamma x_j\theta - \gamma\theta x_j$$
$$= \theta\left(\gamma x_j - \beta w_{ij}\theta^2\right) > 0$$

The last equality holds, because $\theta < \sqrt{\frac{\gamma x_j}{\beta w_{ij}}}$, so it implies that $\gamma x_j > \beta w_{ij}\theta^2$.

We note here that the derivative is not smooth at the point $\nu_i = \theta$.

- $\nu_i = \sqrt{\frac{\gamma x_j}{\beta w_{ij}}} > \theta$: stable fixed point, because:

$$\left.\frac{\mathrm{d}\left(\Delta w_{ij}(\nu_i)\right)}{\mathrm{d}\nu_i}\right|_{\nu_i=\sqrt{\frac{\gamma x_j}{\beta w_{ij}}}} = -4\beta w_{ij}\frac{\gamma x_j}{\beta w_{ij}}\sqrt{\frac{\gamma x_j}{\beta w_{ij}}} + 3\beta\theta w_{ij}\frac{\gamma x_j}{\beta w_{ij}} + 2\gamma x_j\sqrt{\frac{\gamma x_j}{\beta w_{ij}}} - \gamma\theta x_j$$
$$= -2\gamma x_j\sqrt{\frac{\gamma x_j}{\beta w_{ij}}} + 2\gamma\theta x_j$$
$$= 2\gamma x_j\left(\theta - \sqrt{\frac{\gamma x_j}{\beta w_{ij}}}\right) < 0$$

Winners, losers, and quasi-orthogonal inputs

We define the winners as the DGCs which become strongly active ($\nu_i > \theta$) during presentation of an input pattern. Since the input patterns are normalized to have an L2-norm of 1 ($||\vec{x}^\mu|| = 1$ by construction), and the L2-norm of the feedforward weight vectors is bounded (see Section Direction and length of the weight vector), the winning units are the ones whose weight vectors $\vec{w}_i$ (row of the feedforward connectivity matrix) align best with the current input pattern $\vec{x}^\mu$. Furthermore, we say that an input pattern $\vec{x}^\mu$ is "quasi-orthogonal" to a weight vector $\vec{w}_i$ if $I_i = \sum_{j=1}^{N_{EC}} w_{ij}x_j + \sum_{k=1}^{N_I} w_{ik}^{EI}\nu_k^I < b_i$. If an input pattern $\vec{x}^\mu$ is quasi-orthogonal to a weight vector $\vec{w}_i$, then neuron i does not fire in response to $\vec{x}^\mu$. Note that for a case without inhibitory neurons and with $b_i \to 0$, we recover the standard orthogonality condition.

Direction and length of the weight vector

Let us denote the ensemble of patterns for which neuron i is a winner by C_i and call this the set of winning patterns ($C_i = \{\mu|\nu_i > \theta\}$). Suppose that neuron i is quasi-orthogonal to all other patterns, so that for all $\mu \notin C_i$ we have $\nu_i = 0$. Then the feedforward weight vector of neuron i converges in expectation to:

$$\vec{w}_i = \frac{\gamma}{\beta}\frac{\langle G_1(\nu_i)\vec{x}\rangle_{\mu\in C_i}}{\langle G_2(\nu_i)\rangle_{\mu\in C_i}} \tag{4.7}$$

where $G_1(\nu_i) = (\nu_i - \theta)\nu_i$ and $G_2(\nu_i) = (\nu_i - \theta)\nu_i^3$. Hence $\vec{w}_i$ is a weighted average over all winning patterns. (To obtain expression (4.7), set equation (4.5) to zero and solve for w_i.)

The squared length of the feedforward weight vector can be computed by multiplying equation (4.7) with $\vec{w}_i$:

$$||\vec{w}_i||^2 = \vec{w}_i \cdot \vec{w}_i = \frac{\gamma}{\beta} \frac{\langle G_1(\nu_i)\,(\vec{w}_i \cdot \vec{x})\rangle_{\mu\in C_i}}{\langle G_2(\nu_i)\rangle_{\mu\in C_i}} \tag{4.8}$$

Since input patterns have length one, the scalar product on the right-hand side can be rewritten as $\vec{w}_i \cdot \vec{x} = ||\vec{w}_i|| \cos(\alpha)$ where α is the angle between the weight vector and pattern $\vec{x}$. Division by $||\vec{w}_i||$ yields the L2-norm of the feedforward weight vector:

$$||\vec{w}_i|| = \frac{\gamma}{\beta} \frac{\langle G_1(\nu_i)\cos(\alpha)\rangle_{\mu\in C_i}}{\langle G_2(\nu_i)\rangle_{\mu\in C_i}} \tag{4.9}$$

where the averages run, as before, over all winning patterns.

Let us now derive bounds for $||\vec{w}_i||$. First, since $\cos(\alpha) \leqslant 1$ we have $\langle G_1(\nu_i)\cos(\alpha)\rangle_{\mu\in C_i} \leqslant \langle G_1(\nu_i)\rangle_{\mu\in C_i}$. Second, since for all winning patterns $\nu_i > \theta$, where θ is the LTP threshold, we have $\langle G_2(\nu_i)\rangle_{\mu\in C_i} \geqslant \langle(\nu_i - \theta)\,\nu_i\rangle\theta^2$. Thus the length of the weight vector is finite and bounded by:

$$||\vec{w}_i|| \leqslant \frac{\gamma}{\beta} \frac{\langle G_1(\nu_i)\rangle_{\mu\in C_i}}{\langle G_2(\nu_i)\rangle_{\mu\in C_i}} \leqslant \frac{\gamma}{\beta} \frac{1}{\theta^2} \tag{4.10}$$

It is possible to make the second bound tighter if we find the winning pattern with the smallest firing rate $\nu_{\min}$ such that $\nu_i \geqslant \nu_{\min}\ \forall i \in C_i$:

$$||\vec{w}_i|| \leqslant \frac{\gamma}{\beta} \frac{1}{(\nu_{\min})^2} \tag{4.11}$$

The bound is reached if neuron i is winner for a single input pattern.

We can also derive a lower bound. For a pattern $\mu \in C_i$, let us write the firing rate of neuron i as $\nu_i(\mu) = \bar{\nu}_i + \Delta\nu_i(\mu)$ where $\bar{\nu}_i$ is the mean firing rate of neuron i averaged across all winning patterns and $\langle\Delta\nu_i\rangle_{\mu\in C_i} = 0$. We assume that the absolute size of $\Delta\nu_i$ is small, i.e., $\langle(\Delta\nu_i)^2\rangle_{\mu\in C_i} \ll (\bar{\nu}_i)^2$. Linearization of equation (4.9) around $\bar{\nu}_i$ yields:

$$\begin{aligned} ||\vec{w}_i|| &= \frac{\gamma}{\beta} \frac{G_1(\bar{\nu}_i)}{G_2(\bar{\nu}_i)} \langle\cos(\alpha)\rangle_{\mu\in C_i} \qquad (4.12)\\ &\quad + \frac{\gamma}{\beta} \frac{G_1'(\bar{\nu}_i)\langle\cos(\alpha)\Delta\nu_i\rangle_{\mu\in C_i} G_2(\bar{\nu}_i) - G_1(\bar{\nu}_i)\langle\cos(\alpha)\rangle_{\mu\in C_i} G_2'(\bar{\nu}_i)\langle\Delta\nu_i\rangle_{\mu\in C_i}}{(G_2(\bar{\nu}_i))^2} \\ &= \frac{\gamma}{\beta} \frac{G_1(\bar{\nu}_i)}{G_2(\bar{\nu}_i)} \langle\cos(\alpha)\rangle_{\mu\in C_i} + \frac{\gamma}{\beta} \frac{G_1'(\bar{\nu}_i)}{G_2(\bar{\nu}_i)} \langle\cos(\alpha)\Delta\nu_i\rangle_{\mu\in C_i} \qquad (4.13) \end{aligned}$$

Elementary geometric arguments for a neuron model with monotonically increasing frequency-current curve yield that the value of $\langle\cos(\alpha)\Delta\nu_i\rangle_{\mu\in C_i}$ is positive (or zero), because an increase

in the angle α lowers both the cosine and the firing rate, giving rise to a positive correlation. Since we are interested in a lower bound, we can therefore drop the term proportional to G_1' and evaluate the ratio G_1/G_2 to find:

$$||\vec{w}_i|| \geqslant \frac{\gamma}{\beta}\frac{1}{(\bar{\nu}_i)^2}\langle\cos(\alpha)\rangle_{\mu\in C_i} \geqslant \frac{\gamma}{\beta}\frac{1}{(\nu_{\max})^2}\cos(\hat{\alpha}) \tag{4.14}$$

where $\nu_{\max}$ is the maximal firing rate of a DGC and $\hat{\alpha} = \max_{\mu\in C_i}\{\alpha\}$ is the angle of the winning pattern that has the largest angle with the weight vector. The first bound is tight and is reached if neuron i is winner for only two patterns.

To summarize we find that the length of the weight vector remains bounded in a narrow range. Hence, for a reasonable distribution of input patterns and weight vectors, the value of $||\vec{w}_i||$ is similar for different neurons i, so that the weight vector will have, after convergence, similar lengths for all DGCs that are winners for at least one pattern. In our simulations with the MNIST data set, we find that the length of feedforward weight vectors lies in the range between 9.3 and 11.1 across all responsive neurons.

Early maturation phase

During the early phase of maturation, the GABAergic input onto a newborn DGC with index l has an excitatory effect. In the model, it is implemented as follows: $w_{lk}^{EI} = -w_{EI} > 0$ with probability p_{EI} for any interneuron k and $w_{lk}^{EI} = 0$ otherwise (no connection). Since newborn cells do not project yet onto inhibitory neurons (Temprana et al. (2015)), we have $w_{kl}^{IE} = 0\ \forall l$. Newborn DGCs are known to have enhanced excitability (Schmidt-Hieber et al. (2004); Li et al. (2017)), so their threshold is kept at $b_l = 0\ \forall l$. Presentation of all patterns of the data set once (1 epoch) is sufficient to reach convergence of the feedforward weights onto newborn DGCs.

Because the newborn DGCs receive lateral excitation via interneurons and their thresholds are zero during the early phase of maturation, the lateral excitatory GABAergic input is always sufficient to activate them. Hence, if the firing rate of a newborn DGC exceeds the LTP threshold θ, the feedforward weights grow towards the presented input pattern, cf. equation (4.5).

Late maturation phase

During the late phase of maturation (starting at about 3 weeks (Ge et al. (2006))), the GABAergic input onto newborn DGCs switches from excitatory to inhibitory. In terms of our model, it means that all existing w_{lk}^{EI} connections switch their sign to $w_{EI} < 0$. Furthermore, since newborn DGCs develop lateral connections to inhibitory neurons in the late maturation phase (Temprana et al. (2015)), we set $w_{kl}^{IE} = w_{IE}$ with probability p_{IE}, and $w_{kl}^{IE} = 0$ otherwise. The thresholds of newborn DGCs are updated after presentation of pattern μ at time $n \cdot T$ ($b_l^{(n)} = b_l^{(n-1)} + \eta_b \Delta b_l$) according to $\Delta b_l = \nu_l - \nu_0$, where ν_0 is a reference rate, to mimic the decrease of excitability as newborn DGCs mature. Therefore the distribution of firing rates of

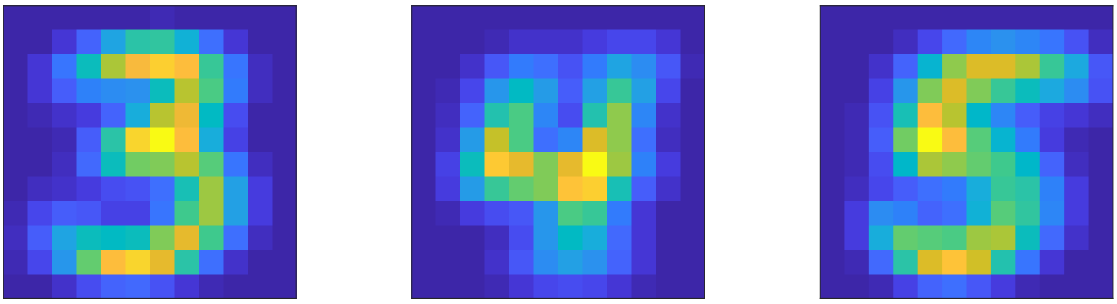

Figure 4.4 – Center of mass for three ensembles of patterns from the MNIST data set, visualized as 12x12 pixel patterns.

newborn DGCs is shifted to the left (towards lower firing rates) at the end of the late phase of maturation compared to the early phase of maturation (Figure 4.2). A sufficient condition for a newborn DGC to win the competition upon presentation of patterns of the novel cluster is that the scalar product between a pattern of the novel cluster and the feedforward weight vector onto the newborn DGC is larger than the scalar product between the pattern of the novel cluster and the feedforward weight vector onto any of the mature DGCs. Analogous to the early phase of maturation, presentation of all patterns of the data set once (1 epoch) is sufficient to reach convergence of the feedforward weights onto newborn DGCs.

4.2.3 Input patterns

We use the MNIST 12x12 patterns (LeCun et al. (1998)) ($N_{EC} = 144$), normalized such that the L2-norm of each pattern is equal to 1. The training set contains approximately 6000 patterns per digit ($K = 10$ clusters), while the testing set contains about 1000 patterns per digit (Figure 4.4).

4.2.4 Classification performance

To evaluate whether the newborn DGCs contribute to the dentate gyrus network function, we study classification performance and compare it to control cases (see next section). First, the feedforward weights are learned upon presentation of many input patterns from the training set, as described in Section 4.2.2. After convergence, we keep them fixed and determine classification on the test set using artificial readout units (RO).

To do so, the readout weights w_{ki}^{RO} from model DGC i to readout unit k are initialized to random values drawn from a uniform distribution: $w_{ki}^{RO} \sim \sigma\mathcal{U}(0,1)$, with $\sigma = 0.1$. The number of readout units, N_{RO}, corresponds to the number of learned classes. To adjust the readout weights, all patterns of the training data set that belong to the learned classes are presented one after the other. For each pattern $\vec{x}^{\mu}$, we let the firing rate of the DGCs converge (values at

convergence: ν_i^μ). The activity of a readout unit k is given by:

$$\nu_k^{RO,\mu} = g\left(I_k^{RO,\mu}\right) = g\left(\sum_{i=1}^{N_{DGC}} w_{ki}^{RO}\nu_i^\mu\right) \tag{4.15}$$

As we aim to assess the performance of the network of DGCs, the readout weights are adjusted by an artificial supervised learning rule. The loss function, which corresponds to the difference between the activity of the readout units and a one-hot representation of the corresponding pattern label (Hertz et al. (1991)):

$$\begin{aligned} L(W^{RO}) &= \frac{1}{2}\sum_{k=1}^{N_{RO}}(L_k^\mu - \nu_k^{RO,\mu})^2 \\ &= \frac{1}{2}\sum_{k=1}^{N_{RO}}\left(L_k^\mu - g\left(\sum_{i=1}^{N_{DGC}} w_{ki}^{RO}\nu_i^\mu\right)\right)^2 \end{aligned} \tag{4.16}$$

with L_k^μ the element k of a one-hot representation of the real label of pattern $\vec{x}^\mu$, $\vec{L}^\mu$, is minimized by stochastic gradient descent:

$$\begin{aligned} \Delta w_{ki}^{RO,\mu} &= -\eta\frac{\mathrm{d}L(W^{RO})}{\mathrm{d}w_{ki}^{RO}} \\ &= \eta\left(L_k^\mu - g\left(\sum_{i=1}^{N_{DGC}} w_{ki}^{RO}\nu_i^\mu\right)\right)g'\left(\sum_{i=1}^{N_{DGC}} w_{ki}^{RO}\nu_i^\mu\right)\nu_i^\mu \\ &= \eta(L_k^\mu - \nu_k^{RO,\mu})g'\left(I_k^{RO,\mu}\right)\nu_i^\mu \end{aligned} \tag{4.17}$$

The readout units have a rectified tangent hyperbolic gain function: $g(x) = \tanh(2[x]_+)$, whose derivative is: $g'(x) = 2\left(1-(\tanh(2[x]_+))^2\right)$. We learn the weights of the readout units over 100 epochs of presentations of all training patterns with $\eta = 0.01$, which is sufficient to reach convergence.

Thereafter, the readout weights are fixed. Each test set pattern belonging to one of the learned classes is presented once, and the firing rates of the DGCs are let to converge. Finally, the activity of the readout units $\nu_k^{RO,\mu}$ is computed and compared to the correct label L_k^μ of the presented pattern. If the readout unit with the highest activity value is the one that represents the class of the presented input pattern, the pattern is said to be correctly classified. Classification error is given by the number of misclassified patterns divided by the total number of test patterns of the learned classes.

4.2.5 Control cases

In our standard setting, patterns from a third digit are presented to a network that has previously only seen patterns from two digits. The question is whether neurogenesis helps when adding the third digit. We use several control cases to compare with the neurogenesis case.

In two control cases, we either keep all feedforward connections towards the DGCs plastic (Figure 4.8c), or fix the feedforward connections for all selective DGCs but keep unselective neurons plastic (as in the neurogenesis case) (Figure 4.8b). However, in both instances, the DGCs do not mature in the two-step process of our model of neurogenesis. Finally, in the third control case, all three digits are learned in parallel (Figure 4.8a).

4.2.6 Pretraining with two digits

As we are interested by neurogenesis at the adult stage, we pretrain the network with patterns from two digits, such that it already stores some memories before neurogenesis takes place. To do so, we randomly initialize the EC neurons to DGCs connection weights: they are drawn from a uniform distribution ($w_{ij} \sim U[0,1]$). The L2-norm of the feedforward weight vector onto each DGC is then normalized to 1, to ensure fair competition between DGCs during learning. Then, we present all patterns from digits 3 and 4 in random order, as many times as needed for convergence of the weights. During each pattern presentation, the firing rates of the DGCs are computed (Section 4.2.1) and their feedforward weights are updated according to our plasticity rule (Section 4.2.2). We find that we need approximately 40 epochs for convergence of the weights, and use 80 epochs to make sure that all weights are stable. Such a large number of epochs is needed (versus only 2 epochs in total for newborn DGCs), because weights are randomly initialized (while for newborn DGCs, the early phase acts as a "smart initialization" procedure). At the end of the pretraining, our network is considered to correspond to an adult stage, because some DGCs are selective for prototypes of the pretrained digits (Figure 4.5a).

4.3 Results

We hypothesize that the two-step maturation process caused by the switch of GABA from excitation to inhibition is crucial for functional integration of newborn cells: since excitatory GABAergic input potentially increases cooperativity within the dentate gyrus network, we predicted that newborn DGCs would respond to familiar representations during the early phase of maturation, but not during the late phase, when inhibitory GABAergic input leads to competition.

To test this prediction, we pretrained a model network of 100 mature DGCs with input from 144 EC cells to respond to patterns representing two different digits from MNIST, a standard data set in artificial intelligence (LeCun et al. (1998)). Even though we do not expect EC neurons to show a 2-dimensional arrangement, the use of 2-dimensional patterns provides a simple way to visualize the activity of all 144 EC neurons in our model in the form of digits (Figure 4.4). Pretraining was based on a synaptic plasticity rule which combines LTP and LTD (Section 4.2.2, Figure 4.3) (Bienenstock et al. (1982); Artola et al. (1990); Pfister and Gerstner (2006)).

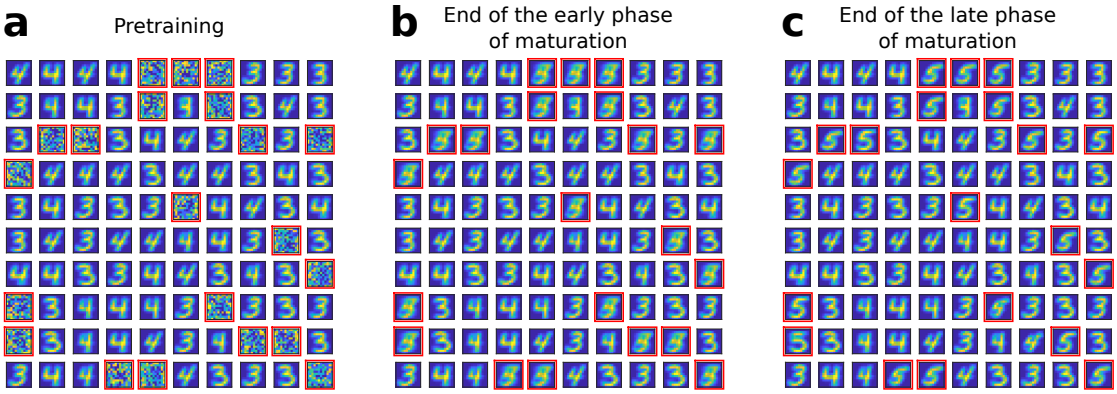

Figure 4.5 – Newborn DGCs learn novel patterns. Receptive fields, defined as the set of feedforward weights, are visually represented in a 2-dimensional organization. (**a**) During pretraining, patterns from MNIST digits 3 and 4 are presented to the network. At the end of the pretraining, some DGCs have receptive fields corresponding to the two learned digits, while others remain non-selective (highlighted by red frames). (**b**) Unselective neurons are replaced by newborn DGCs, which adapt their feedforward weights while patterns from digits 3, 4, and 5 are presented. At the end of the early phase of maturation, the receptive fields of all newborn DGCs (red frames) show mixed selectivity. (**c**) At the end of the late phase of maturation, newborn DGCs are selective for patterns from the novel digit 5, with different writing styles.

4.3.1 Different prototypes are learned during the pretraining period

After pretraining with patterns from digits 3 and 4, we examined the receptive fields of the DGCs, defined as the set of feedforward weights of connections from all 144 EC neurons onto one DGC. We observed that out of the 100 DGCs, some developed a receptive field that corresponds to digit 3, others a receptive field that corresponds to digit 4, and the remaining ones stayed unselective (Figure 4.5a). We classified the DGCs with non-selective receptive fields as unresponsive units. Because they are barely activated by any of the input patterns, their incoming synaptic weights show LTD whenever they are slightly active, causing a further reduction of the cells responsiveness. In our model, reduction of the cell responsiveness below a critical value eventually causes cell death. The number of unresponsive units depends on the value of the θ parameter in equation (4.5): the higher it is, the larger the number of unresponsive units. The selective DGCs represent different prototypes of the two digits, visible here as different writing styles and inclinations (Figure 4.5a). At the end of pretraining, the classification error (Section 4.2.4) was low: 0.75% (classification error on digit 3: 1.29%; digit 4: 0.20%), indicating that nearly all input patterns in the two digits are well represented by the network of mature DGCs.

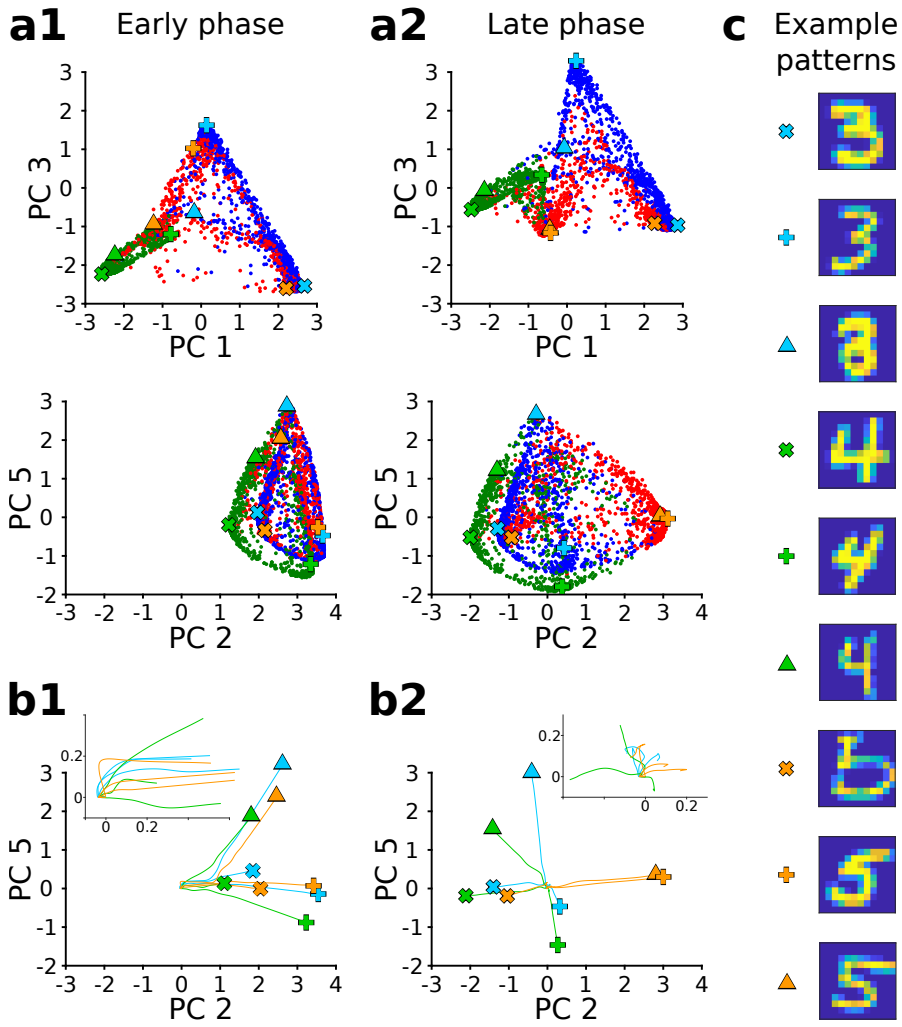

Figure 4.6 – The representation of novel patterns occupies a previously empty subspace. (**a**) Projections of the final firing rates of all 100 DGCs on PC 1 and PC 3 (first row), and on PC 2 and PC 5 (second row), at the end of the early (a1) or late (a2) phase of maturation of the newborn DGCs. Each pattern of the MNIST test set corresponds to one point. Color indicates digit 3 (blue), 4 (green) or 5 (red). The PCs were determined at the end of the late phase of maturation. (**b**) Trajectories of firing rates in the PC-space for a few example patterns at the end of the early (b1) and late (b2) phase of maturation of newborn DGCs. It takes about 170 to 200 ms for the trajectories to converge to their final points, where symbols mark the corresponding patterns. Insets: zoom on trajectories during the first 10 ms. (**c**) Example patterns from the test set, with the symbols used in (a) and (b).

4.3.2 Newborn neurons can learn novel patterns

After convergence of synaptic weights during pretraining, unresponsive model neurons died and were replaced by newborn DGCs. Model newborn DGCs go through two maturation phases (Section 4.2.2). The early phase of maturation is cooperative because, for each pattern presentation, the activated mature DGCs laterally excite the newborn DGCs via GABAergic interneurons. This indirect activation of newborn DGCs drives the growth of their receptive fields in a direction similar to those of the currently active DGCs. As a result, at the end of the early phase of maturation, newborn DGCs show a receptive field corresponding to a mixture of several patterns (Figure 4.5b).

In the late phase of maturation, model newborn DGCs receive inhibitory GABAergic input from interneurons, similar to the input received by mature DGCs. Given that at the end of the early phase, newborn DGCs have receptive fields similar to those of mature DGCs, lateral inhibition induces competition with mature DGCs for activation during presentation of patterns from the novel digit. Model newborn DGCs start the late phase of maturation with a low threshold (high excitability), while mature DGCs are less excitable. Since the distribution of firing rates of newborn DGCs is shifted to lower firing rates in the late phase of maturation compared to the early phase (Figure 4.2), the activation of newborn DGCs is facilitated for those input patterns for which no mature DGC has selectivity. Therefore, in the late phase of maturation, competition drives the synaptic weights of newborn DGCs towards a receptive field corresponding to patterns from the ensemble of novel input patterns, i.e. digit 5 (Figure 4.5c).

To better characterize how DGCs represent various input patterns, we performed Principal Component Analysis (PCA) on the vector of final firing rates of all DGCs in response to MNIST test patterns representing digits 3, 4, and 5. We then projected the firing rates on Principal Components (PCs) that were selected based on their visualization power (Figure 4.6a2). The firing rates of all DGCs were also projected on the same PCs at the end of the early phase of maturation (Figure 4.6a1). First, we observe that ensembles of firing rate responses are scattered, rather than concentrated, even for a single digit, indicating good pattern separation abilities of our network within and between digits (Figure 4.6a2). Second, the representations of a few extremal patterns seem to span a low-dimensional manifold (Figure 4.6c). Third, we observe that the representation of novel patterns lies close to the representation of pretrained patterns at the end of the early phase of maturation of newborn DGCs (Figure 4.6a1), while during the late phase of maturation the representation expands into a previously empty subspace (Figure 4.6a2), consistent with the experimentally observed promotion of pattern separation of newborn DGCs. Finally, by examining the trajectories of the firing rates of the DGCs for a few example patterns, we further notice that two novel patterns (orange swiss cross and orange triangle) are located far from each other (and close to the familiar patterns) at the end of the early phase (Figure 4.6b1), but close to each other (and far from familiar patterns) at the end of the late phase of maturation of newborn DGCs (Figure 4.6b2). The zoom in insets further show the influence of lateral inhibition on firing rate dynamics of the DGCs.

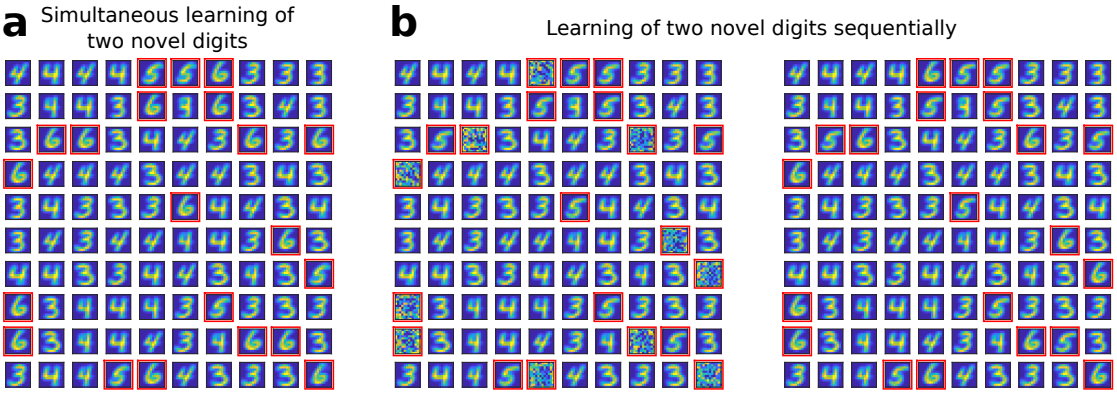

Figure 4.7 – Two novel digits can be learned. (**a**) Several novel digits can be learned simultaneously. After pretraining with patterns from digits 3 and 4 as in Figure 4.5a, unresponsive neurons are replaced by newborn DGCs. When patterns from digits 3, 4, 5, and 6 are presented in random order, newborn DGCs exhibit after maturation receptive fields with selectivity for the novel digits 5 and 6. (**b**) Several novel digits can be learned sequentially. After pretraining with digits 3 and 4, ten randomly selected unresponsive neurons are replaced by newborn DGCs. Patterns from digits 3, 4, and 5 are presented in random order, while newborn DGCs mature and develop selectivity for the novel digit 5, with different writing styles. Later, the eleven remaining unresponsive neurons of the network are replaced by newborn DGCs. When patterns from the novel digit 6 are presented intermingled with patterns from digits 3, 4, and 5, the newborn DGCs develop selectivity for digit 6.

To test the quality of the learned representations, we compute classification performance by a linear classifier for the three ensembles of digits (Section 4.2.4). We obtain an overall classification error of 5.44% (classification error for digit 3: 9.50%; digit 4: 1.83%; digit 5: 4.82%). We compare this performance with that of a network where all three digit ensembles are simultaneously pretrained (Figure 4.8a). In this case, the overall classification error is 7.91% (classification error for digit 3: 13.17%; digit 4: 1.22%; digit 5: 9.30%). Classification performance is therefore slightly better when a novel ensemble of patterns is learned sequentially by newborn DGCs, than if all patterns are learned simultaneously. This is due to the fact that there are similar numbers of unresponsive neurons at the end of pretraining, regardless of how many patterns are learned (compare Figure 4.5a and Figure 4.8a). The number of unresponsive units is similar because the θ parameter value in equation (4.5) is identical in all simulations, and the MNIST patterns all lie in a comparable subspace.

Furthermore, if two novel ensembles of digits (instead of a single one) are introduced during maturation of newborn DGCs, we observe that some newborn DGCs become selective for one of the novel digits, while others become selective for the other novel digit (Figure 4.7a). Therefore, newborn DGCs can ultimately promote separation of several novel ensembles of patterns, no matter if they are learned simultaneously (Figure 4.7a) or sequentially (Figure 4.7b).

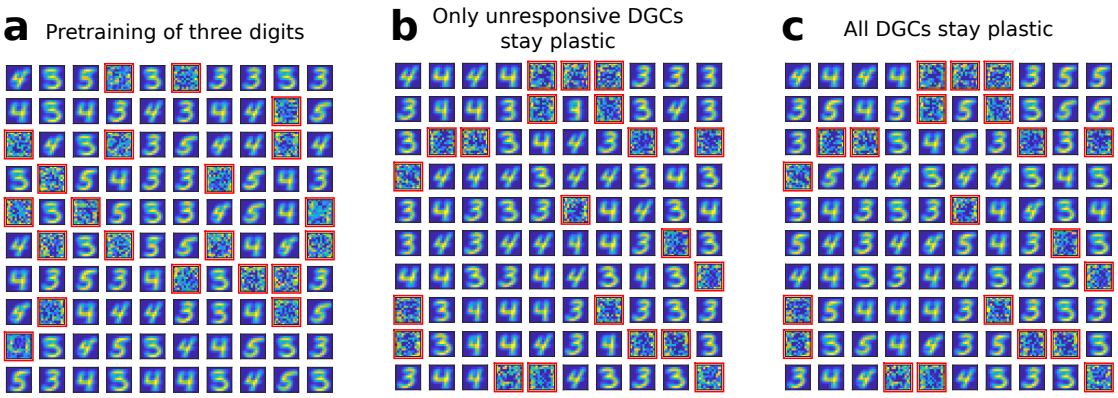

Figure 4.8 – Control cases. (**a**) Training on digits 3, 4 and 5 simultaneously without neurogenesis. Patterns from digits 3, 4 and 5 are presented to the network while all DGCs learn their feedforward weights. After pretraining, some DGCs have receptive fields corresponding to the three learned digits, while some remain non-selective (as in Figure 4.5a). (**b**) Control without maturation. After pretraining with patterns from digits 3 and 4 as in Figure 4.5a, only the unresponsive neurons stay plastic, but they fail to become selective for digit 5 when patterns from digits 3, 4, and 5 are presented in random order. (**c**) If all DGCs stay plastic when patterns from digit 5 are introduced, some of the DGCs previously responding to patterns from digits 3 or 4 become selective for digit 5 because they have strong weights and their receptive fields are close enough from some of the patterns from digit 5 for the cells to be activated.

4.3.3 The switch from excitation to inhibition is necessary for learning of novel representations

To assess whether maturation of newborn DGCs promotes learning of a novel ensemble of digit patterns, we compare with a case without neurogenesis. Similar to the neurogenesis case, patterns from the novel digit 5 are introduced after pretraining with patterns from digits 3 and 4. In the control case, the thresholds and weights of all unresponsive neurons remained plastic after pretraining, similar to the neurogenesis case, while the feedforward weights and thresholds of DGCs that developed selectivity during pretraining were fixed. The only differences with the neurogenesis case are that unresponsive neurons (i) keep their feedforward weights (i.e., no reinitialization), and (ii) keep the same connections from and to inhibitory neurons.

We find that without neurogenesis, the previously unresponsive DGCs do not become selective for the novel digit 5, no matter during how many epochs patterns are presented (we went up to 100 epochs here) (Figure 4.8b). Therefore, if patterns from digit 5 are presented to the network, it fails to discriminate them from the previously learned digits 3 and 4: the overall classification error is 18.31% (classification error for digit 3: 14.06%; digit 4: 2.44%; digit 5: 40.58%). This result suggests that inclusion of newborn DGCs is beneficial for sequential learning of novel patterns.

As a further control, we compare with a case where all DGCs keep plastic feedforward weights.

We observe that in the case where all neurons are plastic, learning of the novel digit occurs at the cost of loss of selectivity of mature neurons. Several DGCs switch their selectivity to become sensitive to the novel ensemble of patterns (Figure 4.8c), while none of the previously unresponsive units becomes selective for the novel digit (compare with Figure 4.5a). This induces a drop of classification performance to 9.08% error (classification error for digit 3: 14.55%; digit 4: 1.63%; digit 5: 11.10%). We observe that the classification error for digit 3 is the one which increases the most. This is due to the fact that many DGCs previously selective for digit 3 are now selective for digit 5.

5 Adult-born dentate granule cells promote discrimination of similar stimuli

5.1 Introduction

Adult dentate gyrus neurogenesis has been shown to affect behavioral pattern separation in a variety of tasks (Clelland et al. (2009); Sahay et al. (2011a); Jessberger et al. (2009)). Clelland and colleagues have shown that ablation of adult dentate gyrus neurogenesis in mice impairs performance in a spatial, navigable, radial arm maze task and a spatial, but non-navigable, touch screen task for spatially close stimuli, but not for stimuli having large spatial separation (Clelland et al. (2009)). Furthermore, Sahay and colleagues have observed that adult dentate gyrus neurogenesis is necessary for discrimination of similar but not distinct contexts in a fear conditioning in mice, and that enhanced neurogenesis promotes better context discrimination (Sahay et al. (2011a)). However, it is still unclear how adult-born DGCs impact pattern separation, if it is through a cell-autonomous function as individual encoding units, or through a modulatory role (Sahay et al. (2011b); Aimone et al. (2011)).

We suggest that adult-born DGCs are properly integrated into the existing network only if the stimuli that are encountered during their maturation are similar enough to already experienced stimuli. Indeed, in the early cooperative phase of maturation, newborn DGCs can only be indirectly activated by interneurons if the latter are sufficiently activated by mature DGCs. And for mature DGCs to be activated, the stimuli that are presented should be similar enough from their receptive field. If instead the presented stimuli are very distinct from the selectivity of mature DGCs, none will be sufficiently activated to indirectly activate the maturing newborn DGCs. Therefore, adult-born DGCs become selective for novel stimuli similar to familiar (already stored) stimuli, but not for distinct stimuli. Consequently, our model reveals why newborn DGCs promote pattern separation of similar stimuli, but do not impact pattern separation of distinct stimuli. We propose that the direct connection from EC to CA3, and then the backprojection from CA3 to dentate gyrus could be used as an alternate

pathway to represent distinct stimuli with high valence in the dentate gyrus.

5.2 Methods

5.2.1 Simplified rate network

We use a toy network and an artificial data set to determine if our theory for the integration of newborn DGCs can explain why adult dentate gyrus neurogenesis helps for the discrimination of similar, but not for distinct patterns.

The rate network described above is simplified as follows. We use K dentate granule cells for K clusters. Their firing rate ν_i is given by:

$$\tau_m \frac{\mathrm{d}\nu_i}{\mathrm{d}t} = -\nu_i + \mathscr{H}(I_i - b_i) \tag{5.1}$$

where $\mathscr{H}$ is the Heaviside step function. As before, b_i is the threshold, and I_i the total input towards neuron i:

$$I_i = \sum_{j=1}^{N_{EC}} w_{ij} x_j + \sum_{k \neq j}^{N_{DGC}} w_{rec} \nu_k \tag{5.2}$$

with x_j the input of presynaptic EC neuron j, w_{ij} the feedforward weight between EC neuron j and DGC i, and ν_k the firing rate of DGC k. Inhibitory neurons are modeled implicitly: each DGC directly connects to all other DGCs via inhibitory recurrent connections of value $w_{rec} < 0$. During presentation of pattern $\vec{x}^\mu$, the firing rates of the DGCs evolve according to equation (5.1). After convergence, the feedforward weights are updated: $w_{ij}^{(\mu)} = w_{ij}^{(\mu-1)} + \eta \Delta w_{ij}$. The synaptic plasticity rule is the same as before, see equation (4.5), but with the parameters reported in Table 5.1. They are different from those for the biologically-plausible network because we now aim for a single winning neuron for each cluster. Note that for an LTP threshold $\theta < 1$ all active DGCs update their feedforward weights, because of the Heaviside function for the firing rate (equation (5.1)).

Asssuming a single winner i^* for each pattern presentation, the input (equation (5.2)) to the winner is:

$$I_{i^*} = \vec{w}_{i^*} \cdot \vec{x}, \tag{5.3}$$

while the input to the losers is:

$$I_i = \vec{w}_i \cdot \vec{x} + w_{\mathrm{rec}}. \tag{5.4}$$

Therefore, two conditions need to be satisfied for a solution with a single winner:

$$\vec{w}_{i^*} \cdot \vec{x} > b_i \tag{5.5}$$

Table 5.1 – Parameters for the simulations of the simplified network

Network	$N_{EC} = 128$	$N_{DGC} = 3$
Connectivity	$w_{rec} = -1.2$	
Dynamics	$\tau_m = 20$ ms	
Plasticity	$\alpha_0 = 0.03$ $\gamma_0 = 1.65$	$\beta = 1$ $\theta = 0.15$
Numerical simulations	$\Delta t = 1$ ms	$\eta = 0.01$

for the winner to actually be active, and:

$$\vec{w}_i \cdot \vec{x} + w_{\text{rec}} < b_i \tag{5.6}$$

to prevent non-winners to become active. The value of b_i is lower in the early maturation phase than in the late maturation phase to mimic enhanced excitability (Schmidt-Hieber et al. (2004); Li et al. (2017)).

5.2.2 Input patterns

We use hand-made artificial patterns designed such that the distance between the centers of any two clusters is the same. All K clusters lie on the positive quadrant of the surface of a hypersphere of dimension $N_{EC} - 1$ (Figure 5.1a). The cluster centers are Walsh patterns shifted along the diagonal (Figure 5.1b):

$$\begin{aligned}
\vec{P}^1 &= \frac{1}{c_0}(1+\xi, 1-\xi, 1+\xi, 1-\xi, ..., 1+\xi, 1-\xi, 1+\xi, 1-\xi) \\
\vec{P}^2 &= \frac{1}{c_0}(1+\xi, 1+\xi, 1-\xi, 1-\xi, ..., 1+\xi, 1+\xi, 1-\xi, 1-\xi) \\
&... \\
\vec{P}^K &= \frac{1}{c_0}(1+\xi, 1+\xi, 1+\xi, 1+\xi, ..., 1-\xi, 1-\xi, 1-\xi, 1-\xi)
\end{aligned} \tag{5.7}$$

with $|\xi| < 1$ a parameter that determines the spacing between clusters. c_0 is a normalization factor to ensure that the center of mass of all clusters has an L2-norm of 1:

$$c_0 = \sqrt{N_{EC}\left(1+\xi^2\right)}. \tag{5.8}$$

The number of input neurons N_{EC} is $N_{EC} = 2^K$. The scalar product, and hence the angle Ω, between the center of mass of any pair of clusters k and l ($k \neq l$) is a function of ξ (Figure 5.1c):

$$\vec{P}^k \cdot \vec{P}^l = \frac{1}{1+\xi^2} = \cos(\Omega) \tag{5.9}$$

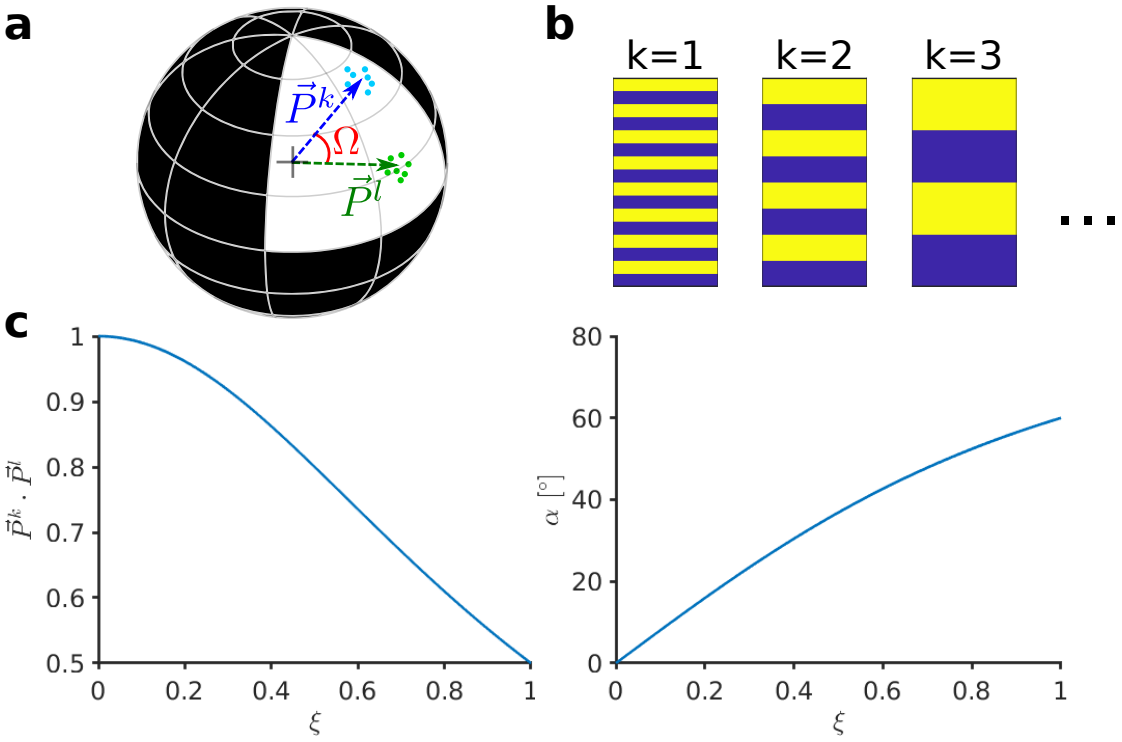

Figure 5.1 – Handmade dataset. (**a**) Center of mass of clusters k and l of the artificial data set ($\vec{P}_k$ and $\vec{P}_l$ respectively, separated by angle Ω) are represented by arrows that point to the surface of the positive quadrant of a hypersphere. Blue and green dots represent individual patterns. (**b**) Center of mass of the first three clusters of an artificial data set visualized as 16x8 pixel patterns. The 2-dimensional arrangement is chosen for visualization only. (**c**) Separation between the center of mass of the clusters of the artificial dataset. Left: The scalar product between the center of mass of any two clusters is a function of ξ. Right: The angle α as a function of ξ.

To make the artificial data set comparable to the MNIST 12x12 data set, we choose $K = 7$, so $N_{EC} = 128$, and we generate 6000 patterns per cluster for the training set and 1000 patterns per cluster for the testing set. The patterns $\vec{x}^{\mu(k)}$ of a given cluster k with center of mass $\vec{P}^k$ are sampled from a Von Mises-Fisher distribution (Mardia and Jupp (2009)):

$$\vec{x}^{\mu(k)} \sim \left(\sqrt{1-a^2}\right)\vec{\zeta} + a\vec{P}^k \tag{5.10}$$

with $\vec{\zeta}$ an L2-normalized vector taken in the space orthogonal to $\vec{P}^k$. The vector $\vec{\zeta}$ is obtained by performing the singular-value decomposition of $\vec{P}^k$ ($U\Sigma V^* = \vec{P}^k$), and multiplying the matrix U (after removing its first column), which corresponds to the left-singular vectors in the orthogonal space to $\vec{P}^k$, with a vector whose elements are drawn from the standard normal distribution. Then the L2-norm of the obtained pattern is set to 1, so that it lies on the surface of the hypersphere. A rejection sampling scheme is used to obtain a (Mardia and Jupp (2009)). The sample a is kept if $\kappa a + (N_{EC} - 1)\ln(1 - \psi a) - c \geqslant \ln(u)$, with κ a concentration parameter,

$\psi = \frac{1-b}{1+b}$, $c = \kappa\psi + (N_{EC}-1)\ln(1-\psi^2)$, u drawn from a uniform distribution $u \sim U[0,1]$, $a = \frac{1-(1+b)z}{1-(1-b)z}$, $b = \frac{N_{EC}-1}{\sqrt{4\kappa^2+(N_{EC}-1)^2}+2\kappa}$, and z drawn from a beta distribution $z \sim \mathscr{B}e(\frac{N_{EC}-1}{2}, \frac{N_{EC}-1}{2})$.

The concentration parameter κ characterizes the spread of the distribution around the center $\vec{P}^k$. In the limit where $\kappa \to 0$, sampling from the Von Mises-Fisher distribution becomes equivalent to sampling uniformly on the surface of the hypersphere, so the clusters become highly overlapping. In dimension $N_{EC} = 128$, if $\kappa > 10^3$ the probability of overlap between clusters is negligible. We use a value $\kappa = 10^4$.

Similar versus distinct patterns with the artificial data set

Using the artificial data set with $|\xi| < 1$ (equation (5.7)), the scalar product between the center of mass of two different clusters, given by equation (5.9), satisfies: $0.5 \leqslant \frac{1}{1+\xi^2} \leqslant 1$. This corresponds to $0° \leqslant \Omega \leqslant \Omega_{\max} = 60°$ (Figure 5.1c).

After stimulation with a pattern $\vec{x}$, it takes some time before the firing rates of the DGCs converge. We call two patterns "similar" if they activate, at least initially, the same output unit, while we consider two patterns as "distinct" if they do not activate the same output unit, not even initially. We now show that, with a large concentration parameter κ, patterns of different clusters are similar if $\xi < \sqrt{\frac{||\vec{w}_i||}{b_i} - 1}$ and distinct if $\xi > \sqrt{\frac{||\vec{w}_i||}{b_i} - 1}$.

We first consider a DGC i whose feedforward weight vector has converged towards the center of mass of cluster k. If an input pattern $\vec{x}^{\mu(k)}$ from cluster k is presented, it will receive the following initial input:

$$I_i = \vec{w}_i \cdot \vec{x}^{\mu(k)} = ||\vec{w}_i|| \cdot ||\vec{x}^{\mu(k)}|| \cdot \cos(\vartheta_{\text{kk}}) = ||\vec{w}_i|| \cdot \cos(\vartheta_{\text{kk}}) \tag{5.11}$$

where ϑ_{kk} is the angle between the pattern $\vec{x}^{\mu(k)}$ and the center of mass $\vec{P}^k$ of the cluster to which it belongs. The larger the concentration parameter κ for the generation of the artificial data set, the smaller the dispersion of the clusters, and thus the larger $\cos(\vartheta_{\text{kk}})$. If instead, an input pattern from cluster l is presented, that same DGC will receive a lower initial input:

$$I_i = \vec{w}_i \cdot \vec{x}^{\mu(l)} = ||\vec{w}_i|| \cdot ||\vec{x}^{\mu(l)}|| \cdot \cos(\vartheta_{\text{kl}}) \approx \frac{||\vec{w}_i||}{1+\xi^2} \tag{5.12}$$

The approximation holds for small dispersion of the clusters (large concentration parameter κ). We note that there is no subtraction of the recurrent input yet, because output units are initialized with zero firing rate before each pattern presentation. By definition, similar patterns stimulate (initially) the same DGCs. A DGC can be active for two clusters only if its threshold is:

$$b_i < \frac{||\vec{w}_i||}{1+\xi^2} \tag{5.13}$$

Therefore, with a high concentration parameter κ, patterns of different clusters are similar if

$\xi < \sqrt{\frac{||\vec{w}_i||}{b_i} - 1}$, while patterns of different clusters are distinct if $\xi > \sqrt{\frac{||\vec{w}_i||}{b_i} - 1}$.

Parameter choice

The upper bound of the expected L2-norm of the feedforward weight vector towards the DGCs at convergence can be computed, see equation (4.11). With the parameters in Table 5.1, the value is $\langle ||\vec{w}_i|| \rangle \leqslant 1.5$. Moreover, the input patterns for each clusters are highly concentrated, hence their angle with the center of mass of the cluster they belong to is close to 0, so we have $\langle ||\vec{w}_i|| \rangle \approx 1.5$. Therefore, at convergence, a DGC selective for a given cluster k receives an input $I_{i^*} = \vec{w}_{i^*} \cdot \vec{x}^{\mu(k)} \approx 1.5$ upon presentation of input patterns $\vec{x}^{\mu(k)}$ belonging to cluster k. We thus set $b_i = 1.2$ to satisfy condition (5.5). The threshold value ξ_{thresh} for which two clusters are similar (and above which two clusters are distinct) can be determined by equation (5.13) : $\xi_{\text{thresh}} = 0.5$. We created a handmade data set with $\xi = 0.2$ for the case of similar clusters, and a handmade data set with $\xi = 0.8$ for the distinct case.

Let us suppose that the weights of DGC i have converged and made this cell respond to patterns from cluster i. If another DGC k of the network is selective for cluster k, it ultimately gets the input $I_k = \vec{w}_k \cdot \vec{x}^{\mu(k)} + w_{\text{rec}} \approx \frac{1.5}{1+\xi^2} + w_{\text{rec}}$ upon presentation of input patterns $\vec{x}^{\mu(k)}$ belonging to cluster k ($k \neq i$). Hence, to satisfy condition (5.6), we need $w_{\text{rec}} < b_i - \max_\xi \left(\frac{1.5}{1+\xi^2} \right)$. Furthermore, a newborn DGC is born with a null feedforward weight vector, hence at birth, its input consists only of the indirect excitatory input from mature DGCs: $I_i = -w_{\text{rec}} > 0$. For the feedforward weight vector to grow, the condition $-w_{\text{rec}} > b_{\text{birth}}$ is also necessary (with b_{birth} the neuronal threshold of a newborn DGC at birth). We set $w_{\text{rec}} = -1.2$ and $b_{\text{birth}} = 0.9$, which satisfy the two above conditions.

Neurogenesis with the handmade data set

To save computation time, we initialize the feedforward weight vectors of two mature DGCs at two randomly chosen training patterns of the first two clusters, normalized such that they have an L2-norm of 1.5. We then present patterns from clusters 1 and 2, and let the feedforward weights evolve according to equation (4.5) until they reach convergence.

We thereafter introduce a novel pattern cluster as well as a newborn DGC in the network. The sequence of presentation of patterns from the three clusters (a novel one and two pretrained ones) is random. The newborn DGC maturation follows the same rules as before: it is born with a null feedforward weight vector. In the early phase, GABAergic input has an excitatory effect (Ge et al. (2006)) and the newborn DGC does not inhibit the mature DGCs (Temprana et al. (2015)). This is modeled by setting $w_{\text{rec}}^{NM} = -w_{\text{rec}}$ for the connections from mature to newborn DGC, and $w_{\text{rec}}^{MN} = 0$ for the connections from newborn to mature DGCs. The threshold of the newborn DGC starts at 0.9 at birth, mimicking enhanced excitability (Schmidt-Hieber et al. (2004); Li et al. (2017)), and increases linearly up to 1.2 (same threshold as the mature DGCs) over 12000 pattern presentations, reflecting loss of excitability with maturation.

The exact time window is not critical. In the late phase of maturation of the newborn DGC, GABAergic input switches to inhibitory (Ge et al. (2006)), and the newborn DGC recruits feedback inhibition onto mature DGCs (Temprana et al. (2015)). It is modeled by switching the sign of the connection from mature to newborn DGC: $w_{\text{rec}}^{NM} = w_{\text{rec}}$, and establishing connections from newborn to mature DGCs: $w_{\text{rec}}^{MN} = w_{\text{rec}}$. All patterns are presented once for the early phase.

The above paradigm is run separately for each of the two handmade data sets: the one where pattern clusters are similar ($\xi = 0.2$), and the one where pattern clusters are distinct ($\xi = 0.8$).

5.3 Results

5.3.1 Newborn dentate granule cells become selective for a similar novel cluster

To investigate whether our theory for integration of newborn DGCs can explain why adult dentate gyrus neurogenesis promotes discrimination of similar stimuli, but does not affect discrimination of distinct patterns (Clelland et al. (2009); Sahay et al. (2011a)), we used a simplified competitive winner-take-all network (Section 5.2.1) and constructed an artificial data set (Section 5.2.2) (Figure 5.1). The MNIST data set is not appropriate to tackle this question, because all digit clusters are similar and highly overlapping, reflected by a high within cluster dispersion compared to the separation between clusters.

After pretraining such that a first mature DGC responds to patterns of cluster 1 and a second mature DGC to those of cluster 2 (Figure 5.2b1,b2), we introduce a newborn DGC in the network, and present patterns from three clusters (the two pretrained ones, as well as a novel one). We observe that the newborn DGC ultimately becomes selective for the novel cluster if it is similar ($\xi = 0.2$) to the two pretrained clusters (Figure 5.2d1), but not if it is distinct ($\xi = 0.8$, Figure 5.2d2). Indeed, if the novel cluster is similar to one or several of the pretrained clusters, a pattern from the novel cluster activates the mature DGC that has a receptive field closest to the novel pattern. This activated mature DGC induces the activation of the newborn DGC via lateral excitatory GABAergic connections, to a level where LTP is triggered at active synapses. Similarly, excitatory GABAergic connections will also cause LTP of a newborn DGC, whenever one of the patterns of the pretrained clusters is presented. Thus, in the early phase of maturation, the feedforward weight vector onto the newborn DGC grows in the direction of the center of mass of all three clusters (the two pretrained ones and the novel one, because for each pattern presentation, one of the mature DGCs becomes active (compare Figure 5.2c1 and Figure 5.2e1). On the other hand, if the novel cluster is distinct (i.e. has a low similarity to pretrained clusters), patterns from the novel cluster do not activate any of the existing mature DGCs, whereas patterns from the pretrained clusters indirectly activate the newborn DGC. Hence plasticity makes the feedforward weight vector onto the newborn DGC move to the center of mass of the pretrained clusters (compare Figure 5.2c2 and Figure 5.2e2).

As a result of the different orientation of the feedforward weight vector onto the newborn DGC

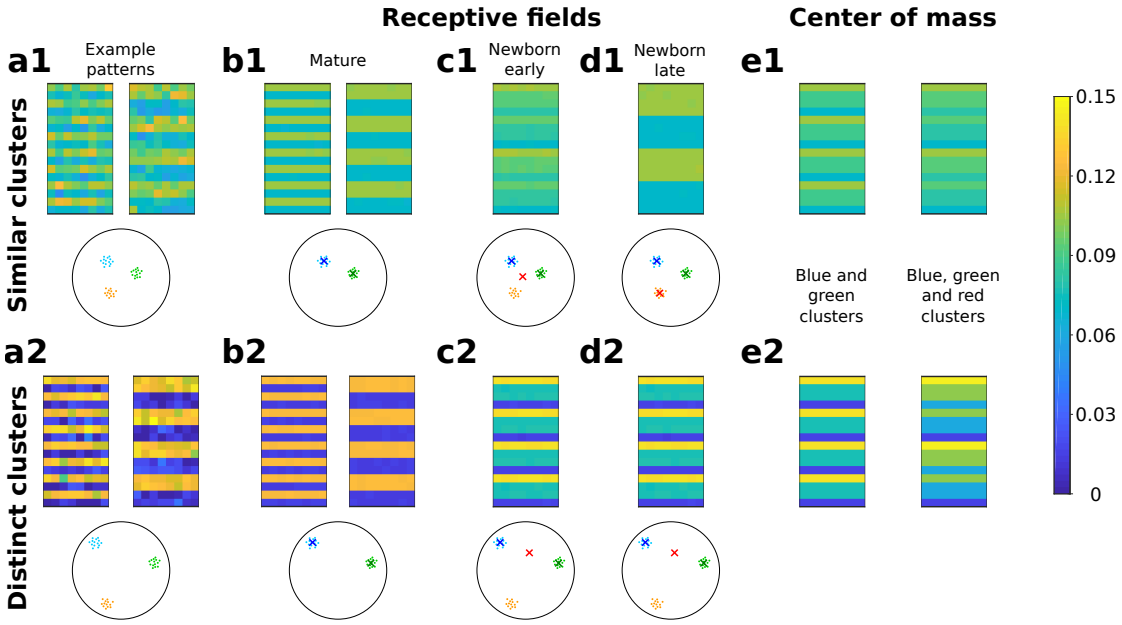

Figure 5.2 – A newborn DGC becomes selective for similar but not distinct novel stimuli. (**a**) Example patterns from clusters 1 and 2 if clusters are similar (a1, $\xi = 0.2$), and distinct (a2, $\xi = 0.8$). The receptive fields of the mature DGCs are initialized at example patterns. The schematics below the receptive fields represent the surface of the hypersphere, with dots corresponding to patterns. (**b**) After pretraining with patterns from two clusters, the receptive fields exhibit the noise-free prototypes (blue and green crosses) of each cluster. (**c**) If the clusters are similar, the receptive field of the newborn DGC (red cross) moves towards the center of mass of the three clusters during its early phase of maturation (c1), and if the clusters are distinct towards the center of mass of the two pretrained clusters (c2). (**d**) Receptive field after the late phase of maturation for the case of similar (d1) or distinct (d2) clusters. (**e**) Center of mass of all patterns of the blue and green clusters (left column) and of the blue, green and red clusters (right column) for the case of similar (e1) or distinct (e2) clusters. For better visualization, the receptive fields are normalized such that they have an L2-norm of 1. Hence the color scale is valid for the example patterns, the receptive fields, and the center of mass of the first two and the first three clusters.

at the end of the early phase of maturation, two different situations arise in the late phase of maturation, when lateral GABAergic connections are inhibitory. If the novel cluster is similar to the pretrained clusters, the weight vector onto the newborn DGC at the end of the early phase of maturation lies at the center of mass of the three clusters, and thus is closer to the novel cluster than the mature DGCs (Figure 5.2c1). So if a novel pattern is presented, the newborn DGC wins the competition between the three DGCs, and its feedforward weight vector moves towards the center of mass of the novel cluster (Figure 5.2d1). To the contrary, if the novel cluster is distinct, the weight vector onto the newborn DGC at the end of the early phase of maturation is located at the center of mass of the two pretrained clusters (Figure 5.2c2). If a novel pattern is presented, no output unit is activated since their receptive fields do not match the input. So the newborn DGC always stays silent and never updates its feedforward weights (Figure 5.2d2). These results are consistent with studies that have suggested that dentate gyrus is only involved in the discrimination of similar stimuli, but not distinct stimuli (Gilbert et al. (2001); Hunsaker and Kesner (2008)). For discrimination of distinct stimuli, another pathway might be used, such as the direct EC to CA3 connection (Fyhn et al. (2007); Vazdarjanova and Guzowski (2004)).

In conclusion, our model suggests that adult dentate gyrus neurogenesis promotes discrimination of similar patterns because newborn DGCs can ultimately become selective for a novel stimulus which is similar to already learned stimuli. On the other hand, newborn DGCs fail to represent a novel distinct stimulus, precisely because it is too distinct from other stimuli already represented by the network. Presentation of the novel distinct stimulus therefore does not induce synaptic plasticity of the newborn DGCs feedforward weight vector toward the novel cluster. In the simplified network, the transition between similar and distinct can be determined analytically (Section 5.2.2).

5.3.2 Similar clusters drive the receptive fields of newborn DGCs closer to the novel cluster in the early phase of maturation

To better characterize the evolution of the receptive field of the newborn DGC, we examine the growth of the norm of the feedforward weight vector onto it, as well as its angle with the center of mass of the novel cluster, as a function of maturation time (Figure 5.3).

In the early phase of maturation, the feedforward weight vector onto the newborn DGC grows, while its angle with the center of mass of the novel cluster stays constant. The norm stabilizes at a higher value in the case of similar patterns ($\xi = 0.2$, Figure 5.2c1 and Figure 5.3a1) than in the case of distinct patterns ($\xi = 0.8$, Figure 5.2c2 and Figure 5.3b1). It is due to the fact that the center of mass of three similar clusters lies closer to the surface of the sphere than the center of mass of two distinct clusters (Section 5.3.3).

In the late phase of maturation, the angle between the center of mass of the novel cluster and the feedforward weight vector onto the newborn DGC decreases in the case of similar patterns (Figure 5.3a2), but not in the case of distinct patterns (Figure 5.3b2), indicating that

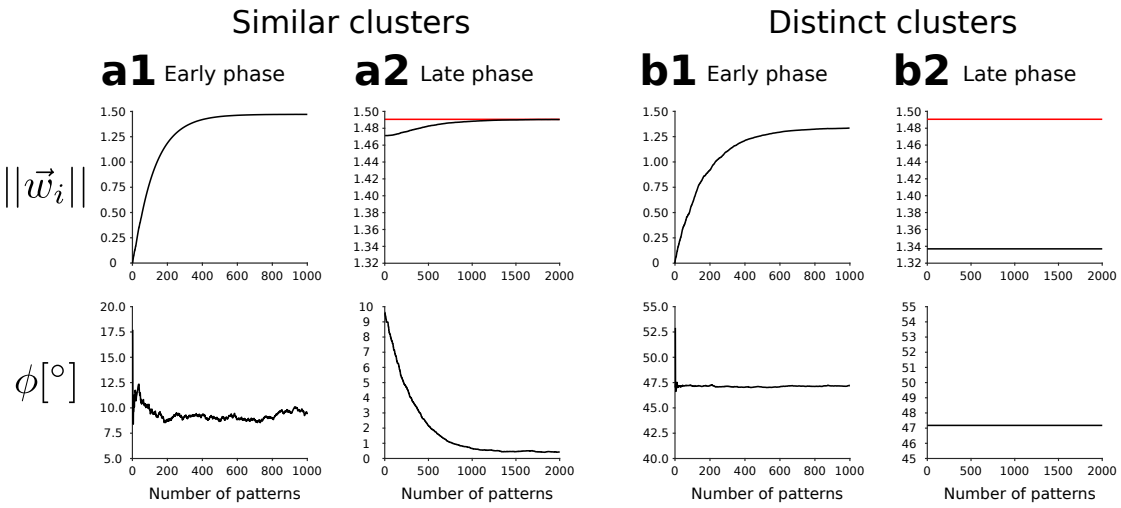

Figure 5.3 – Evolution of the norm and angle. Evolution of the total synaptic strength $||\vec{w}_i|| = \sqrt{\sum_j \left(w_{ij}\right)^2}$ of the newborn DGC (top row), and of its angular separation ϕ with the center of mass of the novel cluster (bottom row), as a function of the number of pattern presentations for the early phase of maturation (a1,b1) and the late phase of maturation (a2,b2). (**a**) The three clusters are similar ($\xi = 0.2$). (**b**) The three clusters are distinct ($\xi = 0.8$). Red line shows the mean value of the synaptic strength of the mature DGCs.

the newborn DGC becomes selective for the novel cluster for similar but not for distinct patterns. In addition, we observe a slight increase of the L2-norm of the feedforward weight vector onto the newborn DGC concomitantly with the decrease of angle with the center of mass of the novel cluster (Figure 5.3a2), because the center of mass of the novel cluster lies closer to the surface of the sphere than the center of mass of the three clusters.

5.3.3 Analytical computation of the L2-norm and angle

We consider the case where two mature DGCs have learned their synaptic connections, such that the first mature DGC with feedforward weight vector $\vec{w}_1$ is selective for cluster 1, and the second mature DGC with feedforward weight vector $\vec{w}_2$ is selective for cluster 2. By construction, we have $\vec{w}_1 = \langle||\vec{w}_1||\rangle\vec{P}_1$ and $\vec{w}_2 = \langle||\vec{w}_2||\rangle\vec{P}_2$, where $\langle||\vec{w}_k||\rangle$ is the expected L2-norm of the feedforward weight vector onto mature DGC k that is selective for pretrained cluster k. We remind here that the L2-norm of the center of mass of the clusters are normalized to one by construction, hence $||\vec{P}_k|| = 1\ \forall k$. In addition, the upper bound for the L2-norm of the weight vectors of the mature DGCs is $\langle||\vec{w}_1||\rangle = \langle||\vec{w}_2||\rangle = 1.5$ (Section 5.2.2). In our case, we obtain $\langle||\vec{w}_1||\rangle = \langle||\vec{w}_2||\rangle \approx 1.49$ because of the dispersion of the patterns around their center of mass, hence we will use this value for the numerical computations below.

We represent the feedforward weight vector $\vec{w}_i$ onto a newborn DGC as an arrow which points below the surface of a sphere with radius $\langle||\vec{w}_1||\rangle$ (Figure 5.4). We compute analytically its L2-norm at the end of the early phase of maturation of the newborn DGC, as well as its angle

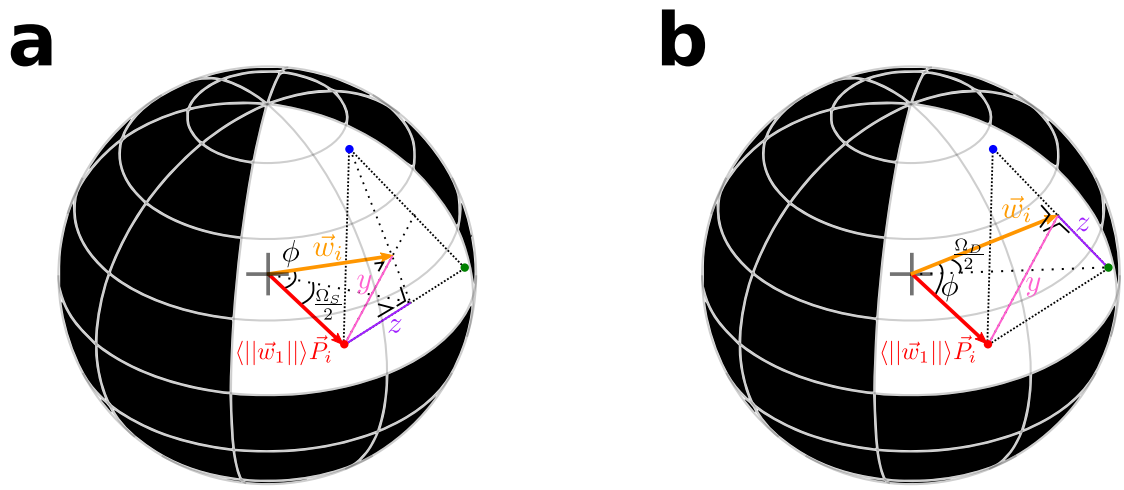

Figure 5.4 – Schematic for computation of the norm and angle. Schematic drawing for the analytical computation of the L2-norm of the weight vector $\vec{w}_i$ onto the newborn DGC at the end of the early phase of maturation, and its angle ϕ with the center of mass of the novel cluster, for (**a**) similar clusters ($\xi = 0.2$), and (**b**) distinct clusters ($\xi = 0.8$). The sphere has a radius $\langle||\vec{w}_1||\rangle$. The projection of the center of mass of the first two clusters (represented by the two mature DGCs) are represented by the blue and green dots. The red dot represents the projection of the center of mass of the novel cluster, $\vec{P}_i$, on the sphere.

ϕ with the center of mass of the novel cluster $\vec{P}_i$, to confirm the results obtained in Figure 5.3.

Similar clusters

The angle between the center of mass of any pair of similar clusters ($\xi = 0.2$) is given by equation (5.9):

$$\Omega_S = \arccos\left(\frac{1}{1+0.2^2}\right) \tag{5.14}$$

Half the distance between the projections of the center of mass of any pair of two similar clusters on a concentric sphere with radius $\langle||\vec{w}_1||\rangle$ is given by (Figure 5.4a):

$$z = \langle||\vec{w}_1||\rangle \cdot \sin\left(\frac{\Omega_S}{2}\right) \tag{5.15}$$

The triangle which connects the projections on the sphere of the center of masses of the three clusters is equilateral, and y separates one of its angle in two equal parts ($\pi/6$ [rad] each). So the length y can be calculated:

$$y = \frac{z}{\cos\left(\frac{\pi}{6}\right)} \tag{5.16}$$

Using Pythagoras formula, we can thus determine the expected L2-norm $\langle||\vec{w}_i||\rangle$ of the feedforward weight vector onto the newborn DGC at the end of the early phase of maturation:

$$\langle||\vec{w}_i||\rangle = \sqrt{\langle||\vec{w}_1||\rangle^2 - y^2}, \tag{5.17}$$

and finally its angle with the center of mass of the novel cluster:

$$\phi = \arccos\left(\frac{\langle||\vec{w}_i||\rangle}{\langle||\vec{w}_1||\rangle}\right) \tag{5.18}$$

The numerical values are: $\langle||\vec{w}_i||\rangle \approx 1.47$ and $\phi \approx 9.21[°]$, which correspond to the values on Figure 5.3a1.

Distinct clusters

In the case of distinct patterns ($\xi = 0.8$), the angle between the center of mass of any pair of clusters is given by equation (5.9):

$$\Omega_D = \arccos\left(\frac{1}{1+0.8^2}\right) > \Omega_S \tag{5.19}$$

We can directly compute the expected L2-norm of the feedforward weight vector onto the newborn DGC at the end of the early phase of maturation (Figure 5.4b):

$$\langle||\vec{w}_i||\rangle = \langle||\vec{w}_1||\rangle \cdot \cos\left(\frac{\Omega_D}{2}\right) \tag{5.20}$$

Using Pythagoras formula, we can then calculate the length z between the projection of the center of mass of one of the two pretrained clusters on a concentric sphere with radius $\langle||\vec{w}_1||\rangle$ and the feedforward weight vector onto the newborn DGC:

$$z = \sqrt{\langle||\vec{w}_1||\rangle^2 - \langle||\vec{w}_i||\rangle^2} \tag{5.21}$$

Analogous to the similar case, we observe that y separates one angle of the equilateral triangle connecting the projections of the center of mass of the clusters on the sphere in two equal parts, consequently:

$$y = \frac{z}{\tan\left(\frac{\pi}{6}\right)} \tag{5.22}$$

Finally, the angle between the center of mass of the novel cluster and the feedforward weight vector onto the newborn DGC at the end of the early phase of maturation is:

$$\phi = \arccos\left(\frac{\langle||\vec{w}_i||\rangle^2 + \langle||\vec{w}_1||\rangle^2 - y^2}{2\langle||\vec{w}_i||\rangle\langle||\vec{w}_1||\rangle}\right) \tag{5.23}$$

We obtain the following approximate values: $\langle||\vec{w}_i||\rangle \approx 1.34$ and $\phi \approx 47.2[^\circ]$, which correspond to the values on Figure 5.3b1. The angle ϕ is smaller in the similar case than in the distinct case, hence the norm is larger in the similar case, as observed in Figure 5.3a1,b1.

Part III Discussion and appendices

6 Discussion

While experimental results stemming from the manipulation of the ratio of NKCC1 to KCC2 suggest that the switch from excitation to inhibition of the GABAergic input onto adult-born DGCs is crucial for their integration into the preexisting circuit (Ge et al. (2006, 2007); Alvarez et al. (2016)) and that adult dentate gyrus neurogenesis promotes better pattern separation (Clelland et al. (2009); Sahay et al. (2011a); Jessberger et al. (2009)), the link between channel properties and behavior has remained puzzling (Sahay et al. (2011b); Aimone et al. (2011)). In this work, we have shown with a model that the switch enables newborn DGCs to become selective for novel features of presented stimuli which are similar to familiar, already stored, representations, consistent with the experimentally-observed function of pattern separation (Clelland et al. (2009); Sahay et al. (2011a); Jessberger et al. (2009)). Previous modeling studies already suggested that newborn DGCs integrate novel inputs into the representation in dentate gyrus (Chambers et al. (2004); Becker (2005); Crick and Miranker (2006); Wiskott et al. (2006); Chambers and Conroy (2007); Appleby and Wiskott (2009); Aimone et al. (2009); Weisz and Argibay (2009); Temprana et al. (2015)). However, they either (i) used an abstract framework of additive or turnover neurogenesis in autoencoders; or (ii) used classical algorithmic competitive Hebbian learning with weight vector normalization, thus bypassing the two-phase integration issue. To our knowledge, we present the first synaptic plasticity model that can explain both how adult-born DGCs integrate into the preexisting network and why they promote pattern separation of similar stimuli.

Our work emphasizes why a two-phase maturation of newborn DGCs is beneficial for proper integration in the preexisting network. From a computational perspective, the early phase of maturation, when GABAergic inputs onto newborn DGCs are excitatory, corresponds to cooperative unsupervised learning. Therefore, the synapses grow in the direction of patterns that indirectly activate the newborn DGCs (via GABAergic interneurons), see Figure 6.1a. At the end of the early phase of maturation, the receptive field of a newborn DGC represents the center of mass of all input patterns that led to its (indirect) activation. In the late phase of maturation, GABAergic inputs onto newborn DGCs become inhibitory, so that lateral interactions change from cooperation to competition, causing a shift of the receptive fields of

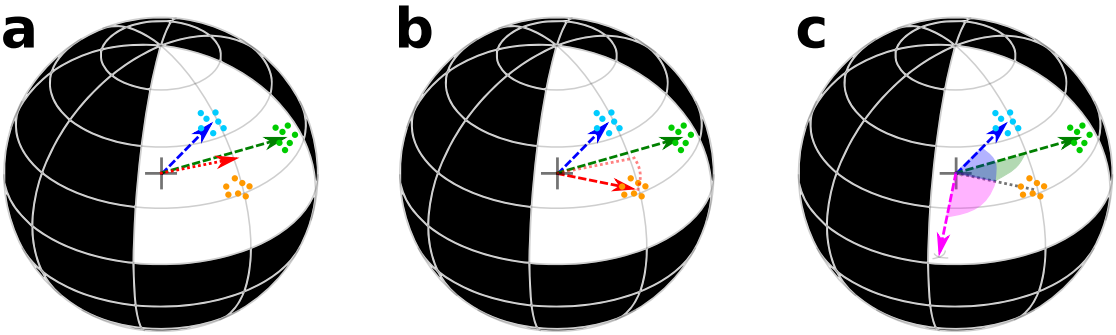

Figure 6.1 – Explanatory schematics. Schematics of the hypersphere surface that contains three clusters of patterns (colored dots) and the feedforward weight vectors towards the output units (colored arrows, ending slightly below the surface). The blue and green clusters are stored before patterns from the novel red cluster are presented to the network. (**a**) During the early phase of maturation, the newborn DGC grows its vector of feedforward weights (red arrow) in the direction of the subspace of active previously stored patterns. (**b**) During the late phase of maturation, the red vector moves away from previously stored patterns and towards the novel cluster. (**c**) Control scenario with an unresponsive neuron (with magenta feedforward weight vector), unable to learn the novel cluster.

the newborn DGCs towards novel features, see Figure 6.1b. At the end of maturation, newborn DGCs are thus selective for novel inputs. This integration mechanism is in agreement with the experimental observation that newborn DGCs are broadly tuned early in maturation, and get high input selectivity at the end of maturation (Marín-Burgin et al. (2012)). Loosely speaking, the cooperative phase of excitatory GABAergic input promotes the growth of the synaptic weights coarsely in the relevant direction, whereas the competitive phase of inhibitory GABAergic input helps to specialize on detailed, but potentially important differences between patterns.

In the context of theories of unsupervised learning, the switch of lateral GABAergic inputs to newborn DGCs from excitatory to inhibitory provides a biological solution to the "problem of unresponsive units" (Hertz et al. (1991)). Unsupervised competitive learning has been used to perform clustering of input patterns into a few categories (Rumelhart and Zipser (1985); Grossberg (1987a); Kohonen (1989); Hertz et al. (1991); Du (2010)). Ideally, after learning of the feedforward weights between an input layer and a competitive network, input patterns that are distinct from each other activate different neuron assemblies of the competitive network. After convergence of competitive Hebbian learning, the vector of feedforward weights of a given neuron points to the center of mass of the cluster of input patterns for which it is selective (Kohonen (1989); Hertz et al. (1991)). Yet, if the synaptic weights are randomly initialized, it is possible that the set of feedforward weights onto some neurons of the competitive network point in a direction "quasi-orthogonal" (Section 4.2.2) to the subspace of the presented input patterns, see Figure 6.1c. Therefore, those neurons called "unresponsive units" will never get active during pattern presentation. Different learning strategies have been developed in the field of artificial neural networks to avoid this problem (Grossberg (1976); Bienenstock

et al. (1982); Rumelhart and Zipser (1985); Grossberg (1987a); DeSieno (1988); Kohonen (1989); Hertz et al. (1991); Du (2010)). However, most of these algorithmic approaches lack a biological interpretation. In our model, the synapses onto newborn DGCs form spontaneously after neuronal birth. The excitatory GABAergic input in the early phase of maturation drives the growth of the synaptic weights in the direction of the subspace of presented patterns that succeeded in activating some of the mature DGCs. Hence the early cooperative phase of maturation can be seen as a smart initialization of the synaptic weights onto newborn DGCs, such that they end up close enough to the novel patterns to be able to become selective for them in the late competitive phase of maturation.

Our results are in line with the classic view that dentate gyrus is responsible for decorrelation of inputs (Marr (1969); Albus (1971); Marr (1971); Rolls and Treves (1998)), a necessary step for storage of similar memories in CA3, and with the observation that dentate gyrus lesions impair discrimination of similar but not distinct stimuli (Gilbert et al. (2001); Hunsaker and Kesner (2008)). To discriminate distinct stimuli, another pathway might be involved, such as the direct EC to CA3 connection (Fyhn et al. (2007); Vazdarjanova and Guzowski (2004)). Our theory for integration of newborn DGCs readily explains why enhanced adult dentate gyrus neurogenesis promotes better discrimination of similar stimuli, but not distinct stimuli, as observed experimentally (Clelland et al. (2009); Sahay et al. (2011a)). In our model, the early cooperative phase of maturation can only drive the growth of synaptic weights onto newborn cells if mature DGCs are activated by presented stimuli. Hence the stimuli should be similar enough to familiar stimuli that are already represented by the network. If only distinct stimuli are presented, none of the mature DGCs becomes active, so the newborn DGCs are not indirectly activated and their synaptic weights do not grow. Consequently, in an experimental paradigm with distinct patterns, newborn cells are, according to our model, silent. As they are poorly integrated into the preexisting circuit, they will probably not survive (Tashiro et al. (2006)).

Experimental observations support the importance of the switch from early excitation to late inhibition of the GABAergic input onto newborn DGCs. An absence of early excitation using NKCC1-knockout mice has been shown to strongly affect synapse formation and dendritic development in vivo (Ge et al. (2006)). Conversely, a reduction in inhibition in the dentate gyrus through decrease in KCC2 expression has been associated with epileptic activity (Pathak et al. (2007); Barmashenko et al. (2011)). An analogous switch of the GABAergic input has been observed during development, and its proper timing has been shown to be crucial for sensorimotor gating and cognition (Wang and Kriegstein (2010); Furukawa et al. (2017)). In addition to early excitation and late inhibition, our theory also critically depends on the duration of the switch. Indeed, it supposes that a sufficient number of newborn DGCs are just about -within a few hours- to switch the effect of their GABAergic input when novel inputs are presented, in order for them to become selective for new features of the environment. Several experimental results have suggested that the switch is indeed sharp and occurs within a single day, both during development (Khazipov et al. (2004); Tyzio et al. (2007); Leonzino et al. (2016)) and adult dentate gyrus neurogenesis (Heigele et al. (2016)). Furthermore, in hippocampal

cell cultures, expression of KCC2 is upregulated by GABAergic activity but not affected by glutamatergic activity (Ganguly et al. (2001)). A similar process during adult dentate gyrus neurogenesis would increase the number of newborn DGCs available for representing novel features by advancing the timing of their switch. In this way, instead of a few thousands of newborn DGCs ready to switch (3 to 6% of the whole population (Van Praag et al. (1999); Cameron and McKay (2001)), divided by 30 days), a larger fraction of newborn DGCs would be made available for coding, if appropriate stimulation occurs.

To conclude, our theory for integration of adult-born DGCs suggests that newborn cells have a coding – rather than a modulatory – role during dentate gyrus pattern separation function. Our theory highlights the importance of GABAergic input in adult dentate gyrus neurogenesis, and links the switch from excitation to inhibition to the integration of newborn DGCs into the preexisting circuit. Finally, it readily illustrates how Hebbian plasticity of the synapses from EC to DGCs makes newborn cells suitable to promote pattern separation of similar but not distinct stimuli, a long-standing question in the field of adult dentate gyrus neurogenesis (Sahay et al. (2011b); Aimone et al. (2011)).

7 My contributions

In adult dentate gyrus neurogenesis, the link between maturation of newborn neurons and their function, such as behavioral pattern separation, has remained puzzling. Besides, unsupervised competitive learning in ANN brought up the beneficial aspect of incorporation of new nodes in the network. But no consensus has been reached about the functional role of adult-born dentate granule cells. Moreover, the question of how new nodes integrate into the network has been avoided by all sorts of initializations of their weights.

In my work, I bridge the gap between biology and theory. I base my model on the accumulated experimental knowledge about the integration of adult dentate granule cells, and show that a solution to the problem of dead units in unsupervised competitive learning directly derives from it. By analyzing a theoretical model, I show that the switch from excitation to inhibition of the GABAergic input onto maturing newborn cells is crucial for their proper functional integration. In the early excitatory phase of GABAergic inputs, cooperativity makes the feedforward connections onto adult-born dentate granule cells grow in the direction of the subspace of previously presented inputs, thus preventing the newborn cells to become unresponding dead units. In the late inhibitory phase, competition kicks in, and drives the feedforward weights away from previously stored clusters. This enables adult-born dentate granule cells in my model to code for concepts that are novel, yet similar to familiar ones. To my knowledge, this theory of maturation of newborn cells is the first that can explain: (i) how adult-born dentate granule cells integrate into the preexisting dentate gyrus network, and (ii) why they promote pattern separation of similar stimuli.

A Determination of the plasticity parameters

To make sure that for a smaller θ, and hence a LTP-dominated regime, the L2-norm of the weight vectors at convergence is slightly bigger than for a higher θ, we choose $\beta = 1$, and set $\gamma = \gamma_0 - \theta$ with $\gamma_0 = 10$. To investigate the dynamics of the system around the high fixed point (LTP regime, where $\nu_i > \theta$), we compute the phase portrait for a simple case without inhibition, bias equal to zero, and $\tau_m = \eta = 1$, see Figure A.1a,b. The trajectories of the dynamical system for a single postsynaptic neuron with a single postsynaptic partner (hence y corresponds to ν_i and x to x_j for convenience) are given by:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -y + \tanh\left(\frac{[wx]_+}{L}\right)$$
$$\frac{\mathrm{d}w}{\mathrm{d}t} = y(y-\theta)\left(\gamma x - \beta w y^2\right)$$

The cases for $\theta = 0$ (Figure A.1a) and $\theta = 0.15$ (Figure A.1b) are represented, with the y-nullcline in red and the w-nullcline in blue. By construction, the y-nullcline corresponds to the neuronal gain function. The w-nullcline has a vertical component at $\nu_i = \theta$, and another component on which the high stable fixed point of the firing rate lies (at the intersection with the red nullcline). We observe on Figure A.1c,d that the bigger the θ, the smaller the w at convergence to the high stable fixed point, hence the smaller the L2-norm of the weight vector. Furthermore, we note that for this choice of parameters, θ has to be smaller than about 0.8, otherwise the high fixed point is lost, and thus all weights ultimately converge to zero.

In addition, the maximum extent of LTP and LTD can be computed. To do so, we determine y^* such that $\frac{\mathrm{d}g(\nu_i)}{\mathrm{d}\nu_i}|_{\nu_i = y^*} = 0$, and then compute $g(y^*)$. In the LTP regime ($\nu_i > \theta$), the maximum amount of LTP is reached at $\nu_i = y^*$. The expression of y^* is complicated because it depends on several parameters: $y^* = f(\theta, \delta, \gamma, x_j, w_{ij})$. Hence, we set δ and γ as above and fix the other parameters at different values, see Figure A.1e,f. Similarly, in the LTD regime ($\nu_i < \theta$), the maximum amount of LTD is reached at $y^* = \frac{\theta}{2}$, which corresponds to an LTD update of $\Delta w_{ij} = -\alpha x_j \frac{\theta^2}{4}$. To get amount of LTP and LTD of similar order of magnitude, we set $\alpha = \frac{\alpha_0}{\theta^3}$, see Figure A.2.

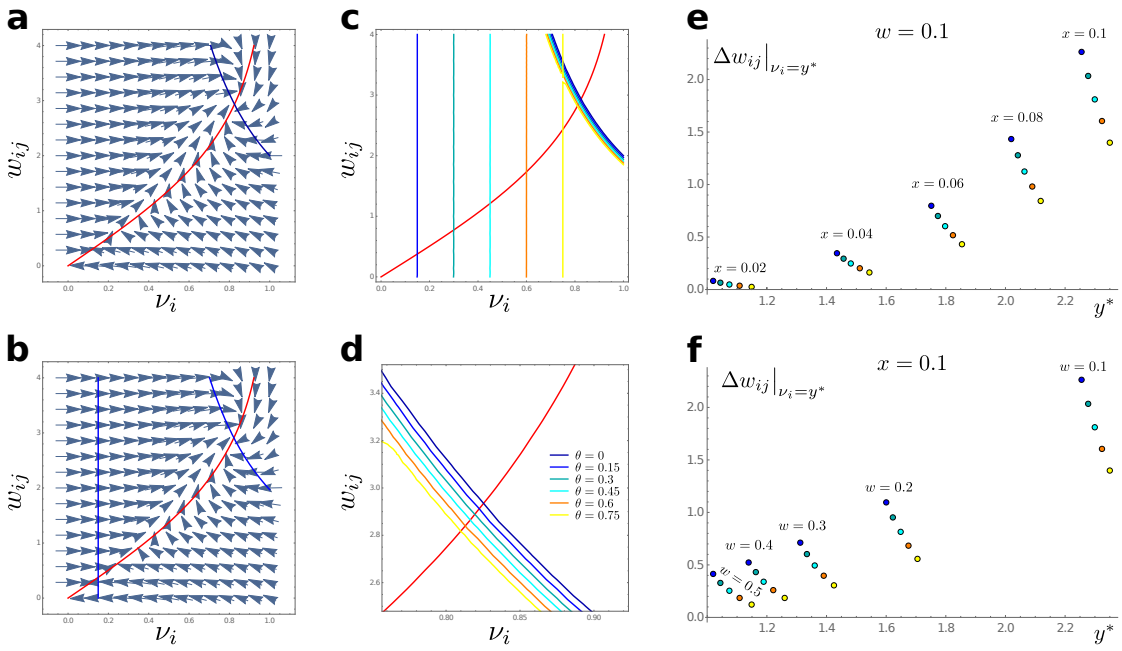

Figure A.1 – High stable fixed point and maximum LTP update. Dynamical trajectories with the y-nullcline in red and the w-nullcline in blue, for (**a**) $\theta = 0$ and (**b**) $\theta = 0.15$. (**c**) w-Nullclines for different values of θ, and (**d**) zoom in around the high stable fixed point. Maximum LTP update $\Delta w_{ij}|_{\nu_i=y^*}$ and corresponding postsynaptic firing rate value y^* as a function of θ and (**e**) the presynaptic rate x, or (**f**) the weight w.

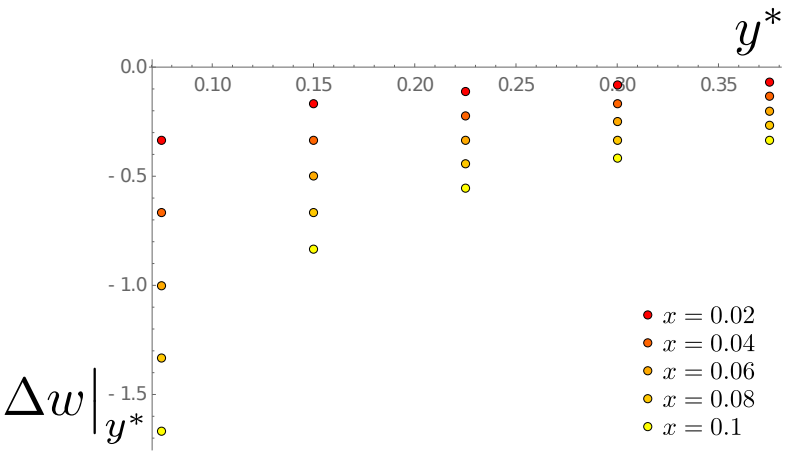

Figure A.2 – Maximum LTD update. Location (y^*) and value of the maximal LTD update ($\Delta w|_{y^*}$) for different values of the presynaptic firing rate x, and threshold values θ (keep in mind that $y^* = \frac{\theta}{2}$).

B Single Winner-Take-All network

Initially, the theory for newborn DGCs integration was implemented in a simple network to thoroughly investigate the conceptual aspects of the model before moving to a more biologically plausible network, which is presented in the main text. The original network was a single WTA network with Oja's learning rule for the feedforward connections.

B.1 Methods

B.1.1 The network of mature DGCs as a WTA network

In our model, we describe the network of mature DGCs as a WTA circuit consisting of N_{DGC} laterally coupled neurons. The WTA network has the task to represent K clusters of input patterns. For each input pattern $\vec{x}$, only one of the N_{DGC} neurons is activated. An input pattern consists of the activity of N_{EC} neurons in EC. We assume that inhibition controls EC activity such that $||\vec{x}|| = 1$.

The input layer (EC neurons) is fully connected to the layer of DGCs neurons. The feedforward connections from EC to DGCs are excitatory (weights $w_{ij}^{ff} \geqslant 0$). They are initialized randomly $w_{ij}^{ff} \in [0,1]$, and normalized such that $\sum_j \left(w_{ij}^{ff}\right)^2 = 1$. In rodent hippocampus, the spiking of mature DGCs activates interneurons in dentate gyrus, which in turn inhibit other mature DGCs (Leutgeb et al. (2007); Alvarez et al. (2016)). In our model, we replace the indirect DGC-to-interneuron-to-DGC connection by a direct inhibitory DGC-to-DGC connection: the output neurons are laterally connected all-to-all with identical inhibitory weights w^{rec}. No neuron in the network has self-connections.

The total input to output neuron i is given by:

$$I_i = \sum_{j=1}^{N_{EC}} w_{ij}^{ff} x_j + \sum_{k=1}^{N_{DGC}} m_{ik} w_{ik}^{rec} v_k \tag{B.1}$$

with $\vec{x} = \left(x_1, x_2, ..., x_{N_{EC}}\right)$ the input pattern, and ν_k the firing rate of the output neuron k. In a mature network, the elements m_{ik} are -1 (inhibitory lateral connections). W^{ff} is the feedforward connectivity matrix. W^{rec} is the recurrent connectivity matrix with diagonal elements $w_{ii}^{rec} = 0$ and off-diagonal elements $w_{ik}^{rec} = w^{rec} > 0$.

Before a stimulus is presented, all rates are initialized at zero. Upon stimulation, the lowpass-filtered firing rate of output neuron i evolves according to (Miller and Fumarola (2012)):

$$\tau \frac{\mathrm{d}\nu_i}{\mathrm{d}t} = -\nu_i + f(I_i) \tag{B.2}$$

with τ a time constant, and f a Heaviside function

$$f(z) = \mathscr{H}(z - b) \tag{B.3}$$

with some bias b.

Once the firing rates in the output layer have converged, the feedforward weights from each EC neuron j to each DGC i are updated in our model according to Oja's rule (Oja (1982)):

$$\Delta w_{ij}^{ff} = \eta \nu_i \left(x_j - w_{ij}^{ff} \nu_i\right) \tag{B.4}$$

with η a learning rate. Since in a WTA circuit all neurons except one have a rate $\nu_i = 0$, only the winning unit ($\nu_i = 1$) updates its inwards feedforward weights in direction of the input pattern $\vec{x}$. Note that the weight update rule (B.4) is applied once for each pattern presentation, after convergence of equation (B.2) to a stationary state. After the weight updates, all firing rates are reset to zero. Therefore, the WTA network is ultimately performing clustering: each output unit feedforward weight vector will move towards the center of mass of the cluster of patterns it is selective for (Hertz et al. (1991)).

Winners, losers, and quasi-orthogonal inputs

We define the winner as the neuron for which the scalar product of feedforward weights with the input pattern is the largest:

$$\max_i I_i = \max_i \left[\sum_{j=1}^{N_{EC}} w_{ij}^{ff} x_j \right] \tag{B.5}$$

Since both the input patterns and the feedforward weight vectors are normalized to have an L2-norm of 1, the winning unit is the one whose weight vector $\vec{W}_i^{ff}$ (row of the feedforward connectivity matrix) aligns best with the input pattern $\vec{x}$. Furthermore, we say that an input pattern $\vec{x}$ is "quasi-orthogonal" to a weight vector $\vec{W}_i^{ff}$ if $\sum_{j=1}^{N_{EC}} w_{ij}^{ff} x_j < b$. Note that for $b \to 0$,

we recover the standard orthogonality condition. Under the condition

$$w^{rec} + b > 1, \tag{B.6}$$

we get at most one active neuron in the output layer. In addition, if the condition $\sum_{j=1}^{N_{EC}} w_{ij}^{ff} x_j > b$ is satisfied for at least one neuron, then the winner will become active and thus will inhibit the other output neurons through the inhibitory recurrent connections w^{rec} of equation (B.1).

To prove the above claims we assume that we have inhibitory lateral connections ($m_{ik} = -1$). At steady state we obtain $\nu_i = f(I_i) \ \ \forall i$, cf. equation (B.2). Assuming n active neurons i^*, we get

$$\nu_{i^*} = \mathscr{H}\left(\sum_{j=1}^{N_{EC}} w_{i^*j}^{ff} x_j - (n-1)\, w^{rec} - b\right) \tag{B.7}$$

for the active neurons, and

$$\nu_i = \mathscr{H}\left(\sum_{j=1}^{N_{EC}} w_{ij}^{ff} x_j - n w^{rec} - b\right) \tag{B.8}$$

for the inactive neurons. Hence $\sum_{j=1}^{N_{EC}} w_{i^*j}^{ff} x_j > b$ is a necessary condition for the winner. Furthermore, knowing that $||\vec{x}|| = 1$ (by construction), and that $||\vec{W}_i^{ff}|| \leqslant 1 \quad \forall i$ (the feedforward weights are initialized such that $||\vec{W}_i^{ff}|| = 1$, and Oja's rule enforces that $||\vec{W}_i^{ff}||$ will stay smaller or equal to 1 during learning), the condition $w^{rec} + b > 1$ is sufficient to ensure that there is only one active neuron. As we initialize all output firing rates to 0 before pattern presentation, equation (B.1) is initially $I_i = \sum_{j=1}^{N_{EC}} w_{ij}^{ff} x_j$, and therefore the winner of the competition is the output unit i^* with the largest input I_{i^*}.

Figure B.1b shows the evolution of the firing rate of four output neurons during presentation of an input pattern. In this example, three units initially show an increasing firing rate, because the total input I_i they receive is bigger than the bias b, see Figure B.1a, whereas the fourth (orange) unit is quasi-orthogonal to the input pattern, and thus stays silent the whole time. With time, as activity of the output neurons increases, lateral inhibition increases as well. It has the effect of lowering the total input to all neurons, and consequently the firing rate. The blue unit, which is the one whose feedforward weight vector matches best the input pattern, ultimately wins the competition, while input to the purple and the yellow units falls below the firing threshold.

Two exceptional cases can arise: (1) all inward weight vectors towards the output units are quasi-orthogonal to the input pattern $\left(\sum_{j=1}^{N_{EC}} w_{ij}^{ff} x_j < b \ \ \forall i\right)$. In this first case, all output units will stay silent ($\nu_i = 0 \ \ \forall i$). However, the more output units we have, the lower the probability that this situation arises since the feedforward weight vectors are initially randomly chosen from a uniform distribution on the positive quadrant of the surface of the hypersphere. (2) The scalar product between the weight vector and the input pattern is identical for several output

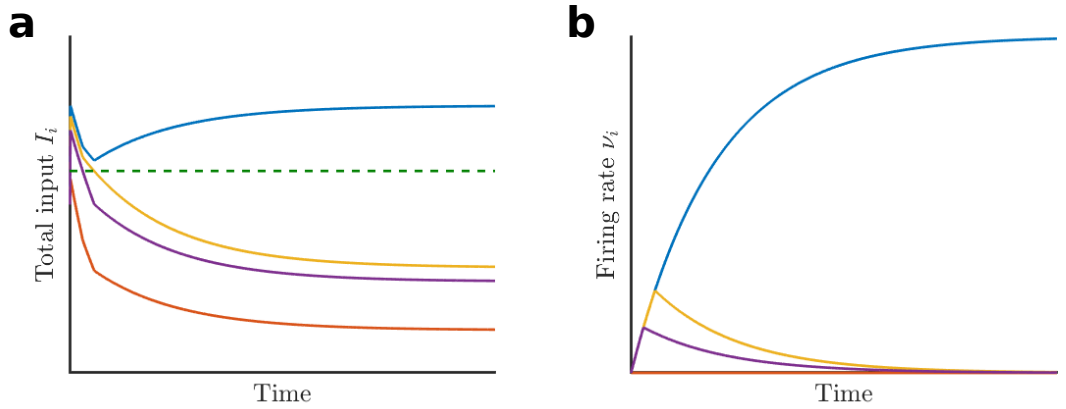

Figure B.1 – Activity dynamics in a single WTA network. Activity dynamics of four DGCs ($i \in \{1,2,3,4\}$) during presentation of an input pattern. The four output units have different feedforward weight vectors. (**a**) Total input I_i towards each output neuron (the green dotted line represents the bias b), and (**b**) resulting firing rate ν_i as a function of time.

units. If the output units that share the same identical scalar product with the input patterns are also the ones with largest scalar product, it will follow that all of them will become winners of the competition. Yet, this case becomes more and more unlikely as the dimension of the input patterns is increased. Moreover, such a symmetric solution of equations (B.1) (B.2) (B.3) is dynamically unstable in the presence of noise.

B.1.2 Neurogenesis and integration of newborn neurons

Before neurogenesis occurs, we store K prototype patterns in the existing network of DGCs ($N_{DGC} = K$). To do so, the feedforward weights are initialized randomly $w_{ij}^{ff} \in [0,1]$, and normalized such that $\sum_j \left(w_{ij}^{ff}\right)^2 = 1$. Patterns of the K different clusters are then presented to the network, and the feedforward weights are updated as explained in section B.1.1. At the end of pretraining, each output unit has a feedforward weight vector that aligns with the center of mass of one of the clusters Hertz et al. (1991), see Figure B.2a.

The feedforward weights towards the mature DGCs are then frozen, and a newborn DGC is added to the output layer. This corresponds to the neurogenesis step. The feedforward connections to this newborn DGC are plastic and initialized to small positive values: $w_{ij}^{ff} \sim |\mathcal{N}(0, 10^{-4})|$. The lateral connections onto a newborn cell are set to the same value w^{rec} as for the other DGCs, while the outward connections from a newborn cell to mature cells are set to 0. Once the newborn DGC has been added to the network, a cluster of novel patterns is presented, intermingled with the K previously stored clusters. Thus we present $K+1$ clusters of patterns onto K mature DGCs and one newborn DGC. Learning of the feedforward weights onto the newborn DGC occurs in two phases.

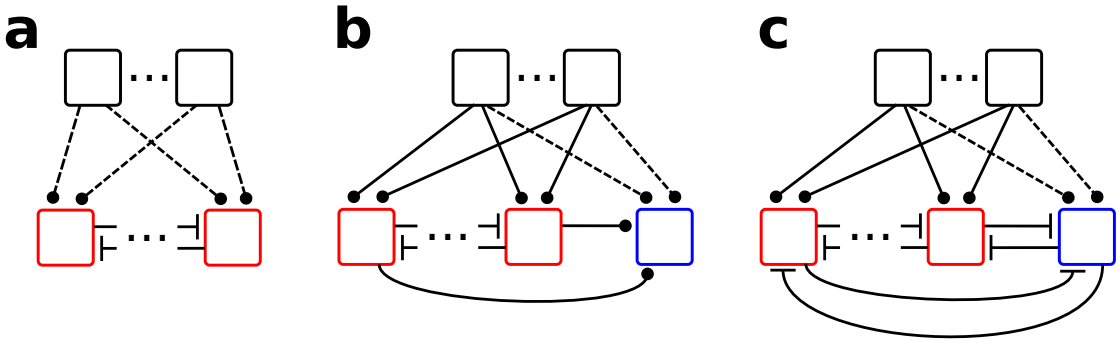

Figure B.2 – Architecture of the single WTA network. The entorhinal cortex input layer (black) is fully feedforwardly connected to the output layer, consisting of mature DGCs (red) and a newborn DGC (blue). (**a**) Pretraining: input patterns are presented to a network of DGCs, and all feedforward connections are learned. (**b**) Early phase of maturation of the newborn DGC: lateral input from mature DGCs (red) has an excitatory effect onto the newborn DGC (blue). (**c**) Late phase of maturation: all lateral connections are inhibitory. Full lines depict fixed connections, dashed lines represent plastic connections; round endings mean excitation, while T endings represent inhibition.

Early maturation phase: cooperative

During the early phase of maturation, the GABAergic input onto a newborn cell has an excitatory effect, see Figure B.2b. In the model, using the subscript notation m for mature DGCs and n for the newborn DGC, it is implemented as follows: $m_{nm} = +1$ (and $m_{mm} = -1$ for two different mature DGCs, as before). Furthermore, since newborn cells do not recruit feedback inhibition onto the mature DGCs, we have $m_{mn} = 0$.

After each pattern presentation, the dynamics of the firing rate of the output units evolve until convergence as before. However, the newborn DGC receives indirect lateral excitation, instead of inhibition, from the mature DGCs. Based on equation (B.1) and the elements $m_{nm} = 1$, the total input towards the newborn DGC is given by:

$$I_n = \sum_j w_{nj}^{ff} x_j + \sum_{k \neq n} w_{nk}^{rec} v_k \tag{B.9}$$

where w_{nj}^{ff} (resp. w_{nj}^{rec}) stands for the feedforward (resp. recurrent) connectivity weight from neuron j to the newborn DGC with index n. If we assume that the conditions to have a single active neuron among the mature DGCs are satisfied, the newborn neuron is indirectly excited by the winner of the competition. The necessary condition for the lateral excitatory GABAergic input to be sufficient to activate the newborn neuron is: $w^{rec} > b$. If this requirement is met, the activation of the newborn DGC drives the growth of its feedforward weights. Therefore, together with condition (B.6), we require $w^{rec} > \frac{1}{2}$ and $w^{rec} > b$.

From the perspective of the newborn DGC, the early phase is cooperative because, for each

pattern presentation, the winning mature DGC laterally excites the newborn DGC. Since in our model mature DGCs have non-plastic connections, the winning mature DGC will not update its feedforward weights, but the newborn DGC will. Under these conditions, the feedforward weights onto the newborn DGC will grow in the direction of the subspace of the feedforward weights of those output units that win the competition for some input patterns. In other words, they will move towards the subspace of the weight vectors of DGCs that are winners for the presented patterns, hence the learning is cooperative.

Late maturation phase: competitive

During the second phase of maturation, the GABAergic input onto the newborn DGC switches from excitatory to inhibitory. In terms of our model, it means that m_{nm} switches from +1 to −1. Equation (B.9) becomes:

$$I_n = \sum_j w_{nj}^{ff} x_j - \sum_{k \neq n} w_{nk}^{rec} v_k \tag{B.10}$$

which is equivalent to the input onto the mature DGCs, hence competitive. Furthermore, in this second phase, inhibitory lateral connections from the newborn DGC to the mature DGCs are added ($m_{mn} = -1$), see Figure B.2c.

B.1.3 Control case

In the control case, we start directly with a number of output neurons equal to the final number of clusters of patterns, that is $N_{EC} = K + 1$. The feedforward connections to all DGCs are plastic. As for the mature DGCs of the neurogenesis model, they are initialized randomly $w_{ij}^{ff} \in [0,1]$, and normalized such that $\sum_j \left(w_{ij}^{ff}\right)^2 = 1$. The DGCs are all-to-all identically inhibitory recurrently connected with value w^{rec}. There are no self-connections. The pretraining, where only K pattern clusters are shown to the network, is identical to the one for the neurogenesis case, except there are $K+1$ output DGCs with plastic feedforward connections instead of K. Once we add the novel cluster of patterns, we keep all feedforward connections towards the DGCs plastic. So the network is equivalent to the one in Figure B.2a, except that there are always $K+1$ output neurons.

B.1.4 Classification performance

After convergence of the feedforward weights, they are fixed, and classification performance of the DGC layer is assessed by computing classification error on the corresponding test set (Diehl and Cook (2015)). First, every DGC neuron of the output layer is assigned its corresponding label (the cluster it represents best). This is done by presenting each pattern of the training set to the network. The index of the output neuron which wins the competition is kept in memory together with the known label of the pattern. Once all training patterns have been presented,

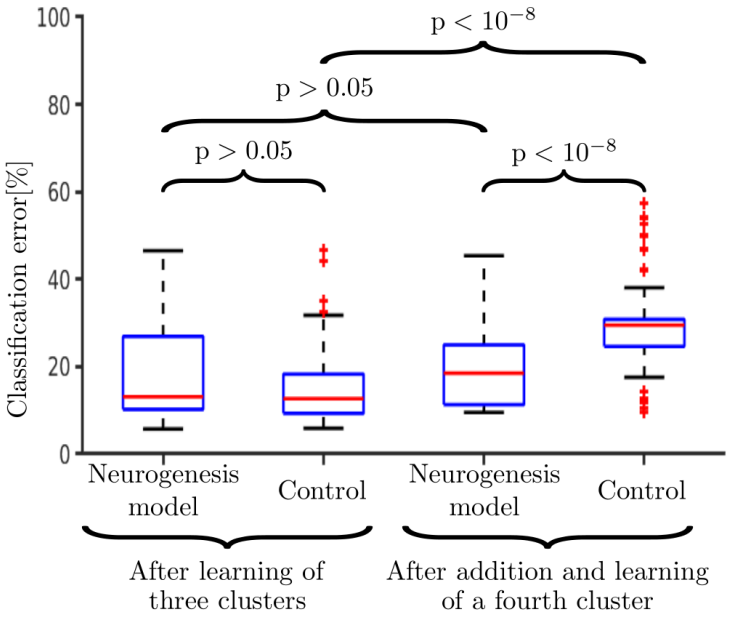

Figure B.3 – Classification error with the single WTA network Boxplots representing the classification error on the test set after learning using the training set (MNIST 12x12 dataset). Levene's test confirmed that the data came from distributions with similar variances, so a two-sample t-test was used to test if the data from two different distributions had similar means or not: the p-values are displayed.

each output neuron is assigned the label of the digit to which it responded most often. Second, all patterns of the testing set whose clusters have been presented during learning are classified by observing which output neuron wins the competition, and comparing its assigned label to the real label of the pattern. The classification error is computed by dividing the number of incorrectly classified patterns over the total number of presented patterns.

B.2 Results

B.2.1 Maturation of a newborn DGC in a WTA network of mature DGCs is better for learning a novel cluster than a population of same size of plastic neurons

We wondered whether a network that was previously trained on three clusters of input patterns was able to learn a novel cluster. In the control network, we used four DGCs with input from 144 EC cells. In the scenario with neurogenesis, we started with three DGCs to which we added a fourth one when patterns from the novel cluster were included. The four clusters of EC inputs correspond to training patterns belonging to four digits of the MNIST 12x12 dataset.

First, the classification error at the end of the learning phase with three clusters is similar

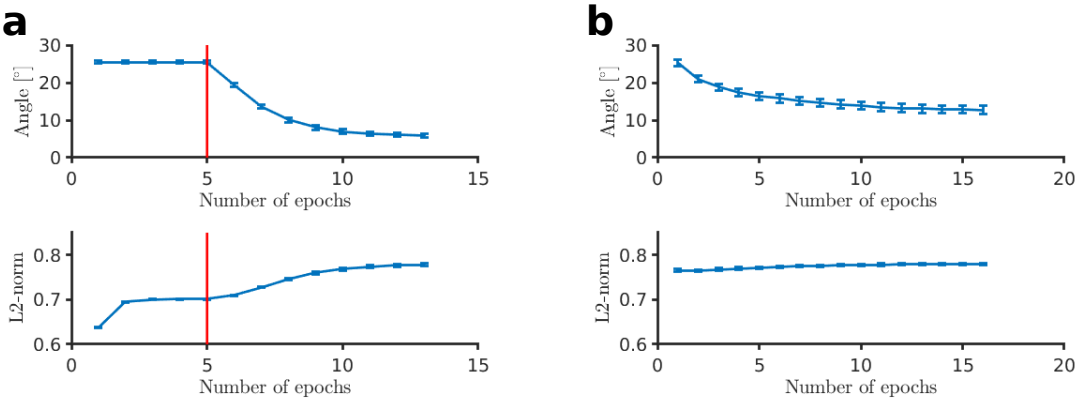

Figure B.4 – Evolution of the angle and norm in the single WTA network. Evolution of the angle between the center of mass of the novel cluster and the feedforward weight vector onto the output neuron of interest (top row), and of the norm of that feedforward weight vector (bottom row). (**a**) Neurogenesis case: evolution of the feedforward weight vector onto the newborn DGC. The red vertical line represents the switch from excitation to inhibition of the GABAergic input onto the newborn DGC. (**b**) Control case: evolution of the feedforward weight vector onto the output unit that ultimately represents the novel cluster, if there is one. One epoch corresponds to the presentation of all the training patterns of the selected clusters once.

for the neurogenesis model and the control, see Figure B.3. Since in the control there are four output units to store only three clusters, the classification error is slightly lower, but the difference is not significant. Second, after presentation of a novel cluster, the network performs significantly better in the neurogenesis model than in the control. This is reflected by a significant increase of the classification error upon inclusion of a novel cluster in the control, but not in the neurogenesis model. We conclude that concurrently including a novel cluster and adding a plastic output unit is beneficial for learning a novel cluster. Even though in the control case all output units always keep plastic feedforward weights, a novel cluster included at a later learning stage is not properly learned.

Next, we asked why there is such a difference in classification performance between the neurogenesis model and the control. We focused on the evolution of the feedforward weight vector onto the newborn DGC as a function of maturation. Similarly, in the control, we looked at the evolution of the feedforward weight vector onto the output unit that ultimately represents the novel cluster (if there is one). First, as expected by design, we observe that the norm of the weight vector is relatively stable in the control, while it increases as a function of maturation in the neurogenesis model, see bottom row of Figure B.4. Second, and more importantly, the angle between the weight vector and the center of mass of the novel cluster decreases more and faster in the neurogenesis model than in the control, see top row of Figure B.4. Consequently, when the weight vector has converged, it is more selective for the novel cluster in the neurogenesis model than in the control.

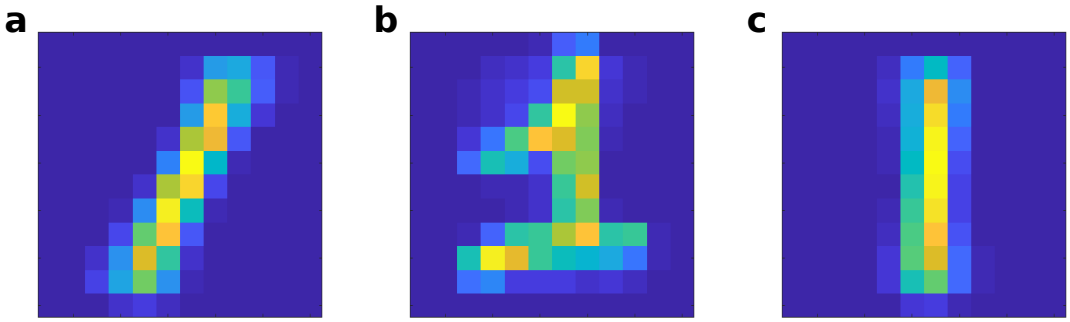

Figure B.5 – Pretrained subclusters of digit one of the MNIST 12x12 training set. (**a**) represents 48% of the training set. (**b**) represents 1% of the training set. (**c**) represents 51% of the training set.

B.2.2 Neurogenesis is a biological solution to the problem of dead units

In the control case, all output units have plastic feedforward weight vectors with a norm close to one. At the end of the learning of the first three clusters, they are each represented by at least one output unit. The fourth output unit is thus either unselective to any of the pretrained clusters (dead unit), or is also selective for one of the three clusters (subclustering of one cluster). Thus, there are two situations where the novel cluster cannot be learned. (1) The dead unit is quasi-orthogonal to the novel cluster, see Figure 6.1c. Most probably, the magenta unit was not updated during pretraining, because the blue and green vectors always won the competition. Then, at the end of learning of the blue and green clusters, the magenta unit still points where it got initialized. When the novel red cluster is added, ideally the magenta unit should become selective for it. However, if the magenta weight vector is quasi-orthogonal to the novel cluster, it will never be activated, and thus stay a dead unit. (2) All units are already assigned to the initial clusters (subclustering). The blue cluster got divided into two subclusters during pretraining: two weight vectors are pointing in its direction. Therefore, when the novel red cluster of patterns is added, it may be quasi-orthogonal to all weight vectors, and thus the novel cluster is never learned. Another possibility is that the novel cluster might actually activate one of the output units. So the feedforward weight vector will move in its direction. However, presentation of any pattern from the corresponding pretrained cluster will bring back the weight vector towards it. In all these cases, classification performance over all presented clusters is impaired, because no output unit represents the novel cluster.

B.2.3 Similar clusters can be learned, while distinct clusters cannot

Depending on how the network parameters are chosen, the newborn DGC is either able to become selective for the novel cluster, or not. For example, we can start by pretraining three subclusters of digit one in three mature DGCs, see Figure B.5. After convergence of the weights, we fix them. Then, we include a novel digit cluster in the presentation, and add a newborn DGC to the network whose feedforward weights are learned. At the end of maturation, we look

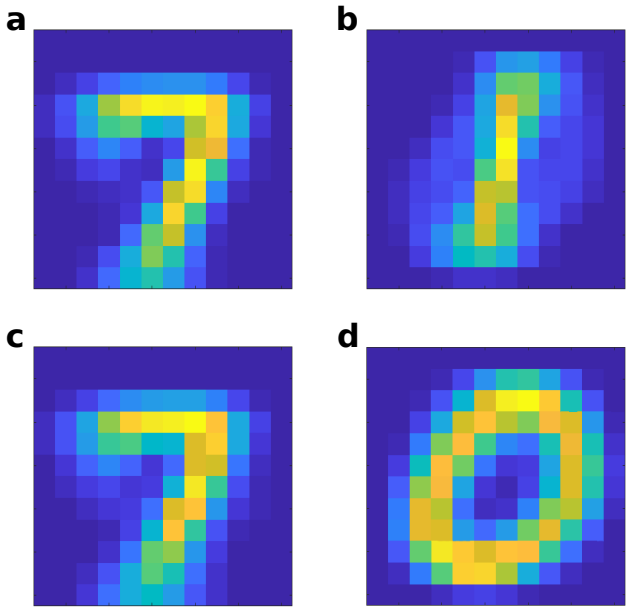

Figure B.6 – Receptive field of the newborn DGC at the end of maturation. The feedforward weight vector onto the newborn DGC is visualized as a 2-dimensional arrangement of its elements. (**a**) The novel cluster is similar (digit seven). (**b**) The novel cluster is distinct (digit zero). (**c**) Center of mass of all patterns of the training set that are labeled as digit seven. (**d**) Center of mass of all patterns of the training set that are labeled as digit zero.

at the feedforward weight vector to the newborn DGC, see Figure B.6. We observe that when the novel cluster is similar (digit seven) to the pretrained clusters, the newborn DGC becomes selective for it, see Figure B.6a. On the other hand, if the novel cluster (digit zero) is distinct, the newborn DGC does not become selective for it, see Figure B.6b. It is reflected by a lower classification error on the test set in the similar case (8.51%) with respect to the distinct case (24.30%).

Bibliography

Acsády, L., Kamondi, A., Sík, A., Freund, T., and Buzsáki, G. (1998). GABAergic cells are the major postsynaptic targets of mossy fibers in the rat hippocampus. *Journal of neuroscience*, 18(9):3386–3403.

Aimone, J. B., Deng, W., and Gage, F. H. (2011). Resolving new memories: a critical look at the dentate gyrus, adult neurogenesis, and pattern separation. *Neuron*, 70(4):589–596.

Aimone, J. B., Wiles, J., and Gage, F. H. (2006). Potential role for adult neurogenesis in the encoding of time in new memories. *Nature neuroscience*, 9(6):723–727.

Aimone, J. B., Wiles, J., and Gage, F. H. (2009). Computational influence of adult neurogenesis on memory encoding. *Neuron*, 61(2):187–202.

Albus, J. (1971). A theory of cerebellar function. *J. Mathematical Biosciences*, 10:25–61.

Altman, J. (1962). Are new neurons formed in the brains of adult mammals? *Science*, 135(3509):1127–1128.

Altman, J. and Das, G. D. (1965). Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *Journal of Comparative Neurology*, 124(3):319–335.

Altman, J. and Das, G. D. (1967). Postnatal neurogenesis in the guinea-pig. *Nature*, 214(5093):1098–1101.

Alvarez, D. D., Giacomini, D., Yang, S. M., Trinchero, M. F., Temprana, S. G., Büttner, K. A., Beltramone, N., and Schinder, A. F. (2016). A disynaptic feedback network activated by experience promotes the integration of new granule cells. *Science*, 354(6311):459–465.

Amaral, D. G., Scharfman, H. E., and Lavenex, P. (2007). The dentate gyrus: fundamental neuroanatomical organization (dentate gyrus for dummies). *Progress in brain research*, 163:3–22.

Anacker, C. and Hen, R. (2017). Adult hippocampal neurogenesis and cognitive flexibility—linking memory and mood. *Nature Reviews Neuroscience*, 18(6):335–346.

Andersen, P., Morris, R., Amaral, D., Bliss, T., and O'Keefe, J., editors (2007). *The hippocampus book*. Oxford university press.

Appleby, P. A., Kempermann, G., and Wiskott, L. (2011). The role of additive neurogenesis and synaptic plasticity in a hippocampal memory model with grid-cell like input. *PLoS computational biology*, 7(1):1–15.

Appleby, P. A. and Wiskott, L. (2009). Additive neurogenesis as a strategy for avoiding interference in a sparsely-coding dentate gyrus. *Network: Computation in Neural Systems*, 20(3):137–161.

Artola, A., Bröcher, S., and Singer, W. (1990). Different voltage dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, 347:69–72.

Barmashenko, G., Hefft, S., Aertsen, A., Kirschstein, T., and Köhling, R. (2011). Positive shifts of the $GABA_A$ receptor reversal potential due to altered chloride homeostasis is widespread after status epilepticus. *Epilepsia*, 52(9):1570–1578.

Bear, M. and Malenka, R. C. (1994). Synaptic plasticity: LTP and LTD. *Current Opinion in Neurobiology*, 4:389–399.

Becker, S. (2005). A computational principle for hippocampal learning and neurogenesis. *Hippocampus*, 15:722–738.

Ben-Ari, Y. (2002). Excitatory actions of GABA during development: the nature of the nurture. *Nature Reviews Neuroscience*, 3(9):728–739.

Biebl, M., Cooper, C. M., Winkler, J., and Kuhn, H. G. (2000). Analysis of neurogenesis and programmed cell death reveals a self-renewing capacity in the adult rat brain. *Neuroscience letters*, 291(1):17–20.

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

Boldrini, M., Fulmore, C. A., Tartt, A. N., Simeon, L. R., Pavlova, I., Poposka, V., Rosoklija, G. B., Stankov, A., Arango, V., Dwork, A. J., Hen, R., and Mann, J. J. (2018). Human hippocampal neurogenesis persists throughout aging. *Cell stem cell*, 22(4):589–599.

Boss, B. D., Peterson, G. M., and Cowan, W. M. (1985). On the number of neurons in the dentate gyrus of the rat. *Brain research*, 338(1):144–150.

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

Carpenter, G. and Grossberg, S. (1988). The art of adaptive pattern recognition by a self-organizing neural network. *Computer*, 21:77–88.

Caviness Jr, V. (1973). Time of neuron origin in the hippocampus and dentate gyrus of normal and reeler mutant mice: an autoradiographic analysis. *Journal of Comparative Neurology*, 151(2):113–119.

Chambers, R. A. and Conroy, S. K. (2007). Network modeling of adult neurogenesis: shifting rates of neuronal turnover optimally gears network learning according to novelty gradient. *Journal of cognitive neuroscience*, 19(1):1–12.

Chambers, R. A., Potenza, M. N., Hoffman, R. E., and Miranker, W. (2004). Simulated apoptosis/neurogenesis regulates learning and memory capabilities of adaptive neural networks. *Neuropsychopharmacology*, 29(4):747–758.

Chancey, J. H., Adlaf, E. W., Sapp, M. C., Pugh, P. C., Wadiche, J. I., and Overstreet-Wadiche, L. S. (2013). GABA depolarization is required for experience-dependent synapse unsilencing in adult-born neurons. *Journal of Neuroscience*, 33(15):6614–6622.

Chapouton, P., Jagasia, R., and Bally-Cuif, L. (2007). Adult neurogenesis in non-mammalian vertebrates. *Bioessays*, 29(8):745–757.

Chawla, M., Guzowski, J., Ramirez-Amaya, V., Lipa, P., Hoffman, K., Marriott, L., Worley, P., McNaughton, B., and Barnes, C. A. (2005). Sparse, environmentally selective expression of Arc RNA in the upper blade of the rodent fascia dentata by brief spatial experience. *Hippocampus*, 15(5):579–586.

Chistiakova, M., Bannon, N., Bazhenov, M., and Volgushev, M. (2014). Heterosynaptic plasticity: Multiple mechanisms and multiple roles. *The Neuroscientist*, 20(5):483–498.

Clelland, C., Choi, M., Romberg, C., Clemenson, G., Fragniere, A., Tyers, P., Jessberger, S., Saksida, L., Barker, R., Gage, F., and Bussey, T. (2009). A functional role for adult hippocampal neurogenesis in spatial pattern separation. *Science*, 325(5937):210–213.

Crick, C. and Miranker, W. (2006). Apoptosis, neurogenesis, and information content in hebbian networks. *Biological cybernetics*, 94(1):9–19.

Dayer, A. G., Ford, A. A., Cleaver, K. M., Yassaee, M., and Cameron, H. A. (2003). Short-term and long-term survival of new neurons in the rat dentate gyrus. *Journal of Comparative Neurology*, 460(4):563–572.

De Paola, V., Holtmaat, A., Knott, G., Song, S., Wilbrecht, L., Caroni, P., and Svoboda, K. (2006). Cell type-specific structural plasticity of axonal branches and boutons in the adult neocortex. *Neuron*, 49(6):861–875.

Debanne, D., Gähwiler, B., and Thompson, S. (1998). Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *Journal of Physiology*, 507:237–247.

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.

Deng, W., Aimone, J. B., and Gage, F. H. (2010). New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nature reviews neuroscience*, 11(5):339–350.

Deshpande, A., Bergami, M., Ghanem, A., Conzelmann, K.-K., Lepier, A., Götz, M., and Berninger, B. (2013). Retrograde monosynaptic tracing reveals the temporal evolution of inputs onto new neurons in the adult dentate gyrus and olfactory bulb. *Proceedings of the National Academy of Sciences*, 110(12):1152–1161.

DeSieno, D. (1988). Adding a conscience to competitive learning. In *IEEE international conference on neural networks*, volume 1, pages 117–124. Institute of Electrical and Electronics Engineers New York.

Diehl, P. U. and Cook, M. (2015). Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Frontiers in Computational Neuroscience*, 9.

Dieni, C. V., Nietz, A. K., Panichi, R., Wadiche, J. I., and Overstreet-Wadiche, L. (2013). Distinct determinants of sparse activation during granule cell maturation. *Journal of Neuroscience*, 33(49):19131–19142.

Du, K.-L. (2010). Clustering: A neural network approach. *Neural networks*, 23(1):89–107.

Eriksson, P. S., Perfilieva, E., Björk-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.

Espósito, M. S., Piatti, V. C., Laplagne, D. A., Morgenstern, N. A., Ferrari, C. C., Pitossi, F. J., and Schinder, A. F. (2005). Neuronal differentiation in the adult hippocampus recapitulates embryonic development. *Journal of Neuroscience*, 25(44):10074–10086.

Faulkner, R. L., Jang, M.-H., Liu, X.-B., Duan, X., Sailor, K. A., Kim, J. Y., Ge, S., Jones, E. G., Ming, G.-l., Song, H., and Cheng, H.-J. (2008). Development of hippocampal mossy fiber synaptic outputs by new neurons in the adult brain. *Proceedings of the National Academy of Sciences*, 105(37):14157–14162.

Fritzke, B. (1995). A growing neural gas network learns topologies. *Advances in neural information processing systems*, 7:625–632.

Furukawa, M., Tsukahara, T., Tomita, K., Iwai, H., Sonomura, T., Miyawaki, S., and Sato, T. (2017). Neonatal maternal separation delays the GABA excitatory-to-inhibitory functional switch by inhibiting KCC2 expression. *Biochemical and biophysical research communications*, 493(3):1243–1249.

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., and Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–194.

Ganguly, K., Schinder, A. F., Wong, S. T., and Poo, M.-m. (2001). GABA itself promotes the developmental switch of neuronal GABAergic responses from excitation to inhibition. *Cell*, 105(4):521–532.

Ge, S., Goh, E. L., Sailor, K. A., Kitabatake, Y., Ming, G.-l., and Song, H. (2006). GABA regulates synaptic integration of newly generated neurons in the adult brain. *Nature*, 439(7076):589–593.

Ge, S., Yang, C.-h., Hsu, K.-s., Ming, G.-l., and Song, H. (2007). A critical period for enhanced synaptic plasticity in newly generated neurons of the adult brain. *Neuron*, 54(4):559–566.

Gilbert, P. E., Kesner, R. P., and Lee, I. (2001). Dissociating hippocampal subregions: A double dissociation between dentate gyrus and CA1. *Hippocampus*, 11(6):626–636.

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.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding II: Feedback, expectation, olfaction, illusions. *Biological Cybernetics*, 23:187–202.

Grossberg, S. (1987a). *The Adaptive Brain I*. Elsevier.

Grossberg, S. (1987b). Competitive learning: From interactive activation to adaptive resonance. *Cognitive science*, 11(1):23–63.

Gu, Y., Arruda-Carvalho, M., Wang, J., Janoschka, S. R., Josselyn, S. A., Frankland, P. W., and Ge, S. (2012). Optical controlling reveals time-dependent roles for adult-born dentate granule cells. *Nature neuroscience*, 15(12):1700–1706.

Gueneau, G., Privat, A., Drouet, J., and Court, L. (1982). Subgranular zone of the dentate gyrus of young rabbits as a secondary matrix. *Developmental neuroscience*, 5(4):345–358.

Hargreaves, E., Rao, G., Lee, I., and Knierim, J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science*, 308:1792–1794.

Heigele, S., Sultan, S., Toni, N., and Bischofberger, J. (2016). Bidirectional GABAergic control of action potential firing in newborn hippocampal granule cells. *Nature neuroscience*, 19(2):263–270.

Henze, D. A., Wittner, L., and Buzsáki, G. (2002). Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. *Nature neuroscience*, 5(8):790–795.

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

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

Hunsaker, M. R. and Kesner, R. P. (2008). Evaluating the differential roles of the dorsal dentate gyrus, dorsal CA3, and dorsal CA1 during a temporal ordering for spatial locations task. *Hippocampus*, 18(9):955–964.

Jarrard, L. E. (1993). On the role of the hippocampus in learning and memory in the rat. *Behavioral and neural biology*, 60(1):9–26.

Jessberger, S., Clark, R. E., Broadbent, N. J., Clemenson, G. D., Consiglio, A., Lie, D. C., Squire, L. R., and Gage, F. H. (2009). Dentate gyrus-specific knockdown of adult neurogenesis impairs spatial and object recognition memory in adult rats. *Learning & memory*, 16(2):147–154.

Johnston, S. T., Shtrahman, M., Parylak, S., Gonçalves, J. T., and Gage, F. H. (2016). Paradox of pattern separation and adult neurogenesis: A dual role for new neurons balancing memory resolution and robustness. *Neurobiology of learning and memory*, 129:60–68.

Káli, S. and Dayan, P. (2004). Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nature neuroscience*, 7(3):286–294.

Kee, N., Teixeira, C. M., Wang, A. H., and Frankland, P. W. (2007). Preferential incorporation of adult-generated granule cells into spatial memory networks in the dentate gyrus. *Nature neuroscience*, 10(3):355–362.

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., Kuhn, H. G., and Gage, F. H. (1997). More hippocampal neurons in adult mice living in an enriched environment. *Nature*, 386(6624):493–495.

Khazipov, R., Khalilov, I., Tyzio, R., Morozova, E., Ben-Ari, Y., and Holmes, G. L. (2004). Developmental changes in GABAergic actions and seizure susceptibility in the rat hippocampus. *European Journal of Neuroscience*, 19(3):590–600.

Kitamura, T., Saitoh, Y., Murayama, A., Sugiyama, H., and Inokuchi, K. (2010). LTP induction within a narrow critical period of immature stages enhances the survival of newly generated neurons in the adult rat dentate gyrus. *Molecular brain*, 3(13):1–8.

Kohonen, T. (1989). *Self-organization and associative memory*. Springer-Verlag, 3rd edition.

Kuhn, H. G., Dickinson-Anson, H., and Gage, F. H. (1996). Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *Journal of Neuroscience*, 16(6):2027–2033.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Leonzino, M., Busnelli, M., Antonucci, F., Verderio, C., Mazzanti, M., and Chini, B. (2016). The timing of the excitatory-to-inhibitory GABA switch is regulated by the oxytocin receptor via KCC2. *Cell reports*, 15(1):96–103.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., and Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*, 315(5814):961–966.

Levy, W. B. and Stewart, D. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in hippocampus. *Neuroscience*, 8:791–797.

Li, L., Sultan, S., Heigele, S., Schmidt-Salzmann, C., Toni, N., and Bischofberger, J. (2017). Silent synapses generate sparse and orthogonal action potential firing in adult-born hippocampal granule cells. *Elife*, 6:e23612.

Li, Y., Stam, F. J., Aimone, J. B., Goulding, M., Callaway, E. M., and Gage, F. H. (2013). Molecular layer perforant path-associated cells contribute to feed-forward inhibition in the adult dentate gyrus. *Proceedings of the National Academy of Sciences*, 110(22):9106–9111.

Lodge, M. and Bischofberger, J. (2019). Synaptic properties of newly generated granule cells support sparse coding in the adult hippocampus. *Behavioural brain research*, 372:112036.

Mardia, K. V. and Jupp, P. E. (2009). *Directional statistics*, volume 494. John Wiley & Sons.

Marín-Burgin, A., Mongiat, L. A., Pardi, M. B., and Schinder, A. F. (2012). Unique processing during a period of high excitation/inhibition balance in adult-born neurons. *Science*, 335(6073):1238–1242.

Marr, D. (1969). A theory of cerebellar cortex. *J. Physiology*, 202:437–470.

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London*, 262:23–81.

Martinetz, T. and Schulten, K. (1991). A "neural-gas" network learns topologies. *Artificial Neural Networks*, pages 397–402.

McHugh, T. J., Jones, M. W., Quinn, J. J., Balthasar, N., Coppari, R., Elmquist, J. K., Lowell, B. B., Fanselow, M. S., Wilson, M. A., and Tonegawa, S. (2007). Dentate gyrus NMDA receptors mediate rapid pattern separation in the hippocampal network. *Science*, 317(5834):94–99.

McNaughton, B. (1980). Evidence for two physiologically distinct perforant pathways to the fascia dentata. *Brain research*, 199(1):1–19.

McNaughton, B., Chen, L., and Markus, E. (1991). "Dead reckoning," landmark learning, and the sense of direction: A neurophysiological and computational hypothesis. *Journal of Cognitive Neuroscience*, 3:190–202.

Miller, K. D. and Fumarola, F. (2012). Mathematical equivalence of two common forms of firing rate models of neural networks. *Neural computation*, 24(1):25–31.

Ming, G.-l. and Song, H. (2005). Adult neurogenesis in the mammalian central nervous system. *Annual Review of Neuroscience*, 28:223–250.

Ming, G.-l. and Song, H. (2011). Adult neurogenesis in the mammalian brain: significant answers and significant questions. *Neuron*, 70(4):687–702.

Neunuebel, J. P. and Knierim, J. J. (2014). CA3 retrieves coherent representations from degraded input: direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron*, 81(2):416–427.

Neves, G., Cooke, S. F., and Bliss, T. V. P. (2008). Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nature Reviews Neuroscience*, 9(1):65–75.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273.

O'Keefe, J. and Nadel, L. (1978). *The Hippocampus as a cognitive map*. Clarendon Press.

Overstreet-Wadiche, L., Bromberg, D. A., Bensen, A. L., and Westbrook, G. L. (2005). GABAergic signaling to newborn neurons in dentate gyrus. *Journal of neurophysiology*, 94(6):4528–4532.

Owens, D. F. and Kriegstein, A. R. (2002). Is there more to GABA than synaptic inhibition? *Nature Reviews Neuroscience*, 3(9):715–727.

Pathak, H. R., Weissinger, F., Terunuma, M., Carlson, G. C., Hsu, F.-C., Moss, S. J., and Coulter, D. A. (2007). Disrupted dentate granule cell chloride regulation enhances synaptic excitability during development of temporal lobe epilepsy. *Journal of Neuroscience*, 27(51):14012–14022.

Pfister, J.-P. and Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *Journal of Neuroscience*, 26:9673–9682.

Piatti, V. C., Davies-Sala, M. G., Espósito, M. S., Mongiat, L. A., Trinchero, M. F., and Schinder, A. F. (2011). The timing for neuronal maturation in the adult hippocampus is modulated by local network activity. *Journal of Neuroscience*, 31(21):7715–7728.

Platt, J. (1991). A resource-allocating network for function interpolation. *Neural computation*, 3(2):213–225.

Preston, A. R. and Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current Biology*, 23(17):764–773.

Rapp, P. R. and Gallagher, M. (1996). Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. *Proceedings of the National Academy of Sciences*, 93(18):9926–9930.

Rolls, E. T. and Treves, A. (1998). *Neural networks and brain function*, volume 572. Oxford university press Oxford.

Rumelhart, D. E. and Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive science*, 9(1):75–112.

Sahay, A., Scobie, K. N., Hill, A. S., O'Carroll, C. M., Kheirbek, M. A., Burghardt, N. S., Fenton, A. A., Dranovsky, A., and Hen, R. (2011a). Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature*, 472(7344):466–470.

Sahay, A., Wilson, D. A., and Hen, R. (2011b). Pattern separation: a common function for new neurons in hippocampus and olfactory bulb. *Neuron*, 70(4):582–588.

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

Scharfman, H. E. (2007). The CA3 "backprojection" to the dentate gyrus. *Progress in brain research*, 163:627–637.

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.

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(11):11–21.

Senzai, Y. and Buzsáki, G. (2017). Physiological properties and behavioral correlates of hippocampal granule cells and mossy cells. *Neuron*, 93(3):691–704.

Shors, T. J., Miesegaes, G., Beylin, A., Zhao, M., Rydel, T., and Gould, E. (2001). Neurogenesis in the adult is involved in the formation of trace memories. *Nature*, 410(6826):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.

Sierra, A., Encinas, J. M., Deudero, J. J., Chancey, J. H., Enikolopov, G., Overstreet-Wadiche, L. S., Tsirka, S. E., and Maletic-Savatic, M. (2010). Microglia shape adult hippocampal neurogenesis through apoptosis-coupled phagocytosis. *Cell stem cell*, 7(4):483–495.

Snyder, J., Hong, N., McDonald, R., and Wojtowicz, J. (2005). A role for adult neurogenesis in spatial long-term memory. *Neuroscience*, 130(4):843–852.

Snyder, J. S. (2019). Recalibrating the relevance of adult neurogenesis. *Trends in neurosciences*, 42(3).

Sorrells, S. F., Paredes, M. F., Cebrian-Silla, A., Sandoval, K., Qi, D., Kelley, K. W., James, D., Mayer, S., Chang, J., Auguste, K. I., Chang, E. F., Gutierrez, A. J., Kriegstein, A. R., Mathern, G. W., Oldham, M. C., Huang, E. J., Garcia-Verdugo, J. M., Yang, Z., and Alvarez-Buylla, A.

(2018). Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature*, 555(7696):377–381.

Tartt, A. N., Fulmore, C. A., Liu, Y., Rosoklija, G. B., Dwork, A. J., Arango, V., Hen, R., Mann, J. J., and Boldrini, M. (2018). Considerations for assessing the extent of hippocampal neurogenesis in the adult and aging human brain. *Cell stem cell*, 23(6):782–783.

Tashiro, A., Makino, H., and Gage, F. H. (2007). Experience-specific functional modification of the dentate gyrus through adult neurogenesis: a critical period during an immature stage. *Journal of Neuroscience*, 27(12):3252–3259.

Tashiro, A., Sandler, V. M., Toni, N., Zhao, C., and Gage, F. H. (2006). NMDA-receptor-mediated, cell-specific integration of new neurons in adult dentate gyrus. *Nature*, 442(7105):929–933.

Temprana, S. G., Mongiat, L. A., Yang, S. M., Trinchero, M. F., Alvarez, D. D., Kropff, E., Giacomini, D., Beltramone, N., Lanuza, G. M., and Schinder, A. F. (2015). Delayed coupling to feedback inhibition during a critical period for the integration of adult-born granule cells. *Neuron*, 85(1):116–130.

Toni, N., Laplagne, D. A., Zhao, C., Lombardi, G., Ribak, C. E., Gage, F. H., and Schinder, A. F. (2008). Neurons born in the adult dentate gyrus form functional synapses with target cells. *Nature neuroscience*, 11(8):901–907.

Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E., and Svoboda, K. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420(6917):788–794.

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

Tronel, S., Belnoue, L., Grosjean, N., Revest, J.-M., Piazza, P.-V., Koehl, M., and Abrous, D. N. (2012). Adult-born neurons are necessary for extended contextual discrimination. *Hippocampus*, 22(2):292–298.

Tronel, S., Fabre, A., Charrier, V., Oliet, S. H., Gage, F. H., and Abrous, D. N. (2010). Spatial learning sculpts the dendritic arbor of adult-born hippocampal neurons. *Proceedings of the National Academy of Sciences*, 107(17):7963–7968.

Turrigiano, G. G. and Nelson, S. B. (2000). Hebb and homeostasis in neuronal plasticity. *Current Opinion in Neurobiology*, 10:358–364.

Tyzio, R., Holmes, G. L., Ben-Ari, Y., and Khazipov, R. (2007). Timing of the developmental switch in $GABA_A$ mediated signaling from excitation to inhibition in CA3 rat hippocampus using gramicidin perforated patch and extracellular recordings. *Epilepsia*, 48:96–105.

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

Varela, J. A., Sen, K., Gibson, J., Fost, J., Abbott, L. F., and Nelson, S. B. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *Journal of Neuroscience*, 17:7926.

Vazdarjanova, A. and Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *Journal of Neuroscience*, 24(29):6489–6496.

Vitureira, N., Letellier, M., and Goda, Y. (2012). Homeostatic synaptic plasticity: from single synapses to neural circuits. *Current opinion in neurobiology*, 22(3):516–521.

Vivar, C., Potter, M. C., Choi, J., Lee, J.-y., Stringer, T. P., Callaway, E. M., Gage, F. H., Suh, H., and Van Praag, H. (2012). Monosynaptic inputs to new neurons in the dentate gyrus. *Nature Communications*, 3:1107.

Wang, D. D. and Kriegstein, A. R. (2010). Blocking early GABA depolarization with bumetanide results in permanent alterations in cortical circuits and sensorimotor gating deficits. *Cerebral cortex*, 21(3):574–587.

Wang, 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.

Weisz, V. I. and Argibay, P. F. (2009). A putative role for neurogenesis in neurocomputational terms: Inferences from a hippocampal model. *Cognition*, 112(2):229–240.

Weisz, V. I. and Argibay, P. F. (2012). Neurogenesis interferes with the retrieval of remote memories: Forgetting in neurocomputational terms. *Cognition*, 125(1):13–25.

West, M., Slomianka, L., and Gundersen, H. J. G. (1991). Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *The Anatomical Record*, 231(4):482–497.

Wiskott, L., Rasch, M. J., and Kempermann, G. (2004). What is the functional role of adult neurogenesis in the hippocampus? *Cognit Sci EPrint Archive (CogPrints)4012*.

Wiskott, L., Rasch, M. J., and Kempermann, G. (2006). A functional hypothesis for adult hippocampal neurogenesis: avoidance of catastrophic interference in the dentate gyrus. *Hippocampus*, 16(3):329–343.

Yingwei, L., Sundararajan, N., and Saratchandran, P. (1997). A sequential learning scheme for function approximation using minimal radial basis function neural networks. *Neural computation*, 9(2):461–478.

Zenke, F., Agnes, E., and Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, 6:6922.

Zenke, F. and Gerstner, W. (2017). Hebbian plasticity requires compensatory processes on multiple timescales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1715):1–17.

Zhao, C., Teng, E. M., Summers, R. G., Ming, G.-l., and Gage, F. H. (2006). Distinct morphological stages of dentate granule neuron maturation in the adult mouse hippocampus. *Journal of Neuroscience*, 26(1):3–11.

Olivia Gozel

EPFL IC IINFCOM LCN1
AAB 1 33 (Bâtiment AAB)
Station 15
CH – 1015 Lausanne
Switzerland

E-mail: olivia.gozel@epfl.ch
Phone: +41 21 693 18 07

Education

2014 – 2019	**Ecole Polytechnique Fédérale de Lausanne (EPFL)** **Research and Teaching assistant** PhD position supervised by Prof. Wulfram Gerstner, Laboratory of Computational Neuroscience
2016	**Summer school** **Methods in Computational Neuroscience** Marine Biological Laboratory, Woods Hole, MA
2013 – 2014	**EPFL** **Research and Teaching assistant** Lab rotation co-supervised by Prof. Olaf Blanke, Laboratory of Cognitive Neuroscience, and Prof. José Millán, Laboratory of Non-Invasive Brain-Machine Interface
2012 - 2013	**Harvard Medical School** Master thesis: “Local Field Correlates of Visual Attention”, supervised by Prof. John Maunsell, Neurobiology department
2011 - 2013	**EPFL** Master in Life Sciences and Technology, Neuroscience track
2008 - 2011	**EPFL** Bachelor in Life Sciences and Technology

Fellowships and awards

2016	Scholarship for Methods in Computational Neuroscience course, Marine Biological Laboratory, Woods Hole, MA
2013	Award of Excellence for the quality of my Master studies
2012 - 2013	Fellowship from the Bertarelli program in Translational Neuroscience and Neuroengineering for my master thesis at Harvard Medical School

Publications

Gozel O, Gerstner W. Adult dentate gyrus neurogenesis: A functional model. BioRxiv 704791 (preprint), 2019.

Schurger A*, Gale S*, **Gozel O**, Blanke O. Performance monitoring for brain-computer-interface actions. *Brain and Cognition*, 111, p. 44-50, 2017.

Selected presentations

Gozel O, and Gerstner W. From adult dentate gyrus neurogenesis to unsupervised clustering. Computational and Systems Neuroscience (CoSyNe) conference, Lisbon, Portugal, 2019 (poster)

Gozel O. Tutorial on adult dentate gyrus neurogenesis for machine learners. European Institute for Theoretical Neuroscience (EITN) Workshop « From Neuroscience to Machine Learning », Paris, France, 2018 (talk)

Gozel O, and Gerstner W. A functional role for the switch from excitation to inhibition in neurogenesis. International Conference on Mathematical Neuroscience, Boulder, CO, 2017 (poster)

Gozel O. A model of synaptic consolidation. HBP CDP5 Plasticity Workshop, Fürberg, Austria, 2016 (talk)

Mayo JP, **Gozel O**, and Maunsell JHR. Predictive power of area V4 local field potentials in determining the state of attention. Society for Neuroscience conference, 261.08, San Diego, CA, 2013 (poster)

Teaching activities

2018-2019	Supervisor for internship students
2015	Teaching assistant for *General Physics II*, EPFL
2014	Teaching assistant for *General Physics I*, EPFL
2014	Teaching assistant for *Neuroscience for Engineers*, EPFL
2009-2010	Student assistant for *Linear Algebra*, EPFL

Technical skills

Programming	Matlab, Python, BRIAN2, C/C++, Mathematica

Languages

English (fluent), French (mother tongue), German (intermediate), Italian (basic)