Adult dentate gyrus neurogenesis: A functional model

Olivia Gozel and Wulfram Gerstner
École Polytechnique Fédérale de Lausanne (EPFL)

1 Abstract

In adult dentate gyrus neurogenesis, the link between maturation of newborn neurons and their function, such as behavioral pattern separation, has remained puzzling. By analyzing a theoretical model, we show that the switch from excitation to inhibition of the GABAergic input onto maturing newborn cells is crucial for their proper functional integration. When the GABAergic input is excitatory, cooperativity drives the growth of synapses such that newborn cells become sensitive to stimuli similar to those that activate mature cells. When GABAergic input switches to inhibitory, competition pushes the configuration of synapses onto newborn cells towards representations that are away from previously stored memories. This enables the maturing newborn cells to code for concepts that are novel, yet similar to familiar ones. Our theory of newborn cell maturation can explain how adult newborn cells integrate into the preexisting network and why they promote pattern separation of similar stimuli.

2 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 [1]. The dentate gyrus is the entry point of input from cortex, primarily entorhinal cortex (EC), to the hippocampus [2], which is believed to be a substrate of learning and memory [3]. Adult-born cells in dentate gyrus mostly develop into dentate granule cells (DGCs), the main excitatory cell type [1].

The properties of rodent adult-born DGCs change as a function of their maturation stage, until they become undistinguishable from other mature DGCs at approximately 8 weeks [1, 4] (Fig. 1B). Many of them die before they fully mature [5]. Their survival is experience-dependent, and depends on NMDA receptor activation [6]. Initially, newborn DGCs have enhanced excitability [7, 8] and stronger synaptic plasticity than mature DGCs, reflected by a larger LTP amplitude and a lower threshold for induction of LTP [7, 9, 10]. 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 [11].

Newborn DGCs receive no direct connections from mature DGCs [12, 13], (yet see [14]), but are indirectly activated via interneurons [13, 15]. 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 [1] (Fig. 1B). Analogous to a similar transition during embryonic and

Table 1: Parameters for the simulations

	Biologically-plausible network		**Simplified network**	
Network	$N_{EC} = 144$ $N_I = 25$	$N_{DGC} = 100$	$N_{EC} = 128$	$N_{DGC} = 3$
Connectivity	$w_{IE} = 1$ $p_{IE} = 0.9$	$w_{EI} = -\frac{1}{p_{EI} * N_I}$ $p_{EI} = 0.9$	$w_{rec} = -1.2$	
Dynamics	$\tau_m = 20$ ms $L = 0.5$	$\tau_{\text{inh}} = 2$ ms $p^* = 0.1$	$\tau_m = 20$ ms	
Plasticity	$\alpha_0 = 0.05$ $\gamma_0 = 10$ $\nu_0 = 0.2$	$\beta = 1$ $\theta = 0.15$	$\alpha_0 = 0.03$ $\gamma_0 = 1.65$	$\beta = 1$ $\theta = 0.15$
Numerical simulations	$\Delta t = 0.1$ ms $\eta_b = 0.01$	$\eta = 0.01$	$\Delta t = 1$ ms	$\eta = 0.01$

early postnatal stages [16], 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 [17–19]. Importantly, it has been shown that GABAergic inputs are crucial for the integration of newborn DGCs into the preexisting circuit [13, 15, 19, 20].

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) [21]. Even though the amount of newly generated cells per month is rather low, 3 to 6% of the total DGCs population [22, 23], adult-born DGCs are critical for behavioral pattern separation [24–26], in particular in tasks where similar stimuli or contexts have to be discriminated [24, 25]. However, the functional role of adult-born DGCs is controversial [27, 28]. One view is that newborn DGCs contribute to pattern separation through a modulatory role [27], via a combination of enhanced excitability [7, 8], and absence of inhibition [13, 15]. 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 [21, 29]. 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 [28].

To address this controversy, we present a model of how newborn DGCs integrate into the preexisting 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 why newborn DGCs favor pattern separation of similar stimuli, but do not impact pattern separation of distinct stimuli.

3 Methods

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 [2, 11, 66, 67]. 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 (Fig. 1A). The activity of N_{EC} neurons in EC represents an input pattern $\vec{x} = (x_1, x_2, ..., x_{N_{EC}})$. Because the perforant path also induces strong feedforward inhibition in the dentate gyrus [68], 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, spiking mature DGCs activate interneurons in DG, which in turn inhibit other mature DGCs [11, 13, 69]. In our model, the DGCs are thus recurrently connected with inhibitory neurons (Fig. 1A). 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 1(Biologically-plausible network).

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 [70]:

$$\tau_m \frac{\mathrm{d}\nu_i}{\mathrm{d}t} = -\nu_i + \tanh\left(\frac{[I_i - b_i]_+}{L}\right) \tag{1}$$

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

$$I_i = \sum_{j=1}^{N_{EC}} w_{ij} x_j + \sum_{k=1}^{N_I} w_{ik}^{EI} \nu_k^I \tag{2}$$

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 measured in units of Hz.

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

$$\tau_{\mathrm{inh}} \frac{\mathrm{d}\nu_k^I}{\mathrm{d}t} = -\nu_k^I + [I_k^I - p^* N_{DGC}]_+ \tag{3}$$

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

$$I_k^I = \sum_{i=1}^{N_{DGC}} w_{ki}^{IE} \nu_i \tag{4}$$

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 [71]) for each pattern presentation, and less than 10% are highly active (firing rate > 9 Hz) (Fig. 1F), consistent with the experimentally observed activity sparsity in dentate gyrus [47].

Plasticity rule

Projections from EC onto newborn DGCs exhibit Hebbian plasticity [7, 10, 72], and NMDA receptors in DGCs are necessary for discrimination of similar contexts in a fear conditioning task in mice [72]. 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 ν_i of the postsynaptic cell [32–34]. 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:

$$\Delta w_{ij} = -\alpha x_j \nu_i [\theta - \nu_i]_+ + \gamma x_j \nu_i [\nu_i - \theta]_+ - \beta w_{ij} [\nu_i - \theta]_+ \nu_i^3 \tag{5}$$

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 1(Biologically-plausible network). The weights are hard-bounded from below at 0, i.e. if equation (5) leads to a new weight smaller than zero, w_{ij} is set to zero. The first two terms of expression (5) are a variation of the BCM rule [32]. The third term implements heterosynaptic plasticity [73, 74]. 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 [48, 51], as we will discuss now.

For a given input pattern $\vec{x}^\mu$, there are three fixed points for the postsynaptic firing rate: $\nu_i = 0$, $\nu_i = \theta$, and $\nu_i = \sqrt{\frac{\gamma x_j^\mu}{\beta w_{ij}}} =: \hat{\nu}_i$ (the negative root is omitted, because $\nu_i \geqslant 0$ due to equation (1)). 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: $w_{ij} \to \frac{\gamma x_j^\mu}{\beta \nu_i^2}$ (Fig. 1C). If a pattern $\vec{x}^\mu$ is presented only for a short time these fixed points are not reached during a single pattern presentation.

After presentation of many input patterns $\vec{x}$ that all drive the postsynaptic neuron to fire at a rate $\nu_i > \theta$, the expectation of the L2-norm of the feedforward weight vector onto neuron i converges to:

$$\langle \left(\sqrt{\sum_j w_{ij}^2} \right) \rangle = \langle \left(\frac{\gamma}{\beta \nu_i^2} \sqrt{\sum_j x_j^2} \right) \rangle = \frac{\gamma}{\beta} \langle \left(\frac{1}{\nu_i^2} \right) \rangle \tag{6}$$

where we have exploited that $||\vec{x}|| = \sqrt{\sum_j x_j^2} = 1$. Therefore, the L2-norm of the feedforward weight vector onto DGCs is bounded and reaches similar values for all those DGCs that strongly respond to at least one input pattern. The actual L2-norm will be a bit smaller if neuron i responds weakly ($0 \leqslant \nu_i \leqslant \theta$) to other input patterns of the data set. Equation (6) is exact if patterns for which neuron i is not a "winner" are quasi-orthogonal (see definitions in next paragraph).

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 (cf. equation (6)), 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.

Early maturation phase

During the early phase of maturation, the GABAergic input onto a newborn DGC with index n has an excitatory effect. In the model, it is implemented as follows: $w_{nk}^{EI} = -w_{EI} > 0$ with probability p_{IE} for any interneuron k and $w_{nk}^{EI} = 0$ otherwise (no connection). Since newborn cells do not project yet onto inhibitory neurons [11], we have $w_{kn}^{IE} = 0 \ \forall n$. Newborn DGCs are known to have enhanced excitability [7, 8], so their threshold is kept at 0 ($b_n = 0 \ \forall n$). 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 (5).

Late maturation phase

During the late phase of maturation (roughly at about 3 weeks [19]), the GABAergic input onto newborn DGCs switches from excitatory to inhibitory. In terms of our model, it means that all existing w_{nk}^{EI} connections switch their sign to $w_{EI} < 0$. Furthermore, since newborn DGCs develop lateral connections to inhibitory neurons in the late maturation phase [11], we set $w_{kn}^{IE} = w_{IE}$ with probability p_{IE}, and $w_{kn}^{IE} = 0$ otherwise. The thresholds of newborn DGCs are updated ($b_i^{(\mu)} = b_i^{(\mu-1)} + \eta_b \Delta b_i$) according to $\Delta b_i = \nu_i - \nu_0$, where ν_0 is a reference rate, to mimic the decrease of excitability as newborn DGCs mature. It can be observed that the distribution of firing rates of newborn DGCs is indeed shifted to the right (towards higher firing rates) in the early phase of maturation compared to the late phase of maturation (Fig. 1F). A sufficient condition for newborn DGCs to win the competition upon presentation of patterns of the novel cluster is that the scalar product between the patterns of the novel cluster and the newborn DGCs feedforward weight vectors is larger than the scalar product between the patterns of the novel cluster and the mature DGCs feedforward weight vectors. Analogous to the early phase, presentation of all patterns of the data set once (1 epoch) is sufficient to reach convergence of the feedforward weights onto newborn DGCs.

Input patterns

Two different sets of input patterns are used. Both data sets have a number K of clusters and several thousand patterns per cluster. As a first data set, we use the MNIST 12x12 patterns [31] ($N_{EC} = 144$), normalized such that the L2-norm of each pattern is equal to 1. The training set contains approximately 6000 patterns per digit, while the testing set contains about 1000 patterns per digit (Fig. 1D1).

As a second data set, we use hand-made artificial patterns designed such that the distance between the centers of any two clusters is the same. All clusters lie on the positive quadrant of the surface of a hypersphere of dimension $N_{EC} - 1$. The cluster centers are Walsh patterns shifted along the diagonal (Fig. 1D2):

$$\begin{aligned}
\vec{P}^1 &= \frac{1}{c_0}\,(1+\xi, 1-\xi, 1+\xi, 1-\xi, ..., 1+\xi, 1-\xi, 1+\xi, 1-\xi) \\
\vec{P}^2 &= \frac{1}{c_0}\,(1+\xi, 1+\xi, 1-\xi, 1-\xi, ..., 1+\xi, 1+\xi, 1-\xi, 1-\xi) \\
&... \\
\vec{P}^K &= \frac{1}{c_0}\,(1+\xi, 1+\xi, 1+\xi, 1+\xi, ..., 1-\xi, 1-\xi, 1-\xi, 1-\xi)
\end{aligned} \tag{7}$$

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:

$$c_0 = \sqrt{N_{EC}\,(1+\xi^2)}. \tag{8}$$

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 (Fig. 1E):

$$\vec{P}^k \cdot \vec{P}^l = \frac{1}{1+\xi^2} = \cos(\Omega) \tag{9}$$

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 [75]:

$$\vec{x}^{\mu(k)} \sim \left(\sqrt{1-a^2}\right)\vec{\zeta} + a\vec{P}^k \tag{10}$$

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 [75]. 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 $z \sim \mathcal{B}e(\frac{N_{EC}-1}{2}, \frac{N_{EC}-1}{2})$.

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 $\kappa = 10^4$.

Classification performance

To evaluate whether the newborn DGCs contribute to the dentate gyrus network function, we study classification performance. First, the feedforward weights are learned upon presentation of many input patterns from the training set, as described in Section Plasticity rule. 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:

$$\nu_k^{RO,\mu} = g\left(I_k^{RO,\mu}\right) = g\left(\sum_{i=1}^{N_{DGCs}} w_{ki}^{RO}\nu_i^{\mu}\right) \tag{11}$$

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 [48]:

$$L(W^{RO}) = \frac{1}{2}\sum_{k=1}^{N_{RO}}(L_k^{\mu} - \nu_k^{RO,\mu})^2 \tag{12}$$

$$= \frac{1}{2}\sum_{k=1}^{N_{RO}}\left(L_k^{\mu} - g\left(\sum_{i=1}^{N_{DGCs}} w_{ki}^{RO}\nu_i^{\mu}\right)\right)^2 \tag{13}$$

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:

$$\Delta w_{ki}^{RO,\mu} = \eta(L_k^{\mu} - \nu_k^{RO,\mu})g'\left(I_k^{RO,\mu}\right)\nu_i^{\mu} \tag{14}$$

The readout units have a rectified tangent hyperbolic gain function: $g(x) = \tanh\left(2[x]_+\right)$, whose derivative is: $g'(x) = 2\left(1 - \left(\tanh\left(2[x]_+\right)\right)^2\right)$. 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.

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 (Fig. 2F), or fix the feedforward connections for all selective DGCs but keep unselective neurons plastic (as in the neurogenesis case) (Fig. 2E). 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 (Suppl. Fig. S1A).

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 (Methods, Section Network architecture and rate neurons dynamics) and their feedforward weights are updated according to our plasticity rule (Methods, Section Plasticity rule). 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. 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 (Fig. 2A).

Simplified rate network

We use a toy network and the 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:

$$\tau_m \frac{\mathrm{d}\nu_i}{\mathrm{d}t} = -\nu_i + \mathcal{H}\left(I_i - b_i\right) \tag{15}$$

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

$$I_i = \sum_{j=1}^{N_{EC}} w_{ij} x_j + \sum_{k \neq j}^{N_{DGC}} w_{rec} \nu_k \tag{16}$$

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 implicitely: 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 (15). 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 (5), but with the parameters reported in Table 1(Simple network). 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 (15)).

Asssuming a single winner i^* for each pattern presentation, the input (equation (16)) to the winner is:

$$I_{i^*} = \vec{w}_{i^*} \cdot \vec{x}, \tag{17}$$

while the input to the losers is:

$$I_i = \vec{w}_i \cdot \vec{x} + w_{\text{rec}}. \tag{18}$$

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

$$\vec{w}_{i^*} \cdot \vec{x} > b_i \tag{19}$$

for the winner to actually be active, and:

$$\vec{w}_i \cdot \vec{x} + w_{\text{rec}} < b_i \tag{20}$$

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 [7, 8].

Similar versus distinct patterns with the artificial data set

Using the artificial data set with $|\xi| < 1$ (equation (7)), the scalar product between the center of mass of two different clusters, given by equation (9), satisfies: $0.5 \leqslant \frac{1}{1+\xi^2} \leqslant 1$. This corresponds to $0° \leqslant \Omega \leqslant \Omega_{\text{max}} = 60°$.

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:

$$I_i = \vec{w}_i \cdot \vec{x}^{\mu(k)} = ||\vec{w}_i|| \cdot ||\vec{x}^{\mu(k)}|| \cdot \cos(\vartheta_{\text{kk}}) = ||\vec{w}_i|| \cdot \cos(\vartheta_{\text{kk}}) \tag{21}$$

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:

$$I_i = \vec{w}_i \cdot \vec{x}^{\mu(l)} = ||\vec{w}_i|| \cdot ||\vec{x}^{\mu(l)}|| \cdot \cos(\vartheta_{\text{kl}}) \approx \frac{||\vec{w}_i||}{1+\xi^2} \tag{22}$$

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:

$$b_i < \frac{||\vec{w}_i||}{1+\xi^2} \tag{23}$$

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 expected L2-norm of the feedforward weight vector towards the DGCs at convergence can be computed, see equation (6). With the parameters in Table 1(Simple network), the value is $\langle ||\vec{w}_i|| \rangle = 1.5$. 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 (19). The threshold value ξ_{thresh} for which two clusters are similar (and above which two clusters are distinct) can be determined by equation (23) : $\xi_{\text{thresh}} = 0.5$. 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 (20), 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 (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 [19] and the newborn DGC does not inhibit the mature DGCs [11]. 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 [7, 8], 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 [19], and the newborn DGC recruits feedback inhibition onto mature DGCs [11]. 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 ($\xi = 0.8$).

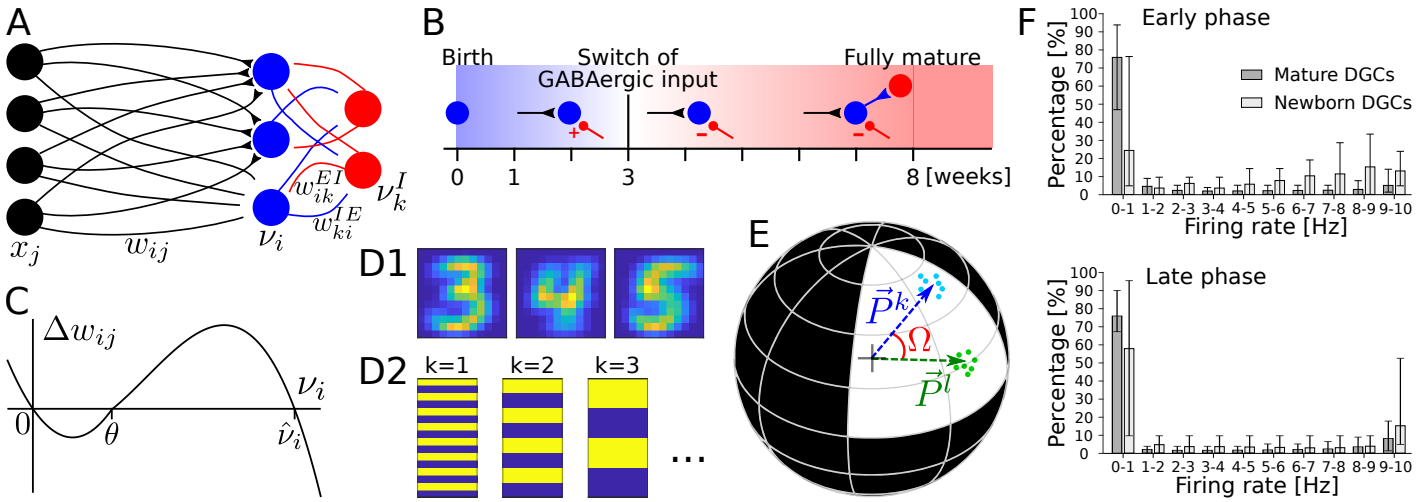

Figure 1: Model network and input patterns. (A) Network structure. EC neurons (black, rate x_j) are fully connected with weights w_{ij} to DGCs (blue, rate ν_i). DGCs and interneurons (red, rate ν_k^I) are mutually connected with probability p_{IE} and p_{EI} and weights w_{ki}^{IE} and w_{ik}^{EI}, respectively. Connections with a triangular end (black and blue) are glutamatergic, the others (red) are GABAergic. (B) Integration of an adult-born DGC (blue) into the preexisiting network as a function of time. Around 1 week after birth, the cell receives GABAergic synaptic input (red) [12, 19, 30], and a few days later also glutamatergic synaptic input from EC (black) [14, 19]. At about 3 weeks, the GABAergic input switches from excitatory to inhibitory [19]. At 7 weeks, the newborn DGC connects to local interneurons (red) [11]. (C) 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 $\theta < \nu_i < \hat{\nu}_i$. (D1) Center of mass for three ensembles of patterns from the MNIST data set, visualized as 12x12 pixel patterns. (D2) Center of mass of the first three clusters of an artificial data set visualized as 16x8 pixel patterns. In both cases, the 2-dimensional arrangement is chosen for visualization only. (E) 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. (F) Distribution of the percentage of DGCs (mean with 10th and 90th percentiles) in each bin of firing rate at the end of the early and late phase of maturation upon presentation of MNIST patterns. Percentages are obtained by normalizing by the number of neurons in each subpopulation (mature in dark grey, newborn in light grey).

4 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 [31]. 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 (Fig. 1D1). Pretraining was based on a synaptic plasticity rule which

combines LTP and LTD [32–34] (Methods, Fig. 1C).

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 (Fig. 2A). 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 selective DGCs represent different prototypes of the two digits, visible here as different writing styles and inclinations (Fig. 2A). At the end of pretraining, the classification error (Methods, Section Classification performance) 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.

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 (Methods). 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 (Fig. 2B).

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 (Fig. 1F), 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 (Fig. 2C).

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 (Fig. 3B1). The firing rates of all DGCs were also projected on the same PCs at the end of the early phase of maturation (Fig. 3A1). 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 (Fig. 3B1). Second, the representations of a few extremal patterns seem to span a low-dimensional manifold (Fig. 3C).

Third, we observe that the representation of novel patterns lays close to the representation of pretrained patterns at the end of the early phase of maturation of newborn DGCs (Fig. 3A1), while it moves away along the axis of PC 2 to a previously empty subspace in the late phase of maturation (Fig. 3B1), 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 (Fig. 3A2), but close to each other (and far from familiar patterns) at the end of the late phase of maturation of newborn DGCs (Fig. 3B2). The zoom in insets further show the influence of lateral inhibition on firing rate dynamics of the DGCs.

To test the quality of the learned representations, we compute classification performance by a linear classifier for the three ensembles of digits (Methods, Section Classification performance). 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 (Suppl. Fig. S1A). 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 Fig. 2A and Suppl. Fig. S1A).

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 (Fig. 2D). Therefore, newborn DGCs can ultimately promote separation of several novel ensembles of patterns, no matter if they are learned simultaneoulsy (Fig. 2D) or sequentially (Suppl. Fig. S1B).

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) (Fig. 2E). 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 (Fig. 2F), while none of the previously unresponsive units becomes selective for the novel digit (compare with Fig. 2A). 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.

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 [24, 25], we used a simplified competitive winner-take-all network (Methods, Section Simplified rate network) and constructed an artificial data set (Methods, Section Input patterns) (Fig. 1D2). 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 (Fig. 4B1,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 (Fig. 4D1), but not if it is distinct ($\xi = 0.8$, Fig. 4D2). 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 Fig. 4C1 and Fig. 4E top). 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 Fig. 4C2 and Fig. 4E bottom).

As a result of the different orientation of the feedforward weight vector onto the newborn DGC 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 (Fig. 4C1). 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 (Fig. 4D1). 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 (Fig. 4C2). 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 (Fig. 4D2). 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 [35, 36]. For discrimination of distinct stimuli, another pathway might be used, such as the direct EC to CA3 connection [37, 38].

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 (Methods, Section Similar versus distinct patterns with the artificial data set).

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 (Suppl. Fig. S2).

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 ($\xi = 0.2$, Fig. 4C1 and Suppl. Fig. S2A1) than in the case of distinct patterns ($\xi = 0.8$, Fig. 4C2 and Suppl. Fig.S2B1). 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 (Supplementary Material).

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 (Suppl. Fig. S2A2), but not in the case of distinct patterns (Suppl. Fig. S2B2), indicating that 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 (Suppl. Fig. S2A2), 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 Discussion

While experimental results steming 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 [10, 13, 19] and that adult dentate gyrus neurogenesis promotes better pattern separation [24–26], the link between channel properties and behavior has remained puzzling [27, 28]. In this paper, 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 [24–26]. Previous modeling studies already suggested that newborn DGCs integrate novel inputs into the representation in dentate gyrus [11, 39–46]. 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) (Fig. 5A). 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 the newborn DGCs towards novel features (Fig. 5B). At the end of maturation, newborn DGCs are thus selective for novel inputs, in agreement with the experimental observation that DGCs generate distinct codes for different environments [47]. Loosely speaking, the cooperative phase of excitatory GABAergic input promotes the growth of the synaptic weights coarsly 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" [48]. Unsupervised competitive learning has been used to perform clustering of input patterns into a few categories [48–52]. 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 [48, 51]. 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" (Methods, Section Winners, losers, and quasi-orthogonal inputs) to the subspace of the presented input patterns (Fig. 5C). Therefore, those neurons called "unresponsive units" will never get active during pattern presentation. Different learning strategies have been developped in the field of artificial neural networks to avoid this problem [32, 48–54]. 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 [55–58], 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 [35, 36]. To discriminate distinct stimuli, another pathway might be involved, such as the direct EC to CA3 connection [37,

38]. 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 [24, 25]. 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 [6].

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 [19]. Conversely, a reduction in inhibition in the dentate gyrus through decrease in KCC2 expression has been associated with epileptic activity [59, 60]. 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 [16, 61]. 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 [62–64] and adult dentate gyrus neurogenesis [15]. Furthermore, in hippocampal cell cultures, expression of KCC2 is upregulated by GABAergic activity but not affected by glutamatergic activity [65]. 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 [22, 23], 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 EC to DGC synapses 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 [27, 28].

6 Acknowledgments

This research was supported by Swiss National Science Foundation (no. 200020_184615) and by the European Union Horizon 2020 Framework Program under grant agreement no. 785907 (HumanBrain Project, SGA2).

References

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

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

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

[4] S. T. Johnston, M. Shtrahman, S. Parylak, J. T. Gonçalves, and F. H. Gage. 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, 2016.

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

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

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

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

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

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

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

[12] A. Deshpande, M. Bergami, A. Ghanem, K.-K. Conzelmann, A. Lepier, M. Götz, and B. Berninger. 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, 2013.

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

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

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

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

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

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

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

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

[21] A. Tashiro, H. Makino, and F. H. Gage. 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, 2007.

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

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

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

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

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

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

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

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

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

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

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

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

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

[35] P. E. Gilbert, R. P. Kesner, and I. Lee. Dissociating hippocampal subregions: A double dissociation between dentate gyrus and ca1. *Hippocampus*, 11(6):626–636, 2001.

[36] M. R. Hunsaker and R. P. Kesner. 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, 2008.

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

[38] A. Vazdarjanova and J. F. Guzowski. 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, 2004.

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

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

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

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

[43] R. A. Chambers and S. K. Conroy. 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, 2007.

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

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

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

[47] M. Chawla, J. Guzowski, V. Ramirez-Amaya, P. Lipa, K. Hoffman, L. Marriott, P. Worley, B. McNaughton, and C. A. Barnes. 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, 2005.

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

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

[50] S. Grossberg. *The Adaptive Brain I.* Elsevier, 1987.

[51] T. Kohonen. *Self-organization and associative memory, 3rd edition.* Springer-Verlag, 1989.

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

[53] S. Grossberg. Adaptive pattern classification and universal recoding ii: Feedback, expectation, olfaction, illusions. *Biol. Cybern.*, 23:187–202, 1976.

[54] D. DeSieno. 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, 1988.

[55] D. Marr. A theory of cerebellar cortex. *J. Physiology*, 202:437–470, 1969.

[56] J. Albus. A theory of cerebellar function. *J. Mathematical Biosciences*, 10:25–61, 1971.

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

[58] E. T. Rolls and A. Treves. *Neural networks and brain function*, volume 572. Oxford university press Oxford, 1998.

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

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

[61] M. Furukawa, T. Tsukahara, K. Tomita, H. Iwai, T. Sonomura, S. Miyawaki, and T. Sato. Neonatal maternal separation delays the GABA excitatory-to-inhibitory functional switch by inhibiting kcc2 expression. *Biochemical and biophysical research communications*, 493(3):1243–1249, 2017.

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

[63] R. Tyzio, G. L. Holmes, Y. Ben-Ari, and R. Khazipov. 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, 2007.

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

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

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

[67] D. A. Henze, L. Wittner, and G. Buzsáki. Single granule cells reliably discharge targets in the hippocampal ca3 network in vivo. *Nature neuroscience*, 5(8):790–795, 2002.

[68] Y. Li, F. J. Stam, J. B. Aimone, M. Goulding, E. M. Callaway, and F. H. Gage. 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, 2013.

[69] J. K. Leutgeb, S. Leutgeb, M.-B. Moser, and E. I. Moser. Pattern separation in the dentate gyrus and ca3 of the hippocampus. *Science*, 315(5814):961–966, 2007.

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

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

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

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

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

[75] K. V. Mardia and P. E. Jupp. *Directional statistics*, volume 494. John Wiley & Sons, 2009.

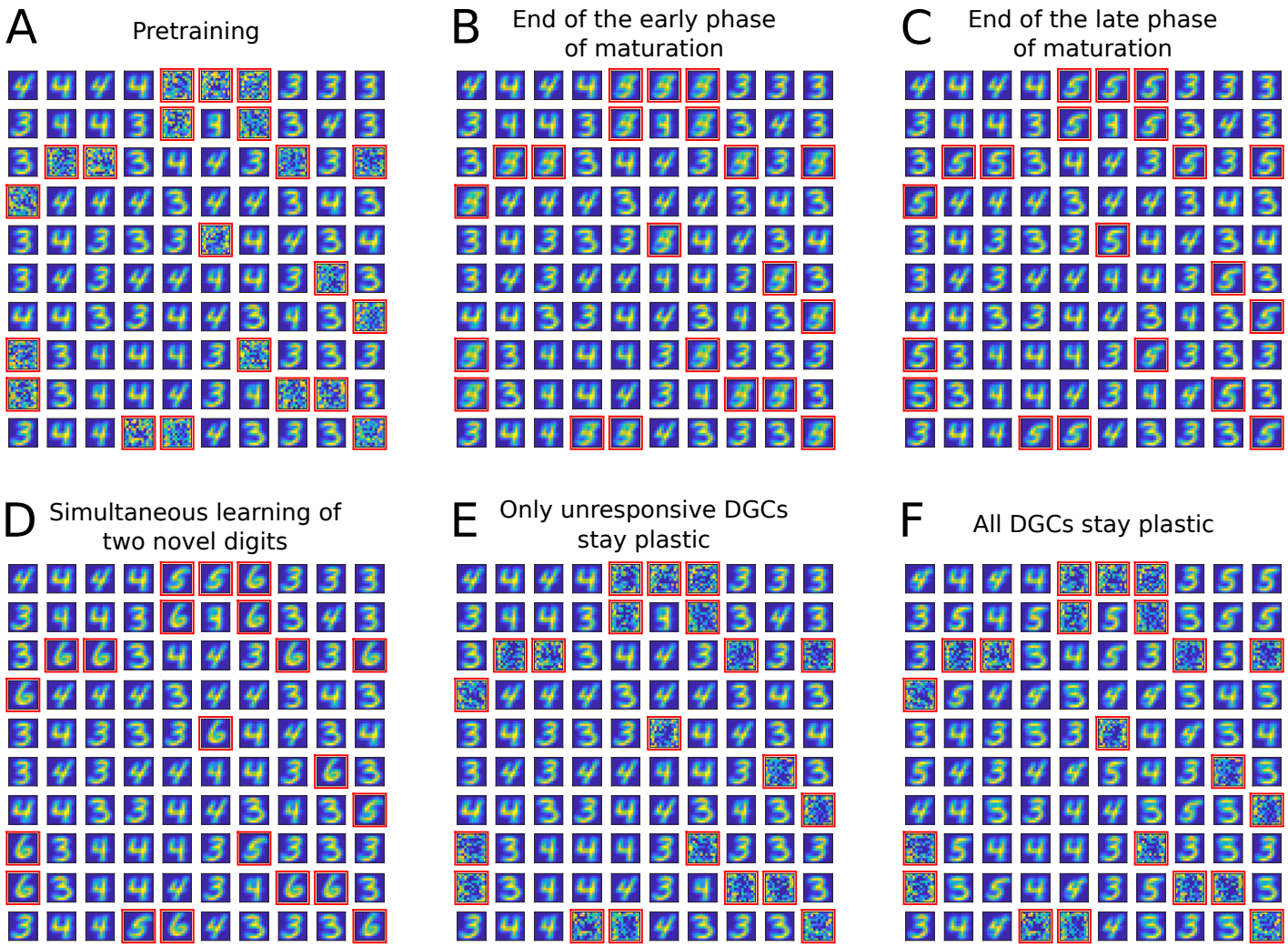

Figure 2: Newborn DGCs learn novel patterns. Receptive fields, defined as the set of feedforward weights, are visually represented in a 2-dimensional organization. (A) During pretraining, patterns from MNIST digits 3 and 4 are presented to the network. At the end of the pretraining, some DGCs have receptive 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. (D) Several novel digits can be learned simultaneously. After pretraining with patterns from digits 3 and 4 as in (A), 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. (E) Control without maturation. After pretraining with patterns from digits 3 and 4 as in (A), 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. (F) 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.

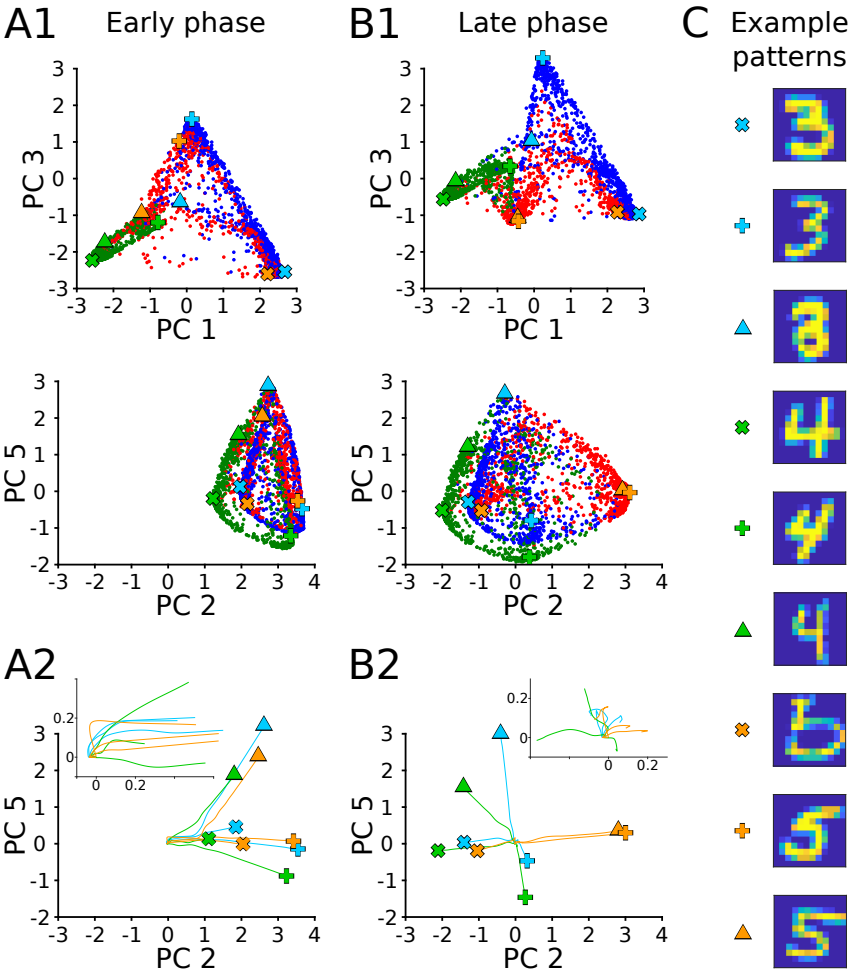

Figure 3: The representation of novel patterns occupies a previously empty subspace. (A1, A2) Projections of the final firing rates of all 100 DGCs on PC 1 and PC 3 (first row), and on PC 2 and PC 5 (second row), at the end of the early (A1) or late (A2) phase of maturation of the newborn DGCs. Each 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. (B1, B2) 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).

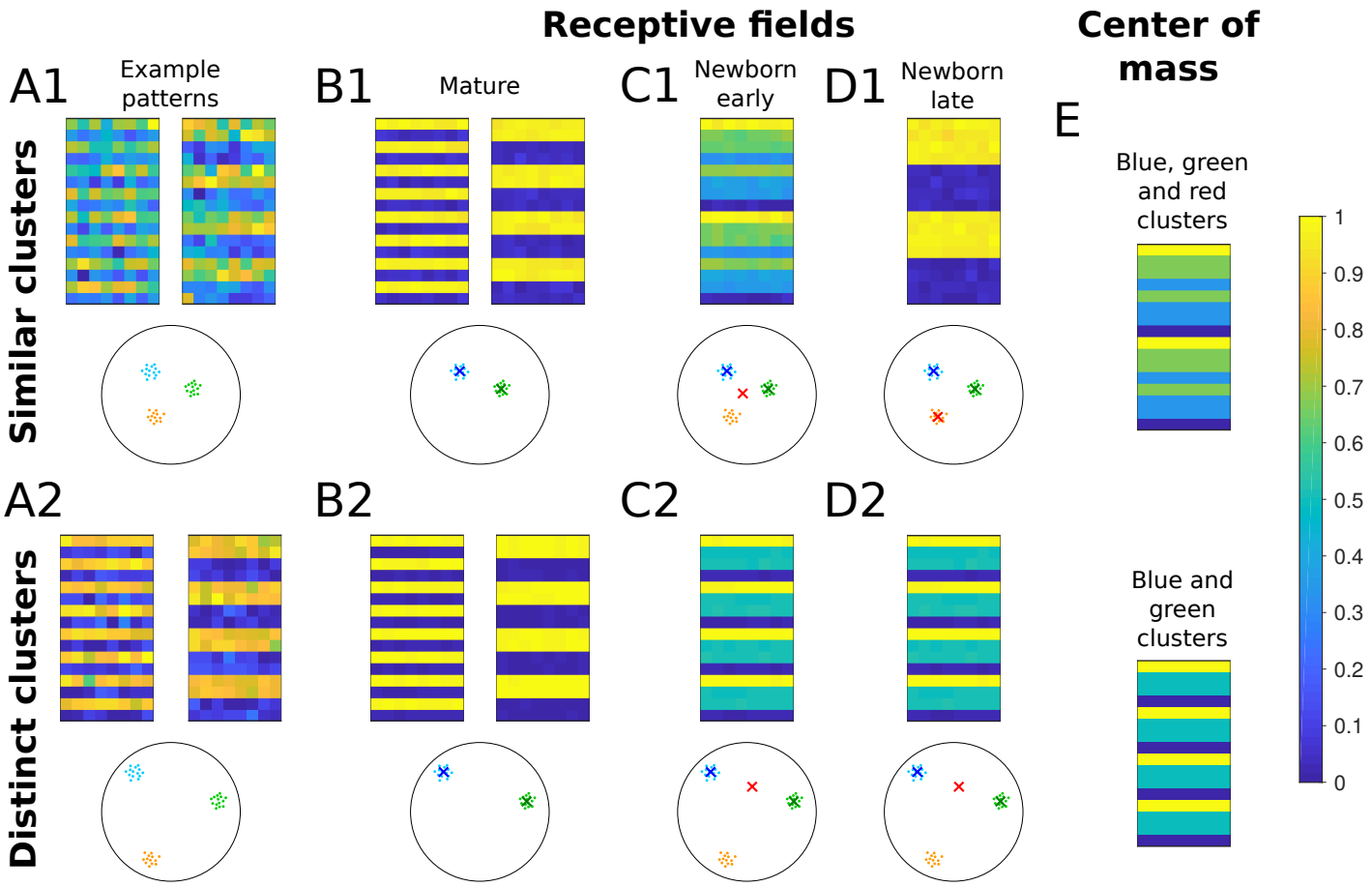

Figure 4: A newborn DGC becomes selective for similar but not distinct novel stimuli. (A) Example patterns from clusters 1 and 2 if clusters are similar (A1, $\xi = 0.2$), and distinct (A2, $\xi = 0.8$). The receptive fields of the mature DGCs are initialized at example patterns. The schematics below 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. For more straightforward comparison of the receptive fields, they are all normalized between 0 and 1: $w_{ij}^{\text{norm}} = \frac{w_{ij} - \min_j(w_{ij})}{\max_j(w_{ij}) - \min_j(w_{ij})}$. (E) Center of mass of all patterns in the first two and first three clusters, normalized to a maximum value of unity (color scale).

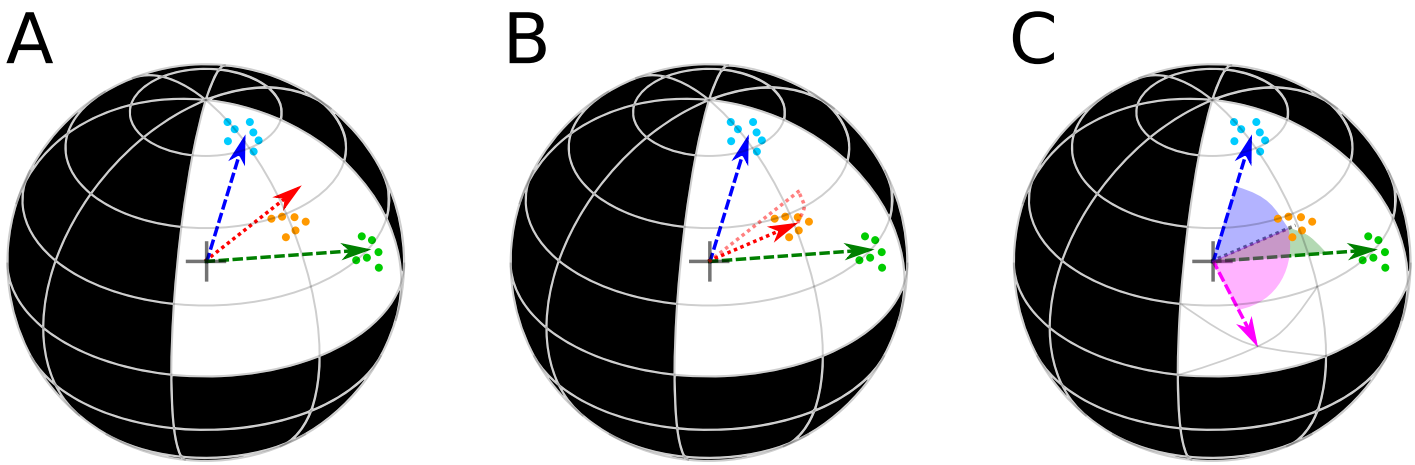

Figure 5: Schematics of the hypersphere surface that contains three clusters of patterns (colored dots) and the feedforward weight vectors towards the output units (colored arrows, ending slightly below the surface). The blue and green clusters are stored before patterns from the novel red cluster are 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.