bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Can neurogenesis act as a neural regularizer?

Lina M. Tran[1-3], Adam Santoro[4], Lulu Liu[1], Sheena A. Josselyn[1-2,5,6], Blake A. Richards[7,8,9,10], Paul W. Frankland[1-2,5,6,11]

[1] Neuroscience and Mental Health, The Hospital for Sick Children, Toronto, ON, Canada

[2] Dept. of Physiology, University of Toronto, Toronto, ON, Canada

[3] Vector Institute, Toronto, ON, Canada

[4] DeepMind, London, United Kingdom

[5] Dept. of Psychology, University of Toronto, Toronto, ON, Canada

[6] Institute of Medical Sciences, University of Toronto, Toronto, ON Canada

[7] Department of Neurology & Neurosurgery, McGill University, Montreal, QC, Canada

[8] School of Computer Science, McGill University, Montreal, QC, Canada

[9] Mila, Montreal, QC, Canada

[10] Learning in Machines and Brains, Canadian Institute for Advanced Research Toronto, ON, Canada

[11] Child and Brain Development Program, Canadian Institute for Advanced Research Toronto, ON, Canada

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

New neurons are continuously generated in the subgranular zone of the dentate gyrus throughout adulthood. These new neurons gradually integrate into hippocampal circuits, forming new naïve synapses. Viewed from this perspective, these new neurons may represent a significant source of 'wiring' noise in hippocampal networks. In machine learning, such noise injection is commonly used as a regularization technique. Regularization techniques help prevent overfitting training data, and allow models to generalize learning to new, unseen data. Using a computational modeling approach, here we ask whether a neurogenesis-like process similarly acts as a regularizer, facilitating generalization in a category learning task. In a convolutional neural network (CNN) trained on the CIFAR-10 object recognition dataset, we modeled neurogenesis as a replacement/turnover mechanism, where weights for a randomly chosen small subset of neurons in a chosen hidden layer were re-initialized to new values as the model learned to categorize 10 different classes of objects. We found that neurogenesis enhanced generalization on unseen test data compared to networks with no neurogenesis. Moreover, neurogenic networks either outperformed or performed similarly to networks with conventional noise injection (i.e., dropout, weight decay, and neural noise). These results suggest that neurogenesis can enhance generalization in hippocampal learning through noise-injection, expanding on the roles that neurogenesis may have in cognition.

Author Summary

In deep neural networks, various forms of noise injection are used as regularization techniques to prevent overfitting and promote generalization on unseen test data. Here, we were interested in whether adult neurogenesis– the lifelong production of new neurons in the hippocampus– might similarly function as a regularizer in the brain. We explored this question computationally, assessing whether implementing a neurogenesis-like process in a hidden layer within a convolutional neural network trained in a category learning task would prevent overfitting and promote generalization. We found that neurogenesis regularization was as least as effective as, or more effective than, conventional regularizers (i.e., dropout, weight decay and neural noise) in improving model performance. These results suggest that optimal levels of

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

hippocampal neurogenesis may improve memory-guided decision making by preventing overfitting, thereby promoting the formation of more generalized memories that can be applied in a broader range of circumstances. We outline how these predictions may be evaluated behaviorally in rodents with altered hippocampal neurogenesis.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Noise reflects random or unpredictable fluctuations that are not part of a signal (Faisal et al 2008). Within the brain there are multiple sources of noise, including processes at the cellular (e.g., protein production and degradation), electrical (e.g., membrane potential) or synaptic levels (e.g., vesicular release) that collectively impact the probability and timing of action potentials (Faisal et al., 2008; McDonnell & Ward, 2011). While neural noise is often considered an obstacle in extracting relevant information from the brain's output activity, optimal levels of neural noise (encapsulated in a broad range of phenomena termed stochastic facilitation (McDonnell & Ward, 2011)) may enhance information transmission and behavior. Similarly, in machine learning noise can also be used to achieve better performance. One of the most common examples is that the addition of some optimized level of noise enhances a model's ability to avoid overfitting and enhances generalization (i.e., avoiding the memorization of training data that is not beneficial to the transfer of learning to unseen data) (Hinton & Camp, 1993). This process of preventing overfitting in order to enhance generalization on unseen data is termed regularization. We, and others, have suggested that neural noise may be one such strategy by which the brain performs regularization to better extract the statistical regularities of our experiences (Hoel, 2021; Richards & Frankland, 2017).

In the hippocampus, another potential source of neural noise is ongoing neurogenesis. Asymmetric division of neural precursor cells in the subgranular zone gives rise to newborn neurons that synaptically integrate into hippocampal networks throughout life (Abrous et al., 2005; Denoth-Lippuner & Jessberger, 2021; Gonçalves et al., 2016; Ming & Song, 2005). This process may be considered a form of noise injection in two respects. First, these new neurons are naive and therefore do not encode any aspects of past experience. Second, their integration gradually reconfigures hippocampal networks. Therefore, we hypothesize that this form of 'wiring' noise may function to regularize learning in the hippocampus, and here we will explore these ideas computationally using categorical learning tasks.

To study whether neurogenesis can act as a regularizer, we implemented neurogenesis in hidden layers of a traditional deep learning architecture (i.e., a convolutional neural network [CNN]). In adult rodents, loss of developmentally-generated granule cells is

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

balanced by the addition of new neurons (Ciric et al., 2019; Cole et al., 2020). Therefore, we modeled neurogenesis as a "replacement/turnover" mechanism, where a randomly chosen small subset of neurons in the middle layer is "turned over" such that their input and output weights are re-initialized to new values (whereas connections of mature neurons remain the same). This turnover/replacement model ensures that the size of the network remains constant. From a computational perspective, this controls for the fact that networks of different sizes can perform very differently (since larger networks have more tunable parameters). From a biological perspective, this mimics the rodent hippocampus where neurogenesis produces negligible net growth (i.e., increase in total number of granule cells) during adulthood (Cole et al., 2020). While deep learning models with neurogenesis have been previously developed (Aimone, 2016; Draelos et al., 2017), these did not explicitly evaluate the role of neurogenesis in generalization.

Results

Neurogenesis improves generalization in CNNs

We implemented neurogenesis in a CNN trained and tested on the CIFAR-10 dataset (Krizhevsky, 2009). CIFAR-10 consists of 60,000 32 × 32 pixel RGB images in 10 different classes representing airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. The CNN consists of 64, 64, and 128 filters, respectively (see Methods for more information), followed by three fully connected layers (**Figure 1A-B**). In order to model neurogenesis, weights of a randomly selected subpopulation of neurons in the middle layer of the fully-connected layers were reinitialized continuously during training, with multiple turnover events. Following hyperparameter tuning, we found that turning over 3.2% of neurons every 640 mini-batch updates to the model had the greatest impact on performance of the network (**Figure 1C**). Neurogenesis was restricted to only one layer, mimicking the occurrence of neurogenesis only in the dentate gyrus layer of the hippocampus. We found that neurogenesis in the second fully connected output layer of the network performed the best during hyperparameter tuning, and therefore used this configuration for all experiments. The performance of neural networks is also dependent on initialization (with some random initializations outperforming others even after undergoing the same training process). To control for potential differences in initialization states between groups, each network initialization

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

was copied and these identical versions were used in the control (no neurogenesis) and experimental (neurogenesis) networks, respectively.

We found that neurogenesis improved network performance on unseen test data. Whereas test accuracy for the control network was 74.36 ± 0.16 %, networks trained with neurogenesis had a test score of 76.20 ± 0.20 % (**Figure 2A**). This improvement in performance with neurogenesis did not depend on initialization states. In particular, the performance-enhancing effects of neurogenesis were not limited to low-scoring variants. Higher-scoring variants also benefited from neurogenesis, indicating that neurogenesis can improve generalization beyond the top scores that a static, non-neurogenic network can achieve (**Figure 2B**).

Adult-generated neurons are more excitable than their developmentally-generated counterparts, with excitability peaking between 4-8 weeks of cell age (Dieni et al., 2016; Doetsch & Hen, 2005; Marín-Burgin et al., 2012; Mongiat et al., 2009; Schmidt-Hieber et al., 2004). To address whether elevated excitability of new neurons might further promote regularization and improve the performance of neurogenic networks, we increased the activation of new neurons by 30% during each forward pass during training (**Figure 2C**). To reflect the transient nature of these changes, following each round of new neuron turnover, excitability of previously turned over neurons returned to baseline. As before, we found that neurogenic networks outperformed our control networks (i.e., non-neurogenic networks) on held-back test data. However, incorporating excitability into the neurogenic CNN did not further improve performance (i.e., neurogenesis + excitability) (**Figure 2D**).

Neurogenesis improves CNN performance via regularization

The improved performance on held-back test data might reflect a more powerful model in the neurogenesis group, rather than a regularization effect. In other words, it could be that learning in general is simply better in the neurogenesis networks, in the same way that networks with more layers tend to learn better. Indeed, in adult rodent studies interventions that elevate hippocampal neurogenesis facilitate learning in many (Creer et al., 2010; Sahay, Scobie, et al., 2011) situations. (Though it should be noted that this does not actually hold in all situations (Frankland, 2013)). Similarly, in artificial neural networks, implementation of neurogenesis may improve learning in general (Chambers

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

et al., 2004; Deisseroth et al., 2004; Meltzer et al., 2005), rather than having a specific effect on generalization. To assess this, we compared training accuracy of the control and neurogenic networks, anticipating that training accuracy would be the same, or even reduced, in the neurogenesis network should improved performance be due to regularization. We found that the neurogenic networks do not improve training accuracy, and in fact, they sacrifice training accuracy in order to achieve the previously observed higher test accuracy on held-back data. Neurogenic networks achieve lower validation loss, but higher training loss relative to control, suggesting that the enhancement in performance is indeed due to a regularization effect (**Figure 2E,F**).

Neurogenesis regularization achieves similar level improvement to dropout

We next compared neurogenesis to conventional regularization methods, including dropout, weight decay, and neural noise (**Figure 3A-C**). Dropout involves the stochastic silencing of a subset of units during each forward pass (Srivastava et al., 2014). Weight decay involves adding a small penalty to the loss function that penalizes large weights, thus resulting in an overall decay of larger weights (Krogh & Hertz, 1992). Neural noise involves the addition of Gaussian noise to all the activations during each forward pass (Bishop, 1995). After identifying the optimal parameter values for each of these regularization methods, we compared the performance of neurogenic networks to networks regularized with these other methods. We found that neurogenic CNNs performed similarly to dropout, and outperformed weight decay and neural noise (**Figure 3D**).

Mechanistically, there are similarities between replacement/turnover neurogenesis and dropout, with both involving information loss at each update. In dropout, a subpopulation of neurons is transiently silenced, but connection weights are otherwise maintained. In contrast, replacement/turnover neurogenesis involves resetting the weights of a subpopulation of neurons. Interestingly, the optimal size of these subpopulations differs markedly for dropout vs. neurogenesis. Optimal performance occurred when dropout was implemented in a subpopulation that corresponded to ~20% of hidden layer neurons, whereas for neurogenesis, optimal performance was achieved when the turnover subpopulation was ~3.2% of hidden layer neurons.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We next asked whether combining regularization techniques might further enhance performance. Surprisingly, neurogenesis, when combined with dropout, weight decay or neural noise, consistently reduced the performance (**Figure 3E**). This suggests that the amount of noise injected by neurogenesis may be suboptimal when combined with other regularization techniques. Consistent with this idea, performance improved when the amount of noise injection for each regularization method was reduced. However, even using these reduced parameter values did not enhance performance beyond that of neurogenesis alone, suggesting that there is a ceiling effect beyond which performance cannot be improved further (**Figure 3F**).

Targeted neurogenesis in CNNs

In our model, neurogenesis occurs in randomly selected subpopulations of middle layer neurons. However, the integration of new neurons might be non-random in nature. For example, in rodents there is evidence for neurogenesis-dependent refinement of synaptic connections in the hippocampus, with the integration of new neurons leading to the elimination of less active synaptic connections (Yasuda et al., 2011). Here we explored whether implementation of a similarly targeted turnover mechanism during training would further enhance performance of our neurogenic networks. Such an approach has previously been explored using dropout regularization, and, in this case, targeting dropout to neurons that are likely to be less important for the task (i.e., neurons with lowest L1 norm values or total weights) led to significant improvement in test performance (Gomez et al., 2019). Using a similar strategy, here we ranked neurons by their L1 norm values at every turnover event, and reinitialized weights of the bottom 3.2% ranking neurons. As a positive control, we reinitialized the weights of the top 3.2% ranking neurons in a separate experiment (**Figure 4A**).

Targeting top-ranked neurons impaired generalization performance below that of the control (no neurogenesis) group. This is expected since these neurons are assumed to be more valuable to the performance of the network. However, targeting the bottom-ranked neurons did not further enhance performance (i.e, compared to random reinitializations affecting the same size subpopulation) (**Figure 4B**). While this finding indicates that targeting turnover neurogenesis to the bottom L1 norm ranking neurons does not further improve generalization, we recognize that there may be alternate ways

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to determine 'importance' of units in a neurogenic network, and it is possible that targeting these might produce performance improvements.

Neurogenesis improves generalization but increases reliance on individual neurons

The degree to which networks depend on single units vs. more distributed, population codes influences generalization (Morcos et al., 2018). Typically, networks that generalize well on held back, unseen test data depend on distributed population codes, rather than a small subset of units. Moreover, because of their distributed nature, these types of networks tend to be more resilient to random ablation of units (Morcos et al., 2018). Conversely, networks that generalize poorly tend to depend on a small subset of units, rather than distributed codes, and these networks are typically less resilient to random ablation of units.

We performed a similar analysis here to assess whether neurogenesis improves generalization by reducing the dependence of networks on a small subset of single units. To evaluate the reliance on single units vs. distributed codes, we sequentially ablated random units following training and compared pre- and post-ablation performance (accuracy was normalized to pre-ablation performance) (**Figure 5A**). Surprisingly, we found that neurogenic networks were less resilient. Ablation of a smaller proportion of units was sufficient to decrease performance (as reflected by the leftward shift of the curve) (**Figure 5B**). This indicates that neurogenesis improves generalization through a mechanism other than reducing the network's reliance on individual neurons. We also measured the class selectivity for each neuron using the metric described in Morcos et al., (2018) to test whether there were fewer highly class selective neurons. We found that the distribution of class selectivities of neurons was shifted to the left (on average neurons were less selective for class) in the neurogenesis group compared to dropout or control networks (**Figure 5C**). To assess whether the reduced resilience to ablation of neurogenic networks translates to increased vulnerability to neurogenesis post-training, we trained networks as before, with and without neurogenesis, and tested performance with and without post-training neurogenesis. One turnover event (replacing 8 neurons) after training/at test-time was used. We found that post-training neurogenesis reduced performance of the neurogenic networks, but not the control networks, further

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

demonstrating a specific sensitivity to perturbation in neurogenic networks (**Figure 5D**).

Discussion

As adult-generated neurons integrate into hippocampal circuits they form naive synapses and therefore can be thought of as a form of 'wiring' noise. Deep learning has found that various forms of noise-injection can reduce overfitting on training data and, as a result, enhance generalization in deep neural networks. We therefore hypothesized that neurogenesis-mediated rewiring would similarly have a regularization effect, i.e., prevent memorization of training data and favor a more flexible, generalized memory that can be applied in a broader range of circumstances. We explored this hypothesis computationally, implementing neurogenesis in a hidden layer of a CNN trained on the CIFAR-10 object recognition task. Consistent with our hypothesis, neurogenesis acted as a regularizer, improving generalization on the test (held-out) data. Performance with neurogenesis regularization matched (dropout) or outperformed (weight decay, neural noise) other conventional regularization techniques.

In our model, we implemented neurogenesis as a replacement/turnover mechanism, where input and output synaptic weights associated with small subsets of middle layer neurons were re-initialized through training. We used this approach since neurogenesis in the rodent hippocampus similarly involves replacement of mature, developmentally-generated neurons with immature, adult-generated neurons, and, therefore, turnover of associated synaptic weights. For instance, in the rat adult dentate gyrus there is significant loss of mature granule cells that were born soon after birth (Ciric et al., 2019). Since there is negligible overall increase in the size of the dentate granule cell layer during adulthood (Rapp & Gallagher, 1996), this implies that this loss of developmentally-generated granule cells in adulthood is balanced by new neuron addition (Cole et al., 2020). Viewed in this way, neurogenesis regularization can be thought of as a form of wiring noise that incrementally alters network connectivity patterns, without impacting overall network size.

In the CNN, we explored whether targeting neurogenesis regularization to less important neurons was beneficial. Using L1 norm values as a metric of "importance" (Gomez et al., 2019), we assessed generalization following three types of neurogenesis

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

regularization: When neurogenesis regularization was applied randomly to all middle layer units versus when neurogenesis regularization was either restricted to the least (i.e., neurons with lowest L1 norm values or weakest weights) or the most (i.e., neurons with highest L1 norm values or strongest weights) important middle layer neurons. As expected, targeting neurons with the highest L1 norm scores decreased generalization scores below control (no neurogenesis) levels, consistent with the idea that neurons with the highest L1 norm values are indeed more important for the task. However, targeting neurogenesis regularization to neurons with the lowest L1 norm scores did not improve generalization beyond networks with non-targeted or random neurogenesis regularization.

In the rodent brain, there is some evidence that the integration of new neurons is a non-random process, with the integration of new neurons leading to the pruning of less active synaptic connections (Yasuda et al., 2011). While our current analyses suggest that such a targeted mechanism may not be critical for improving generalization, nonetheless it is possible that using the L1 norm of each unit may not be the most suitable metric of importance. Alternatively, the lottery ticket hypothesis suggests that within these networks there exist sparse sub-networks that, when trained in isolation, can achieve the same final performance accuracy of the entire network but in the same or fewer training epochs (Frankle & Carbin, 2019). Interestingly, the rates at which these lottery-winning sub-network neurons change weights is much higher compared to other neurons. Therefore, one possibility would be to target neurogenesis regularization to neurons that change their weights the least (i.e., those that contribute least to the loss function) during training. Simply looking at the L1 norm, which is not sensitive to the rate of weight changes, would not capture the same "importance" as described in the lottery ticket hypothesis.

A second issue we explored in the CNN was whether neurogenesis regularization impacts how information is organized within the network. Typically, better generalization is associated with more distributed coding (Morcos et al., 2018). Following dropout regularization, for example, information tends to be coded in a more distributed manner, rather than in single units. One consequence of this organization is that networks tend to be more resilient to random, sequential ablation of units following dropout regularization. In contrast, we found that there was greater reliance on single

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

units following neurogenesis regularization, and networks regularized this way were more vulnerable to random ablations (despite better generalization performance).

One way these findings might be viewed is that while neurogenesis regularization improves generalization it also introduces network vulnerabilities. That is, randomly turning over units pushes the network to rely on neurons that are tuned to single directions (via an unknown mechanism). This may improve generalization performance, but subsequent, random turnover events can also eliminate neurons that are highly-tuned to single directions, and hence, have catastrophic consequences. The idea that neurogenesis might introduce network vulnerabilities is interesting since neurogenesis has been linked to forgetting (Frankland et al., 2013; Ryan & Frankland, 2022). For example, post-training increases in neurogenesis induce forgetting of established hippocampus-dependent memories in adult rodents (Akers et al., 2014; Epp et al., 2016; Gao et al., 2018). Whether (or not) forgetting happens may depend both on the levels and/or the timing of neurogenesis.

With respect to levels of neurogenesis, whereas low levels of neurogenesis may promote overfitting (i.e., memorization or no forgetting), high levels of neurogenesis may promote underfitting (i.e., forgetting). In between these extremes, moderate levels of neurogenesis may prevent overfitting and promote generalization (Ko & Frankland, 2021). According to this scenario, it may be the case that neurogenesis levels need to be tightly regulated in order to balance the costs (e.g., overfitting, forgetting) vs. benefits (improved generalization) of rewiring (Richards & Frankland, 2017).

Our analyses suggest that the timing of increases in neurogenesis (with respect to training) also might matter. When neurogenesis occured in concert with training, we found this led to improved performance on subsequent test data. However, when neurogenesis-like turnover occurs at the time of testing we observed a decrease in performance on the unseen test data. This suggests that neurogenic networks need to keep training if they are to retain high performance. Viewed this way, it might be useful to distinguish between the effects of ongoing neurogenesis on active memories (i.e., those that are undergoing training and invulnerable) vs. inactive (i.e., those where training is ‘complete’ and therefore are vulnerable).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

While our model captures replacement/turnover as a core feature of neurogenesis, we nonetheless recognize that it is highly abstracted and differs significantly from biological networks in terms of architecture, connectivity and sparsity. Another feature that was not captured in the current model, but is relevant especially in the context of consolidation, is replay. During sleep, previous event sequences are "replayed" in the hippocampus and this may provide an opportunity to integrate new experiences with prior, relevant experience and promote generalization (Ji & Wilson, 2007; Louie & Wilson, 2001). Indeed, REM sleep, the period of sleep during which replay events typically occur (Louie & Wilson, 2001), is associated with enhanced memory generalization in humans (Lerner et al., 2021). Interestingly, adult-born dentate granule cells that were active during learning are reactivated during subsequent REM (Kumar et al., 2020). This suggests that new neurons may indeed be contributing to consolidation as a source of noisy replay, potentially promoting neural regularization and generalization of hippocampal memories. Consistent with this view, in a computational model, O'Donnell *et al.* found that noisy replay led to an relative increase in the overlap between input patterns and a particular target pattern (O'Donnell & Sejnowski, 2014). Such a mechanism might underlie generalization through broadening the types of inputs that would drive activation of a given memory.

What are the implications for functional studies of hippocampal neurogenesis? The functional consequences of altering hippocampal neurogenesis (or manipulating the activity of adult-generated granule cells) have largely been studied in rodents (Anacker & Hen, 2017; Cameron & Glover, 2015; Gonçalves et al., 2016). A major focus of these studies has been on the role of hippocampal neurogenesis in pattern separation (Sahay, Wilson, et al., 2011; Santoro, 2013). In a typical experiment, an experimental intervention is introduced to alter levels of hippocampal neurogenesis, and then the ability of mice or rats to make fine spatial or contextual discriminations is assessed. In a touchscreen apparatus, for example, this might involve discriminating between stimuli presented in different spatial locations (Oomen et al., 2013). Rodents with reduced neurogenesis perform poorly under conditions where spatial similarity is high, whereas rodents with elevated neurogenesis exhibit enhanced discrimination (Clelland et al., 2009; Creer et al., 2010).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

These results are consistent with the idea that hippocampal neurogenesis regulates a pattern separation-like process in the hippocampus. However, directly comparing these results to our model is difficult given that spatial discrimination is a computationally different task from category discrimination, as we studied here. While studies of category learning are common in monkeys (Ashby & Spiering, 2004), investigations of category learning have been less common in rodents (where neurogenesis levels can be more readily manipulated). Nonetheless, both object recognition and touchscreen-based category learning tasks have been developed for rodents (Broschard et al., 2021; Creighton et al., 2019). The object category recognition tasks take advantage of rodents' innate preference for novelty. During the study phase, mice are allowed to explore two objects from the same category (e.g., two different toy cars). During the test phase, mice are presented with a choice between a third object from the studied category (i.e., another toy car) and a novel object from a new category (e.g., hair clip). Should the mouse exhibit a preference for the object from the unstudied category (i.e., hair clip), then this suggests some within category generalization (i.e., across different types of toy cars) (Creighton et al., 2019). In the touchscreen-based category learning task, mice are trained to discriminate between 2D visual stimuli presented on a touchscreen that can be categorized according to spatial features (e.g., spatial frequency and orientation of gratings). Generalization is assessed when rodents are subsequently tested on novel visual stimuli from the studied categories (Broschard et al., 2021). Based on the findings presented here, we predict that suppression of adult hippocampal neurogenesis will impair generalization in such tests, whereas increasing adult hippocampal neurogenesis will facilitate performance. All of this would be consistent with the effects proposed for neurogenesis in the pattern separation literature.

Neurogenesis has previously been modeled in the context of deep learning in the form of an autoencoder that can perform continual learning. Neurogenesis was found to prevent memory interference of learning new classes on previously learned classes in an image reconstruction task (Draelos et al., 2017). This aligns with other literature where neurogenesis has been shown to improve memory capacity and prevent catastrophic interference (Appleby & Wiskott, 2009; Finnegan & Becker, 2015; Wiskott et al., 2006). Here, we found that neurogenesis can be applied to CNNs as a novel regularization technique to improve generalization in classification tasks. While the performance of turnover models was superior to neural noise and weight decay, turnover models did

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

not exceed the performance of other regularization methods (namely, dropout). While both methods introduce randomness via the architecture of the network during training, turnover differs from dropout in that turnover consists of the simultaneous permanent loss of learned weights and resetting of weights during the course of training, whereas dropout temporarily silences a random population at each forward pass during training. Neurogenesis as a regularization technique may benefit further by using a different targeted approach to which neurons are turned over, as discussed above. However, there is also the question as to the mechanism by which neurogenesis improves performance. It is possible that by randomly resetting some neurons during training, these neurons' weights have the opportunity to be arranged in such a way that is more beneficial to training than their previous weights.

Conclusions. The current analyses provide evidence that 'wiring noise' in the hippocampus, in the form of ongoing neurogenesis, provides a means to regularize memories, thereby preventing memorization and promoting generalization. These ideas are consistent with recent computational and imaging evidence that the hippocampus supports statistical learning/generalized memories, in addition to detailed memories, in humans (Schapiro & Turk-Browne, 2015; Sučević & Schapiro, 2022), and they provide a potential mechanism. Because neurogenesis can be experimentally manipulated in rodents, these predictions can be evaluated using category learning (or related) paradigms in the future.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

Software

Code for these methods is available on GitHub (https://github.com/linamnt/dnn-neurogenesis).

Python

Models were built and analyzed in Python 3.6 (Van Rossum & Drake, 2009) with custom scripts that are freely available on GitHub, and were developed using the following packages: PyTorch (Paszke et al., 2019), Ax (https://github.com/facebook/Ax), NumPy (Oliphant, 2006), SciPy (Virtanen et al., 2020), Pandas (McKinney & others, 2010), Matplotlib (Hunter, 2007), Seaborn (Waskom et al., 2017), Scikit-learn 0.21.1 (Pedregosa et al., 2011).

Computing resources

These experiments were implemented on the high-performance compute clusters at Compute Canada and Vector Institute for AI.

Convolutional neural network

The CNN involves convolutional layers that have filters that convolve to extract features of the image that are not affected by translation. Each convolutional layer is followed by a pooling layer that downsamples the images to improve computational efficiency. Finally, after multiple sets of convolutional and pooling layers, there are fully connected layers. The input layer was of size $32 \times 32 \times 3$, corresponding to the images (with three color channels) from the CIFAR-10 dataset. The CNN was built with three sets of two convolutional layers, each followed by a max-pooling layer. The convolutional layer sets had 16, 32, and 64 filters, respectively, with a 3×3 filter size and stride (steps the filter moves along) of 1. The max-pooling layers pool a 2×2 region and are applied at strides of 1. The convolutional layers are then connected to three fully connected layers, each with 250 units, using a rectified linear activation function. Neurogenesis occurred in the middle, fully connected layer. Data was split into training (40,000 images), validation (10,000 images) and test sets (10,000 images). Test results are derived from networks

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

trained on the full training set. We tuned the network's hyper-parameters by training on the training set (40, 000 images), and testing on the validation set (10, 000 images). The hyper-parameters, i.e., values that regulate learning in the networks, and other information about the networks are listed in *tbl. 1*.

CIFAR-10

The CIFAR-10 dataset is a collection of images that consists of 50,000 training and 10,000 test images for each of ten classes commonly used to train machine learning and computer vision models. The images are 32×32 color images with the 10 different classes representing airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships and trucks.

Table 1: Table of CNN model parameters.

Hyperparameter	**CNN**
Batch size	4
Learning rate	0.0002
Epochs	14
Turnover Proportion	0.032
Turnover Frequency	1/640 batch updates
Dropout	0.2
Weight Decay	0.00001
Neural Noise (log normal: mean, std)	-0.2, 0.5
Excitation Factor (c)	1.3

Training neurogenic networks

We found that neural networks with neurogenesis can be trained using stochastic gradient descent (Robbins, 2007). During training, neurogenesis was implemented on

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

an ongoing basis, with multiple turnover events occurring. Neurogenesis occured in the second hidden layer (of three). Neurogenesis was implemented as a replacement/turnover mechanism, whereby a randomly chosen subset of neurons in the layer was "turned over" such that their weights were re-initialized to new values. In contrast, the other neurons maintained their learned weights. We initialized the new neurons' weights using the same function used when randomly initializing the network at the start of training, i.e., Kaiming uniform initialization (He et al., 2015). Turnover can occur at a frequency of once every *n* batch updates. We tuned the hyperparameters using Bayesian Optimization which is an iterative parameter tuning process which builds a probability model for the best parameters to try next (Snoek et al., 2012), whereby we input the turnover frequency (ranging from once every update to once every 12500 updates), and how many neurons to add at each turnover event (ranging from 0 to 250 neurons) as parameters to search for. We found that turning over 8 neurons every 640 updates in the second fully-connected layer resulted in the highest performance on the validation set. We implemented neurogenesis in the output layers of a CNN.

Enhancing excitability

To evaluate whether manipulating "excitability" of new neurons might impact regularization, in a separate experiment we incorporated an excitability component into our neural networks in each forward pass during training. We multiplied the activations of the neurogenesis layer by an excitability array. The activations corresponding to the new neurons were set to a factor (1.3; determined by hyperparameter tuning), and all other activations were set to 1. As a result, only the activations corresponding to the turned-over neurons were enhanced, and the remainder of the neurons' activations were unchanged. Whenever a new set of neurons were turned over, the previous set of new neurons became "mature" and were no longer "excited" in the forward pass.

Other Regularization Methods

To compare the performance of neurogenesis to other regularization methods, we also used dropout, weight decay and neural noise in the CNNs. Dropout involves the stochastic silencing of a subset of units during each forward pass (Srivastava et al., 2014). We used a dropout rate of 0.2 in CNNs, with an additional 7 epochs required to

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reach the end of training, compared to non-dropout networks. Weight decay consists of adding a small penalty to the loss function that penalizes large weights, thus resulting in an overall decay of larger weights (Krogh & Hertz, 1992). We used a weight decay value of 0.00001 for our CNN. Neural noise involves the addition of Gaussian noise to all the activations during each forward pass (Bishop, 1995). We defined our noise using a mean of -0.2, and a standard deviation of 0.5 on a log normal distribution for our CNN.

Ablation experiments

To measure the importance of individual neurons in a network, we tested how much the network's performance degrades as progressively more neurons are removed from the network (based on Morcos et al. (2018)). To remove a neuron, we set that neuron's activity to a fixed value of 0, effectively ablating the unit. We progressively ablated neurons in proportional steps of 5% of the neurons in the neurogenesis layer, testing the accuracy of the training data at each step to generate ablation curves. We repeated each ablation five times and randomized the order of neurons ablated each time. Ablation curves plot the degradation in accuracy as more neurons are ablated. Networks that rely more heavily on individual units will drop their accuracy more quickly as units are ablated. Networks that are less sensitive to ablations have been shown to correlate with better generalization (Morcos et al., 2018).

Class selectivity

Class selectivity was calculated using the method described in Morcos et al., 2018. The class-conditional mean activity was calculated from the test set and the class selectivity index was measured as

$$selectivity = \frac{\mu_{□□□} - □_{-□□□}}{□_{□□□} + □_{-□□□}}$$

where $\mu_{□□□}$ is the highest mean class activity and $\mu_{-□□□}$ is the mean class activity across all other classes.

Targeted neurogenesis

To test whether a targeted approach to removing new neurons might improve the performance of neurogenic networks, we implemented a targeted turnover of new neurons. Based on the work of Gomez et al. (2019)(Gomez et al., 2019), we adapted the

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

targeted unit dropout technique to choose which units to replace in the network with naive units. It was found that this method of targeting neurons for dropout performed better than a random dropout process for identifying pruned networks. While identifying pruned networks is not the goal of our work, we could use their methods to identify important and unimportant units to target for turnover. The unit's importance was determined using the unit L1 norm (i.e., the sum of the absolute values of all the weights inputting onto the unit) and ranking them from lowest to highest value and importance. To test whether targeting neurogenesis to the low importance neurons would improve performance, a proportion of the lowest ranking neurons matching the proportion of neural turnover had their weights reset. As a positive control, we also targeted the highest importance neurons (highest L1 norm values) for turnover to confirm that this metric does indeed carry information about the importance of neurons for learning.

Statistical analyses and plotting

All statistics were performed using the scipy.stats module in Python 3.6 (Van Rossum & Drake, 2009; Virtanen et al., 2020). Error bars on graphs represent the standard error of the mean across different initializations of the model, where each experiment is repeated 20 times unless otherwise stated. Comparisons were made using unpaired, two-tailed t-tests, or analysis of variance (ANOVA) followed by Tukey's HSD post-hoc tests where appropriate. Significance indicated by an asterisk (*) for p-values less than 0.01 unless otherwise stated. Graphs were generated using the matplotlib, and seaborn packages in Python 3.6 (Hunter, 2007; Waskom et al., 2017), and figures were compiled using Inkscape (Inkscape Project, 2020).

Acknowledgements

We thank Jason Snyder for comments on earlier drafts of this manuscript. This work was supported by a Canadian Institutes of Health Research (CIHR; FDN143227) grant to PWF and the Canadian Institute for Canadian Institute for Advanced Research (CIFAR) catalyst grant to PWF, BAR and SAJ. LMT was supported by fellowships from The Hospital for Sick Children, the Vector Institute and the Natural Sciences and Engineering Research Council of Canada (NSERC). BAR and PWF are fellows in the Learning in Machines & Brains and the Child, Brain & Development programs, respectively, at CIFAR.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Author contributions

LMT contributed to conceptualization, data curation, formal analyses, investigation, methodology, software, validation, visualization, and writing (original draft preparation, review and editing). AS contributed to conceptualization and writing (review and editing). LL contributed to formal analysis and software. SAJ contributed to funding acquisition and writing (review and editing). BAR contributed to conceptualization, funding acquisition and writing (review and editing). PWF contributed to conceptualization, supervision, funding acquisition and writing (original draft preparation, review and editing).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Abrous, D. N., Koehl, M., & Le Moal, M. (2005). Adult Neurogenesis: From Precursors to Network and Physiology. *Physiological Reviews*, *85*(2), 523–569. https://doi.org/10.1152/physrev.00055.2003

Aimone, J. B. (2016). Computational Modeling of Adult Neurogenesis. *CSH Perspectives*, 1–14.

Akers, K. G., Martinez-canabal, A., Restivo, L., Yiu, A. P., Cristofaro, A. D., Hsiang, H. L., Wheeler, A. L., Guskjolen, A., Niibori, Y., Shoji, H., Ohira, K., Richards, B. A., & Miyakawa, T. (2014). Hippocampal neurogenesis regulates forgetting during adulthood and infancy. *Science*, *344*(May), 598–602. https://doi.org/10.1126/science.1248903

Anacker, C., & Hen, R. (2017). Adult hippocampal neurogenesis and cognitive flexibility—Linking memory and mood. *Nature Reviews Neuroscience*, *18*(6), 335–346. https://doi.org/10.1038/nrn.2017.45

Appleby, P. A., & 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. https://doi.org/10.1080/09548980902993156

Ashby, F. G., & Spiering, B. J. (2004). The neurobiology of category learning. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 101–113. https://doi.org/10.1177/1534582304270782

Bishop, C. M. (1995). Training with Noise is Equivalent to Tikhonov Regularization. *Neural Computation*, *7*(1), 108–116. https://doi.org/10.1162/neco.1995.7.1.108

Broschard, M. B., Kim, J., Love, B. C., & Freeman, J. H. (2021). Category learning in rodents using touchscreen-based tasks. *Genes, Brain and Behavior*, *20*(1), e12665. https://doi.org/10.1111/gbb.12665

Cameron, H. A., & Glover, L. R. (2015). Adult neurogenesis: Beyond learning and memory. *Annual Review of Psychology*, *66*, 53–81. https://doi.org/10.1146/annurev-psych-010814-015006

Chambers, R. A., Potenza, M. N., Hoffman, R. E., & Miranker, W. (2004). Simulated Apoptosis/Neurogenesis Regulates Learning and Memory Capabilities of Adaptive Neural Networks. *Neuropsychopharmacology*, *29*(4), 747–758. https://doi.org/10.1038/sj.npp.1300358

Ciric, T., Cahill, S. P., & Snyder, J. S. (2019). Dentate gyrus neurons that are born at the peak of development, but not before or after, die in adulthood. *Brain and Behavior*, *9*(10), e01435. https://doi.org/10.1002/brb3.1435

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Clelland, C. D., Choi, M., Romberg, C., Clemenson, G. D., Fragniere, a, Tyers, P., Jessberger, S., Saksida, L. M., Barker, R. a, Gage, F. H., & Bussey, T. J. (2009). A functional role for adult hippocampal neurogenesis in spatial pattern separation. *Science*, *325*(5937), 210–213. https://doi.org/10.1126/science.1173215

Cole, J. D., Espinueva, D. F., Seib, D. R., Ash, A. M., Cooke, M. B., Cahill, S. P., O'Leary, T. P., Kwan, S. S., & Snyder, J. S. (2020). Adult-Born Hippocampal Neurons Undergo Extended Development and Are Morphologically Distinct from Neonatally-Born Neurons. *Journal of Neuroscience*, *40*(30), 5740–5756. https://doi.org/10.1523/JNEUROSCI.1665-19.2020

Creer, D. J., Romberg, C., Saksida, L. M., Praag, H. van, & Bussey, T. J. (2010). Running enhances spatial pattern separation in mice. *Proceedings of the National Academy of Sciences*, *107*(5), 2367–2372. https://doi.org/10.1073/pnas.0911725107

Creighton, S. D., Collett, H. A., Zonneveld, P. M., Pandit, R. A., Huff, A. E., Jardine, K. H., McNaughton, B. L., & Winters, B. D. (2019). Development of an "object category recognition" task for mice: Involvement of muscarinic acetylcholine receptors. *Behavioral Neuroscience*, *133*(5), 527–536. https://doi.org/10.1037/bne0000331

Deisseroth, K., Singla, S., Toda, H., Monje, M., Palmer, T. D., & Malenka, R. C. (2004). Excitation-Neurogenesis Coupling in Adult Neural Stem/Progenitor Cells. *Neuron*, *42*, 535–552. https://doi.org/10.1016/S0896-6273(04)00266-1

Denoth-Lippuner, A., & Jessberger, S. (2021). Formation and integration of new neurons in the adult hippocampus. *Nature Reviews Neuroscience*, *22*(4), 223–236. https://doi.org/10.1038/s41583-021-00433-z

Dieni, C. V., Nietz, A. K., Panichi, R., Wadiche, J. I., & Overstreet-Wadiche, L. (2013). Distinct determinants of sparse activation during granule cell maturation. *J. Neurosci.*, *33*(49), 19131–19142. https://doi.org/10.1523/JNEUROSCI.2289-13.2013

Dieni, C. V., Panichi, R., Aimone, J. B., Kuo, C. T., Wadiche, J. I., & Overstreet-Wadiche, L. (2016). Low excitatory innervation balances high intrinsic excitability of immature dentate neurons. *Nature Communications*, *7*(1), 11313. https://doi.org/10.1038/ncomms11313

Doetsch, F., & Hen, R. (2005). Young and excitable: The function of new neurons in the adult mammalian brain. *Current Opinion in Neurobiology*, *15*(1), 121–128. https://doi.org/10.1016/j.conb.2005.01.018

Draelos, T. J., Miner, N. E., Lamb, C. C., Cox, J. A., Vineyard, C. M., Carlson, K. D., Severa, W. M., James, C. D., & Aimone, J. B. (2017). Neurogenesis deep learning: Extending deep networks to accommodate new classes. *2017 International Joint Conference on Neural Networks (IJCNN)*, 526–533. https://doi.org/10.1109/IJCNN.2017.7965898

Epp, J. R., Silva Mera, R., Kohler, S., Josselyn, S. A., & Frankland, P. W. (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Neurogenesis-mediated forgetting minimizes proactive interference. *Nat. Commun.*, *7*, 5–12. https://doi.org/10.1038/ncomms10838

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303. https://doi.org/10.1038/nrn2258

Finnegan, R., & Becker, S. (2015). Neurogenesis paradoxically decreases both pattern separation and memory interference. *Front. Syst. Neurosci.*, *9*(October), 1–12. https://doi.org/10.3389/fnsys.2015.00136

Frankland, P. W. (2013). Neurogenic evangelism: Comment on Urbach et al. (2013). *Behavioral Neuroscience*, *127*(1), 126–129. https://doi.org/10.1037/a0031512

Frankland, P. W., Kohler, S., & Josselyn, S. A. (2013). Hippocampal neurogenesis and forgetting. *Trends Neurosci.*, *36*(9), 497–503. https://doi.org/10.1016/j.tins.2013.05.002

Frankle, J., & Carbin, M. (2019). The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. *ArXiv:1803.03635 [Cs]*. http://arxiv.org/abs/1803.03635

Gao, A., Xia, F., Guskjolen, A., Ramsaran, A., Santoro, A., Josselyn, S., & Frankland, P. (2018). Elevation of Hippocampal Neurogenesis Induces a Temporally Graded Pattern of Forgetting of Contextual Fear Memories. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *38*(13), 3190–3198. https://doi.org/10.1523/jneurosci.3126-17.2018

Ge, S., Yang, C., Hsu, K., Ming, G., & Song, H. (2007). A Critical Period for Enhanced Synaptic Plasticity in Newly Generated Neurons of the Adult Brain. *Neuron*, *54*(4), 559–566. https://doi.org/10.1016/j.neuron.2007.05.002

Gomez, A. N., Zhang, I., Kamalakara, S. R., Madaan, D., Swersky, K., Gal, Y., & Hinton, G. E. (2019). Learning Sparse Networks Using Targeted Dropout. *ArXiv:1905.13678 [Cs, Stat]*. http://arxiv.org/abs/1905.13678

Gonçalves, J. T., Schafer, S. T., & Gage, F. H. (2016). Adult Neurogenesis in the Hippocampus: From Stem Cells to Behavior. *Cell*, *167*(4), 897–914. https://doi.org/10.1016/j.cell.2016.10.021

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *ArXiv:1502.01852 [Cs]*. http://arxiv.org/abs/1502.01852

Hinton, G. E., & Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. *COLT '93*. https://doi.org/10.1145/168304.168306

Hoel, E. (2021). The overfitted brain: Dreams evolved to assist generalization. *Patterns*, *2*(5), 100244. https://doi.org/10.1016/j.patter.2021.100244

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, *10*(1), 100–107. https://doi.org/10.1038/nn1825

Jung, M. W., & McNaughton, B. L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*, *3*(2), 165–182. https://doi.org/10.1002/hipo.450030209

Ko, S. Y., & Frankland, P. W. (2021). Neurogenesis-dependent transformation of hippocampal engrams. *Neuroscience Letters*, *762*, 136176. https://doi.org/10.1016/j.neulet.2021.136176

Krizhevsky, A. (2009). *Learning Multiple Layers of Features from Tiny Images* (p. 60).

Krogh, A., & Hertz, J. A. (1992). A Simple Weight Decay Can Improve Generalization. *Advances in Neural Information Processing Systems*, *4*, 950–957.

Krueppel, R., Remy, S., & Beck, H. (2011). Dendritic Integration in Hippocampal Dentate Granule Cells. *Neuron*, *71*(3), 512–528. https://doi.org/10.1016/j.neuron.2011.05.043

Kumar, D., Koyanagi, I., Carrier-Ruiz, A., Vergara, P., Srinivasan, S., Sugaya, Y., Kasuya, M., Yu, T.-S., Vogt, K. E., Muratani, M., Ohnishi, T., Singh, S., Teixeira, C. M., Chérasse, Y., Naoi, T., Wang, S.-H., Nondhalee, P., Osman, B. A. H., Kaneko, N., … Sakaguchi, M. (2020). Sparse Activity of Hippocampal Adult-Born Neurons during REM Sleep Is Necessary for Memory Consolidation. *Neuron*, *107*(3), 552-565.e10. https://doi.org/10.1016/j.neuron.2020.05.008

Lerner, I., Lupkin, S. M., Tsai, A., Khawaja, A., & Gluck, M. A. (2021). Sleep to Remember, Sleep to Forget: Rapid Eye Movement Sleep Can Have Inverse Effects on Recall and Generalization of Fear Memories. *Neurobiology of Learning and Memory*, 107413. https://doi.org/10.1016/j.nlm.2021.107413

Li, Y., Aimone, J. B., Xu, X., Callaway, E. M., & Gage, F. H. (2012). Development of GABAergic inputs controls the contribution of maturing neurons to the adult hippocampal network. *Proceedings of the National Academy of Sciences*, *109*(11), 4290–4295.

Louie, K., & Wilson, M. A. (2001). Temporally Structured Replay of Awake Hippocampal Ensemble Activity during Rapid Eye Movement Sleep. *Neuron*, *29*(1), 145–156. https://doi.org/10.1016/S0896-6273(01)00186-6

Marín-Burgin, A., Mongiat, L. A., Pardi, M. B., & Schinder, A. F. (2012). Unique Processing During a Period of High Excitation/Inhibition Balance in Adult-Born Neurons. *Science (New York, N.y.)*, *335*(6073), 1238–1242. https://doi.org/10.1126/science.1214956

McDonnell, M. D., & Ward, L. M. (2011). The benefits of noise in neural systems: Bridging theory and experiment. *Nature Reviews Neuroscience*, *12*(7), 415–425. https://doi.org/10.1038/nrn3061

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

McKinney, W. & others. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *445*, 51–56.

Meltzer, L. A., Yabaluri, R., & Deisseroth, K. (2005). A role for circuit homeostasis in adult neurogenesis. *Trends Neurosci.*, *28*(12), 653–660. https://doi.org/10.1016/j.tins.2005.09.007

Ming, G., & Song, H. (2005). Adult Neurogenesis in the Mammalian Central Nervous System. *Annual Review of Neuroscience*, *28*(1), 223–250. https://doi.org/10.1146/annurev.neuro.28.051804.101459

Mongiat, L. A., Espósito, M. S., Lombardi, G., & Schinder, A. F. (2009). Reliable Activation of Immature Neurons in the Adult Hippocampus. *PLOS ONE*, *4*(4), e5320. https://doi.org/10.1371/journal.pone.0005320

Morcos, A. S., Barrett, D. G. T., Rabinowitz, N. C., & Botvinick, M. (2018). On the importance of single directions for generalization. *ArXiv:1803.06959 [Cs, Stat]*. http://arxiv.org/abs/1803.06959

O'Donnell, C., & Sejnowski, T. J. (2014). Selective Memory Generalization by Spatial Patterning of Protein Synthesis. *Neuron*, *82*(2), 398–412. https://doi.org/10.1016/j.neuron.2014.02.028

Oliphant, T. E. (2006). *A guide to NumPy* (Vol. 1). Trelgol Publishing USA.

Oomen, C. A., Hvoslef-Eide, M., Heath, C. J., Mar, A. C., Horner, A. E., Bussey, T. J., & Saksida, L. M. (2013). The touchscreen operant platform for testing working memory and pattern separation in rats and mice. *Nature Protocols*, *8*(10), 2006–2021. https://doi.org/10.1038/nprot.2013.124

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Rapp, P. R., & 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. https://doi.org/10.1073/pnas.93.18.9926

Richards, B. A., & Frankland, P. W. (2017). The Persistence and Transience of Memory. *Neuron*, *94*(6), 1071–1084. https://doi.org/10.1016/j.neuron.2017.04.037

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ryan, T. J., & Frankland, P. W. (2022). Forgetting as a form of adaptive engram cell plasticity. *Nature Reviews Neuroscience*, *23*(3), 173–186. https://doi.org/10.1038/s41583-021-00548-3

Sahay, A., Scobie, K. N., Hill, A. S., O'Carroll, C. M., Kheirbek, M. A., Burghardt, N. S., Fenton, A. A., Dranovsky, A., & Hen, R. (2011). Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature*, *472*(7344), 466–470. https://doi.org/10.1038/nature09817

Sahay, A., Wilson, D. A., & Hen, R. (2011). Pattern Separation: A Common Function for New Neurons in Hippocampus and Olfactory Bulb. *Neuron*, *70*(4), 582–588. https://doi.org/10.1016/j.neuron.2011.05.012

Santoro, A. (2013). Reassessing pattern separation in the dentate gyrus. *Frontiers in Behavioral Neuroscience*, *7*. https://doi.org/10.3389/fnbeh.2013.00096

Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. *Brain Mapping*, *3*, 501–506.

Schmidt-Hieber, C., Jones, P., & Bischofberger, J. (2004). Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. *Nature*, *429*(6988), 184–187. Scopus. https://doi.org/10.1038/nature02553

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems*, *25*. https://proceedings.neurips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. *15*, 1929–1958.

Sučević, J., & Schapiro, A. C. (2022). *A neural network model of hippocampal contributions to category learning* (p. 2022.01.12.476051). bioRxiv. https://doi.org/10.1101/2022.01.12.476051

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Contributors, S. 1. 0. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*.

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., Ruiter, J. de, Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., … Qalieh, A. (2017). *mwaskom/seaborn: V0.8.1 (September 2017)* (v0.8.1) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.883859

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Wiskott, L., Rasch, M. J., & Kempermann, G. (2006). *A Functional Hypothesis for Adult Hippocampal Neurogenesis: Avoidance of Catastrophic Interference in the Dentate Gyrus*. *343*, 329–343. https://doi.org/10.1002/hipo.20167

Yasuda, M., Johnson-Venkatesh, E. M., Zhang, H., Parent, J. M., Sutton, M. A., & Umemori, H. (2011). Multiple Forms of Activity-Dependent Competition Refine Hippocampal Circuits In Vivo. *Neuron*, *70*(6), 1128–1142. https://doi.org/10.1016/j.neuron.2011.04.027

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

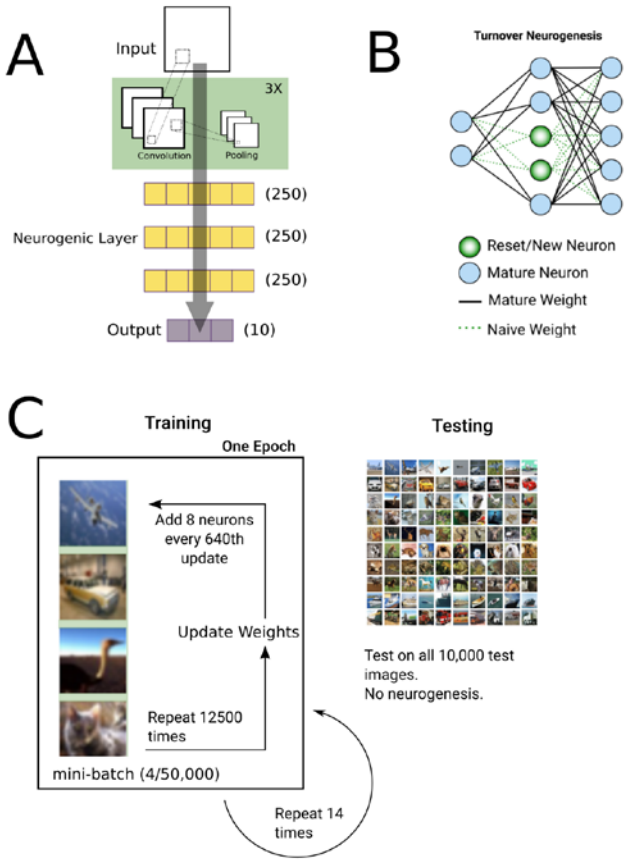

***Figure 1*. Implementing neurogenesis in convolutional neural networks. (A)** Illustration of the CNN used in these experiments. **(B)** A schematic illustrating how replacement/turnover neurogenesis was implemented. **(C)** Illustration of the training and testing process with neurogenesis.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

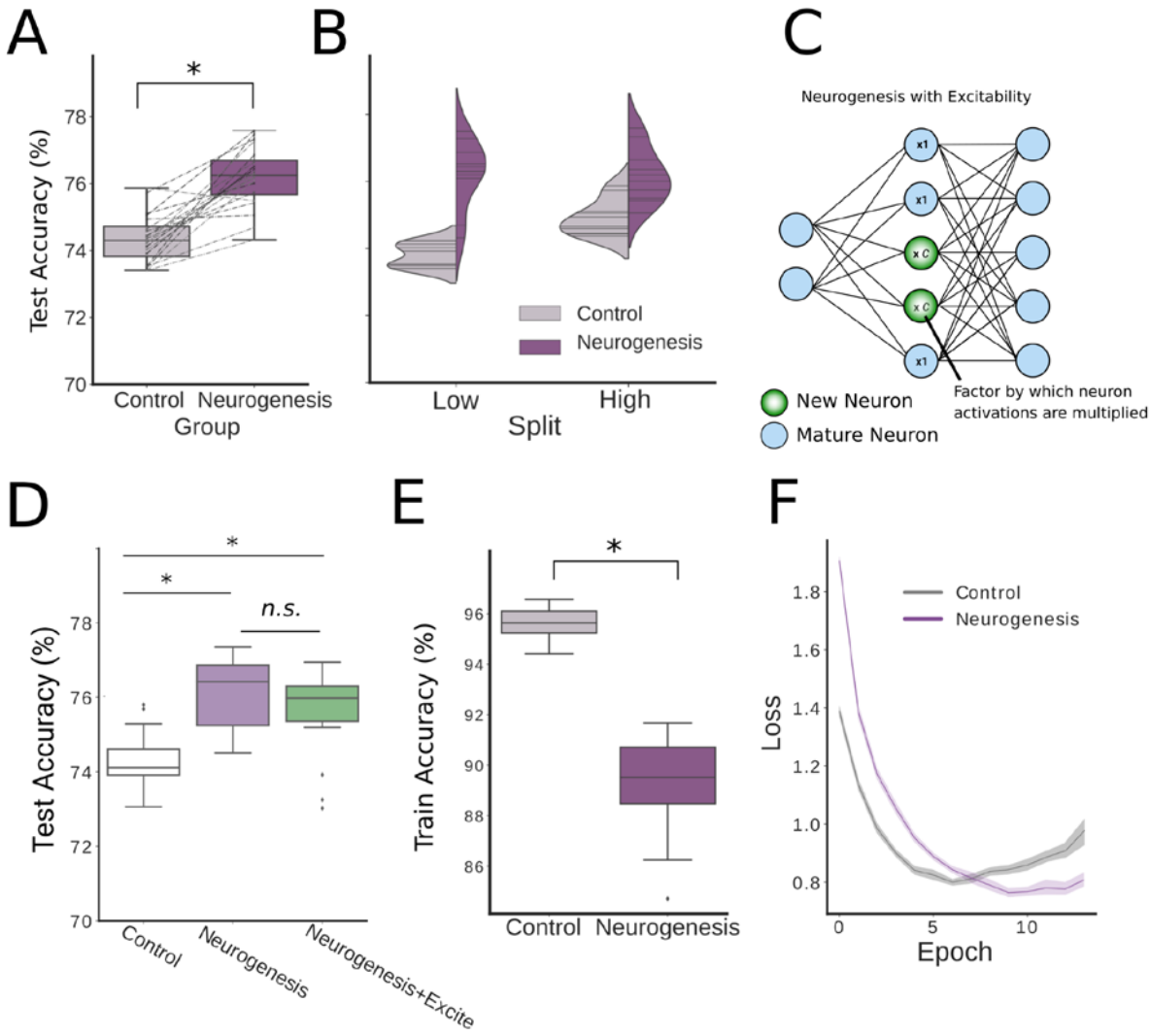

Figure 2. **Neurogenesis improves generalization in CNNs. (A)** Box plot of test accuracy of control and neurogenesis networks after training; t-test: $t_{19} = 7.00$, $p = 1.1\times10^{-6}$. **(B)** Violin plots of the distribution of scores for the lowest scoring (left) and highest scoring (right) halves from each group: control and neurogenesis, from (A). **(C)** An illustration of how we implemented enhanced excitability of new neurons in a neurogenic neural network by multiplying the activations of a new neuron by an excitability factor, c. **(D)** Boxplot of test accuracies of control, neurogenic and neurogenic + excitability CNNs; ANOVA: $F_{2,57} = 19.01$, $p = 4.7\times10^{-7}$; Tukey's HSD: Neurogenesis, Neurogenesis Excite > Control, $p < 0.01$. **(E)** Box plot of the training accuracy of the control and neurogenesis groups at the end of training; t-test: $t_{19} = 4.94$, $p = 9.0 \times 10^{-5}$. **(F)** Plot of validation loss across training. *represents *p-value* below 0.01.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

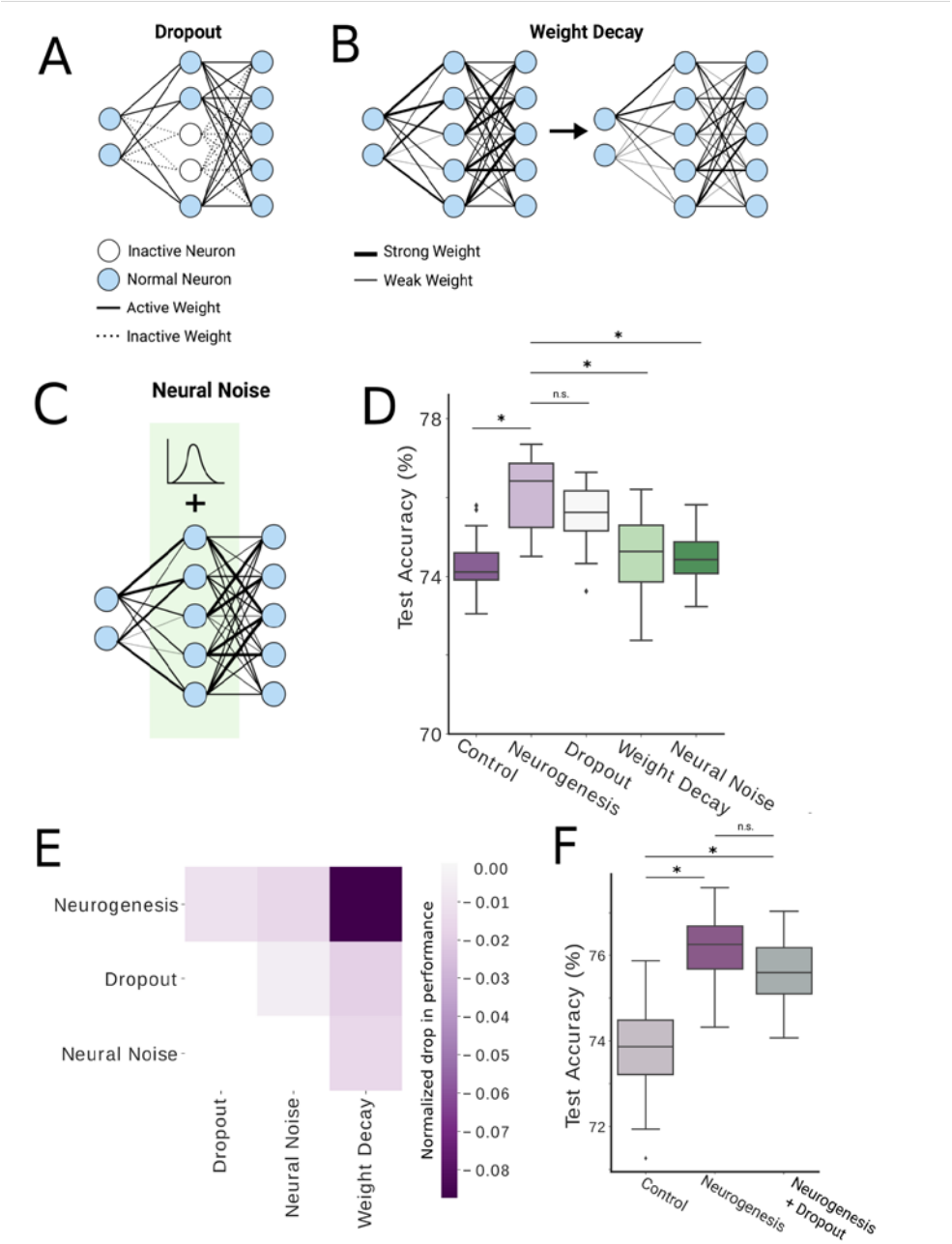

***Figure 3*. Illustration of regularization methods. (A)** Dropout: a random subset of neurons and their weights are inactivated during a given forward pass. **(B)** Weight decay consists of adding a small penalty to the loss function that penalizes large weights, thus resulting in an overall decay of larger weights. **(C)** Neural noise: Gaussian noise is added to the activations of a layer. **(D)** Box plot of the test accuracy of neurogenesis compared to other regularization methods; ANOVA: $F_{4, 95} = 16.54$, $p = 2.5\times10^{-10}$; Tukey's HSD: neurogenesis vs. control $p<0.01$, neurogenesis vs. dropout $p>0.01$, neurogenesis vs. weight decay $p<0.01$, neurogenesis vs. neural noise $p<0.01$. **(E)** Heatmap of z-scores of test performance in networks with combined regularization methods relative to neurogenesis-only networks. **(F)** Plot of neurogenesis and dropout combined using lower parameter values of dropout (0.1), and neurogenesis (turnover every 1000 updates); ANOVA: $F_{2,57} = 65.69$, $p = 1.6\times10^{-15}$; Tukey's HSD: control vs. neurogenesis $p <$ 0.01, control vs. neurogenesis+dropout $p < 0.01$, neurogenesis vs. neurogenesis + dropout $p > 0.01$. * represents *p-value* below 0.01.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

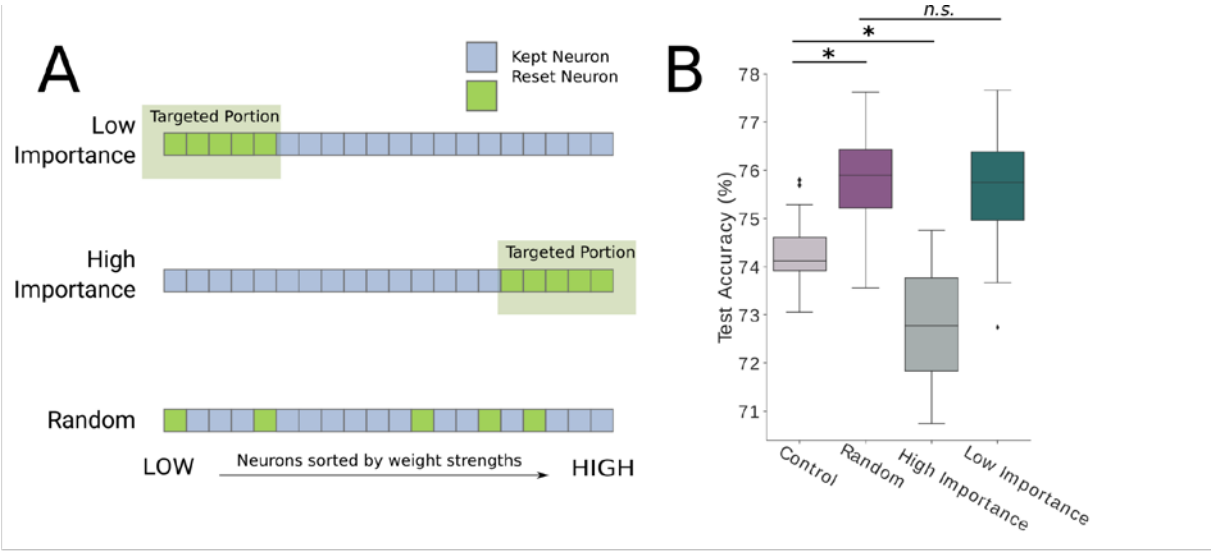

Figure 4. Targeted neurogenesis does not change network performance. (A) An illustration of targeted neurogenesis in a hidden layer. The neurons are ranked in order of input weight strengths, and a proportion of the lowest-ranked (low importance) or highest-ranked (high importance) neurons are targeted for neurogenesis (i.e., having their weights reset), or there is no targeting (random). **(B)** A boxplot of model test accuracy in control, random neurogenesis, targeted neurogenesis of high importance neurons (positive control), and targeted neurogenesis of low importance neurons; ANOVA: $F_{3, 96} = 46.30$ $p = 1.3\times10^{-18}$; Tukey's HSD: (Control vs. Random) $p<0.01$, (Control vs. High Importance) $p<0.01$ (Random vs. Low Importance) $p>0.01$. * represents *p-value* below 0.01.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487582; this version posted April 10, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

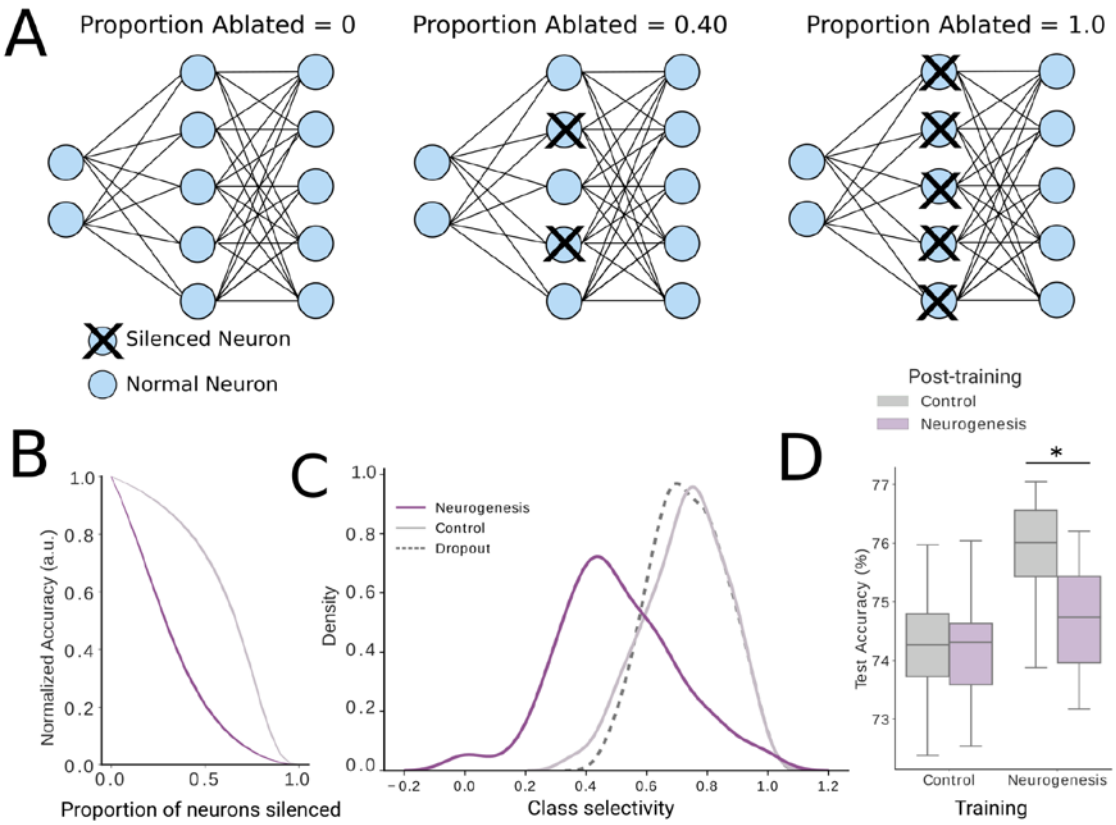

Figure 5. Networks with neurogenesis are less robust to ablation. (A) Illustration of ablation experiments. **(B)** Plot of the mean normalized accuracy across 20 repeats as progressively more neurons are ablated from the network. **(C)** Density plot of the class selectivity of neurons in the second hidden layer in control, neurogenic and dropout networks. **(D)** Boxplot of test accuracy of networks that were trained with and without neurogenesis, and then tested with and without new neurons added post-training; Repeated-measures ANOVA: Training × Post-training interaction, $F_{1,19} = 58.68$, $p < 0.01$; Tukey's HSD: (Neurogenesis/Control vs Neurogenesis/Neurogenesis) $p < 0.01$. * represents *p-value* below 0.01.