

Bridging Neuroscience and AI: Environmental Enrichment as a Model for Forward Knowledge Transfer.

Rajat Saxena¹, Bruce L. McNaughton^{1,2}

1 Department of Neurobiology and Behavior, University of California, Irvine, Irvine, CA 92697, USA

2 Canadian Centre for Behavioural Neuroscience, University of Lethbridge, Lethbridge, AB, T1K 3M4 Canada

*Correspondence: brucemcn@uci.edu, rajat.saxena@uci.edu

Acknowledgments

The authors thank Jeffery Krichmar, Artur Luczak, Aaron Bornstein, Robert Bain, Janina Ferbinteanu, Justin Shobe, Wing Ning, and Aditi Bishnoi for their insightful comments. We also thank Aditi Bishnoi for helping us with graphic illustrations. This work was supported by NIH R01 NS121764 (BLM), NIH R01 MH125557 (BLM) and NIH RF1 NS132041 (BLM).

Keywords: Continual learning, forward transfer, environmental enrichment, learning, memory, complementary learning systems.

Abstract

Continual learning (CL) refers to an agent's capability to learn from a continuous stream of data and transfer knowledge without forgetting old information. One crucial aspect of CL is forward transfer, i.e., improved and faster learning on a new task by leveraging information from prior knowledge. While this ability comes naturally to biological brains, it poses a significant challenge for artificial intelligence (AI). Here, we suggest that environmental enrichment (EE) can be used as a biological model for studying forward transfer, inspiring human-like AI development. EE refers to animal studies that enhance cognitive, social, motor, and sensory stimulation and is a model for what, in humans, is referred to as 'cognitive reserve'. Enriched animals show significant improvement in learning speed and performance on new tasks, typically exhibiting forward transfer. We explore anatomical, molecular, and neuronal changes post-EE and discuss how artificial neural networks (ANNs) can be used to predict neural computation changes after enriched experiences. Finally, we provide a synergistic way of combining neuroscience and AI research that paves the path toward developing AI capable of rapid and efficient new task learning.

Introduction

Forward transfer is the ability to utilize previously acquired knowledge to learn new tasks rapidly and efficiently ^{1,2}. For example, imagine attending your first lecture on deep learning. Rather than nodding off, you're engaged, following the math, and can quickly code a simple linear deep learning model. After the class, your friend comments on how last semester's math course on linear algebra and matrices helped in today's lecture. And it suddenly hits you why the lecture felt easy. You were able to utilize and transfer concepts from math classes to the lecture on deep learning. Previous work has indicated that students with superior mathematics training perform better in computer science and logical reasoning, due to the ability to transfer abstraction skills from one subject matter to another (Fig. 1) ^{3,4}. The mammalian brain is capable of continual learning (CL): learning from a stream of data or experiences continuously, adapting to context changes, and realizing actions throughout its lifetime without forgetting previously acquired knowledge (catastrophic forgetting) ⁵. Although seemingly trivial for humans and animals, CL remains one of the longstanding issues in AI ^{6,7}.

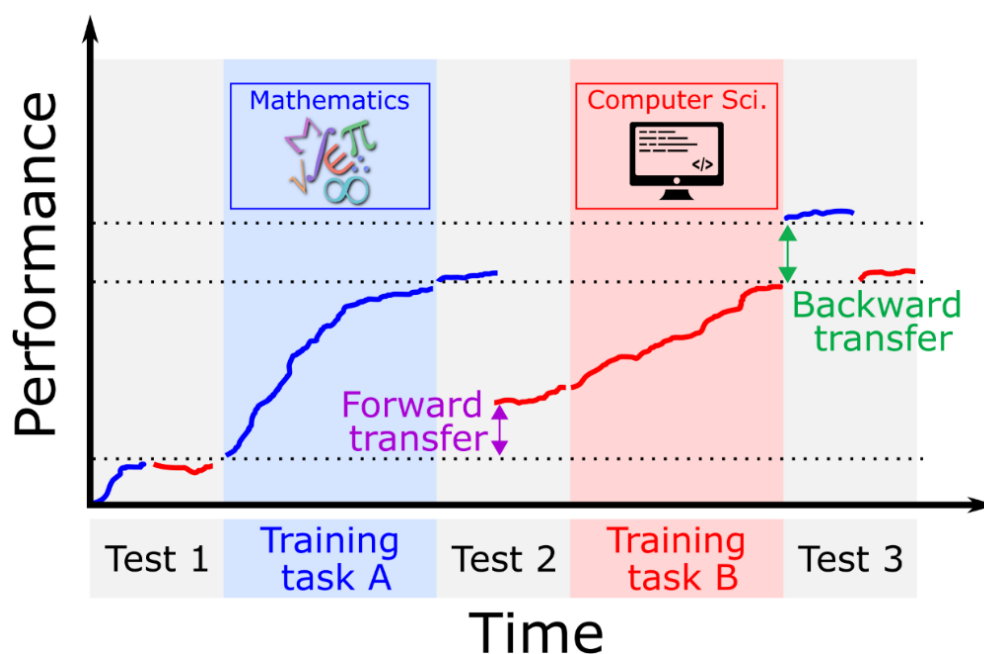


Figure 1: A schematic to demonstrate transfer learning. An agent was trained sequentially on two tasks: task A (mathematics) and task B (computer science) and evaluated on both tasks at three different times (Test 1-3, gray block), with Test 1 as the baseline. Individual task performance is expected to improve during the respective training blocks (Training tasks A and B). Due to forward transfer (purple arrow), the performance on task B (computer science, red curve) is also expected to improve in Test 2 after training on task A (mathematics). Similarly, performance on task A is

likely to improve in Test 3 after training on task B due to backward knowledge transfer (green arrow). This article concentrates solely on forward transfer (purple arrow).

The interconnected history of AI and neuroscience, particularly cognitive neuroscience, has spurred breakthroughs in both fields. While neuroscience has continued to inform the development of brain-like algorithms (although subtly), the road to success has been challenging⁷⁻¹⁰. Despite surpassing expert human performance in challenging tasks such as playing Atari¹¹, Go¹², etc.; current ANNs struggle to capture multiple aspects of human intelligence. Most of the impressive results obtained have been in a static learning setting with a single objective, such as image classification. A CL framework is different since it involves a potentially infinite stream of data stemming from different input domains (visual, audio, etc.) with unclear task boundaries and changing task demands^{6,7}.

How do human brains achieve this feat of CL? Not all your friends would have been good at computer science on the first attempt. Can an individual's ability to learn a new task be influenced by their prior experiences and environment? We draw inspiration from neuroscience research examining the effects of environmental enrichment (EE), especially in rodents. EE refers to environmental modifications to provide better cognitive, social, motor, and sensory stimulation and has shown promising results in improving learning and memory¹³⁻¹⁵. Enrichment was first studied by Donald Hebb¹⁶, who noticed that the rats raised in his home performed better than laboratory-raised animals in problem-solving tasks. Subsequent studies have replicated these results across species, further examining the impact of EE on learning, memory retrieval, multisensory integration, etc. EE also alters different brain mechanisms, leading to increased neurogenesis, population sparsity, long-term potentiation (LTP), etc., all of which are hypothesized to underlie efficient learning^{13-15,17-19}. These studies have helped establish EE as an effective strategy for improving learning, performance, and memory retrieval - desired properties of a CL system exhibiting forward transfer. This article aims to demonstrate how EE serves as fertile ground for studying forward transfer in biological neural networks.

Amidst growing interest in developing ANNs to rival human intelligence, we propose to use EE as an inspiration for studying forward transfer in biological agents. We begin by introducing theories underlying semantic learning and the rapid acquisition of new knowledge in the brain. Then, we discuss EE and its effects on behavior and neural mechanisms. Next, we consider key insights from ANNs exhibiting forward transfer and few-shot learning (learning using very few examples). Finally, we conclude by laying down a path to combine results from EE studies and forward transfer CL in artificial agents to drive further improvement of ANNs and propose hypotheses to be tested in the brain in future experiments.

Faster and efficient learning using schemas and compositionality

The mammalian brain has an incredible capacity to acquire new information rapidly without forgetting pre-existing knowledge. What is the mechanism by which our brain accomplishes this? One prominent theory is the complementary learning systems theory (CLST) which states that the brain relies on complementary learning systems: hippocampus (HC) for rapid encoding of new information, while the neocortex (NC) is thought to be involved in slow learning, gradually forming structured knowledge or “schemas” from new information ^{20,21}. During offline periods, such as awake rest or sleep, HC triggers the replay of new information in NC, and NC spontaneously retrieves and interleaves existing knowledge. This interleaved replay might allow for a gradual creation of context-independent category representations (schemas) in an NC distributed circuit (Fig. 2a). A set of such schemas reflects our generally acquired intelligence or knowledge of the world, representing a mental structure of concepts and the relation between them and can influence the new item encoding, consolidation, and retrieval ^{22,23}. It was proposed that we base our learning of new information on already existing, consolidated schemas, such that the degree of consistency of the new information with the existing schemas affects the rate of acquisition ^{22–24}. Based on this, the proposed slow learning role of NC was updated, with empirical work in both humans and animals, demonstrating that NC can act as a fast-learning system (just like HC) if new information is highly consistent with previously created schemas ^{23,25}. Similarly, in ANN simulations, there was an exponential speedup in learning new items with increasing consistency ^{26,27}. The extent of prior knowledge can differ significantly across the lifespan of an individual, starting with innate knowledge shaped primarily by evolution in childhood and transitioning to a mix of innate and learned structured knowledge driven by experience in adulthood ²⁸. There might be differences in the learning strategy of new items and learning speed between naïve vs. knowledge-rich brains, with the latter exhibiting faster learning (forward transfer) and reasoning-like strategy rather than a trial-and-error approach by the former. The accelerated learning in knowledge-rich agents could be due to schemas that help restrict the extent of internal representation parameter space that must be explored while learning a new task to the most relevant, task-specific, low-dimensional subspace, representing shared features and principles across tasks ^{29,30}.

Recognizing shared features across tasks allows us to leverage prior knowledge and combine it with new information to learn more efficiently. Continuing with the example from Fig. 1, a person with a strong background in mathematics might be able to learn deep learning concepts easily, by breaking down the concepts and recognizing shared features (e.g., matrix multiplication, probability, etc.) between the two subjects (Fig. 2b). This ability to construct complex concepts and representations in novel circumstances by combining smaller, more basic reusable elements (priors) is referred to as “compositionality” ^{31–33}. This can allow a learning agent to generate an essentially infinite number of meaningful concepts from a finite set of basic

building blocks, helping with generalizations and sample-efficient learning ³⁴. Although the concept of compositionality was initially used in language processing research, it is a general property of cognition that allows fast, flexible, generalizable action and reasoning. Despite their differences, compositionality and schemas are closely related. Schemas based on prior knowledge provide basic building blocks to generate novel and complex concepts using compositional processes. At the same time, compositionality also shapes the development and structure of schemas. As we encounter new information, we may combine and recombine elements from existing schemas to create new, more complex schemas. For example, we may combine our schemas for linear algebra, matrices, probability, etc. principles to create a new schema for deep learning (Fig. 2b). The construction of new abstractions or schemas often takes advantage of the reusable and shareable schemas that are formed as a product of learning over time (and evolution) ^{35,36}. In conclusion, a deeper understanding of neurobiological processes underlying schema formation and compositionality might offer insights into the speedup of learning, a topic of interest for both neuroscience and AI.

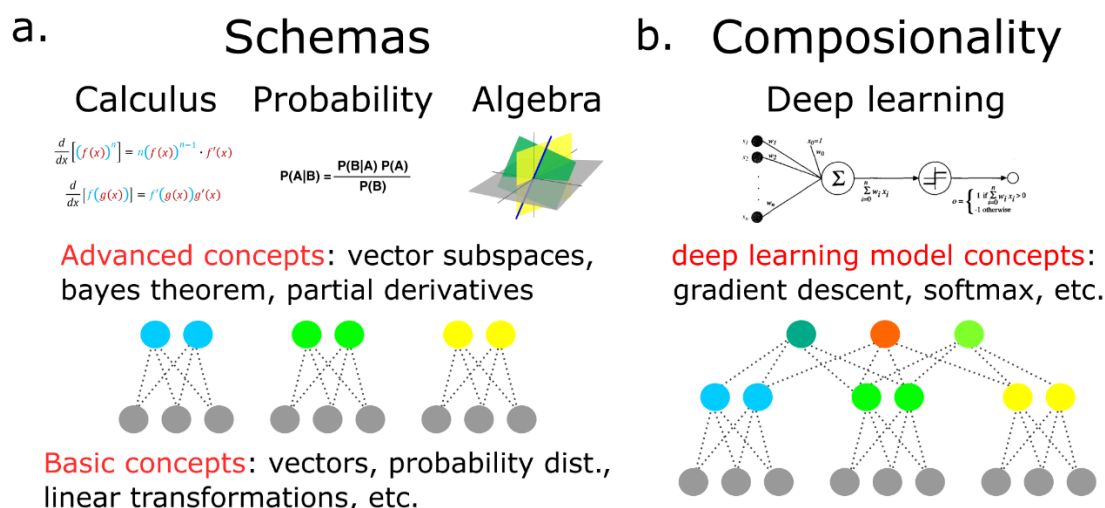


Figure 2: a) Schemas for different math topics, including probability, linear algebra, and matrices. At the bottom are basic schemas common across topics, such as vectors, linear transformation, probability distributions, etc., that are learned first. As the individual gains knowledge of each subject matter, they develop mental abstractions (schemas) for advanced concepts (e.g., vector subspaces, partial derivatives, Bayesian probability) that are specific (and higher-order) to each topic. b) An individual with conceptual knowledge of concepts from linear algebra, probability theory, and calculus can combine them (compositionality) and learn deep learning models much faster than someone with no knowledge of those concepts. For example, stochastic gradient descent in deep learning models uses concepts of partial derivatives, matrix-vector multiplication, etc., from different math topics.

Environmental enrichment enhances learning, memory, and underlying brain mechanisms

The environment around us has a significant impact on our cognitive abilities. EE is most often conceptualized as a method to improve biological and cognitive functions meeting species-specific needs, resulting from modifications to the environment, mental stimulation, social interactions, etc. ^{13,37,38}. Broadly, four types of EE paradigms are frequently used: physical, sensory, cognitive, and social, with each providing an opportunity to improve problem-solving skills, promote exercise, and reduce stress and anxiety or a combination of these (Fig. 3a) ^{14,39–41}. *Note that this article focuses mainly on cognitive EE rather than physical EE.* Apart from the type of EE paradigm, there are other vital factors of the EE paradigm that could contribute to its effectiveness. Out of these factors, age at EE onset ^{13,42}, timing (light vs. dark phase) ⁴³, duration ⁴⁴, and continuous vs. intermittent ⁴⁵ are of utmost importance ³⁸. For example, enrichment spanning weeks at a younger age might lead to more considerable gains than shorter exposure ⁴². Many studies have tried combining different types of EE paradigms while accounting for some of the factors listed above to yield an overall more effective outcome (Fig. 3a) ^{38,39}. A recent study by Gattas et al. ¹⁵ used a novel EE protocol ('obstacle course') that enabled controlled enrichment delivery, accounting for physical exercise (with a control exercise track) and degree of rodent engagement with enrichment materials. The enrichment track obstacle course yielded improved and long-lasting performance on tasks involving multisensory integration, categorization, etc., as compared to standard enriched housing or increased physical exercise-only enrichment paradigm (Fig. 3a).

Extensive work has demonstrated that EE significantly improves spatial learning, spatial memory tasks, recognition memory, increased exploratory behavior, and task learning rate ^{14,17,40,41,45–50} (Fig. 3b, left). One possibility underlying improved behavior after EE could be the emergence of multiple schemas due to exposure to novel complex experiences, which then lead to functionally efficient connectivity, better generalization, and accelerated learning in novel situations ^{22–25}. Most EE studies have focused on rodent models, but similar effects have been demonstrated in humans. Physical EE (e.g., resistance training and aerobic exercise) ⁵¹, cognitive EE (e.g., number of years of education, playing strategy games) ^{19,52}, and social EE (e.g., higher social network density) ⁵³, have all been shown to enhance learning and memory. 'Cognitive reserve' refers to the brain's ability to maintain normal cognitive function in the face of aging or neurological diseases. Engaging in EE practices, such as education, physical exercise, social interactions, etc. providing the brain with a rich and diverse set of experiences has been hypothesized to contribute to the development and maintenance of cognitive reserve ^{54–57}. Altogether, EE is a promising approach that highlights the potential benefits of stimulating and varied experiences on new task learning and cognitive adaptability in both humans and animals.

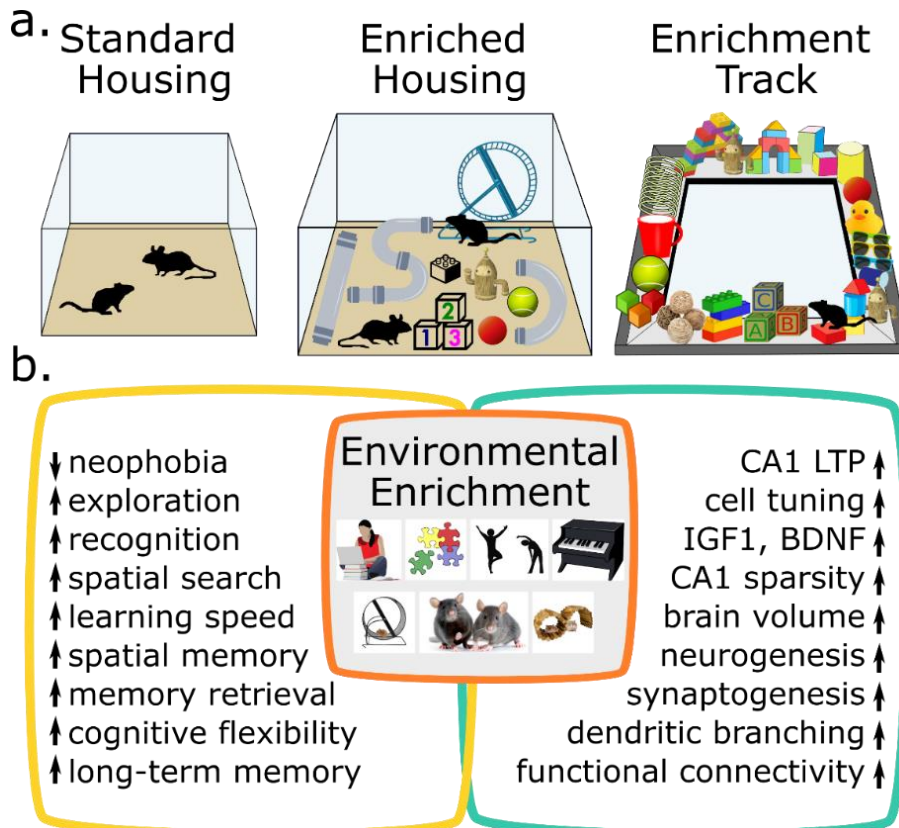


Figure 3: a) Different types of EE strategies. Standard housing allows for no to minimal enrichment opportunities (left), enriched housing involves exposing animals to a larger arena with many objects (middle), and the enrichment track provides controlled delivery of physical and cognitive stimulation (right). b) Effects of EE (physical, cognitive, social, sensory) on behavior (left) and underlying neural dynamics (right). EE typically improves spatial memory, learning rate, exploratory behavior, recognition memory, and multisensory integration (left). EE alters mechanisms underlying learning and memory at all levels, from molecular to brain-circuit wide. Most studies report changes in increased CA1 LTP, adult neurogenesis, synaptogenesis, dendritic branching, etc. (right). Note that only a subset of effects, observed with different enrichment parameters, such as duration, types, age of initiation, etc. are illustrated.

Along with the improvement in behaviors, EE also alters anatomical, molecular, and physiological functions (Fig. 3b, right). EE promotes synaptogenesis, and adult hippocampal neurogenesis^{13,58,59}, and enhances LTP in HC CA1^{60,61}. In addition, it increased spine count in basal dendrites of cortical layer II/III and apical tuft dendrites of layer V pyramidal neurons^{62,63}, and dendritic volume and branching⁶⁴. Such cellular and physiological changes are associated with altered expressions of genes involved in plasticity, including increased levels of brain-derived neurotrophic factor (BDNF), N-methyl-D-aspartate (NMDA) receptor subunit 1 (NR1)⁶⁵, Insulin-like growth factor 1 (IGF-1)⁶⁶ and postsynaptic density protein 95 (PSD95)⁴⁸. Multiple

studies have also reported an increase in levels of acetylcholine, noradrenaline, and serotonin (5-HT), all indicative of triggering different learning and adaptive strategies in enriched animals^{67,68}. EE also sharpens single-unit receptive fields, increases the strength of sensory cortices responses, and the selectivity, sensitivity, and efficiency of population coding (sparser coding)^{69–71}. At the network level, it leads to increased macroscopic functional connectivity efficiency (higher node degree, edge connectivity) in the HC, visual, motor, retrosplenial, and cingulate cortices, which could underlie improved multisensory integration and the emergence of supramodal representations^{43,72,73}. The changes discussed here may allow EE animals to adapt to new experiences and switch between different tasks, leading to improved learning, memory (forward transfer), and cognitive flexibility (Fig. 3b, right).

It is important to state here that more comprehensive experiments are needed to ascertain whether all the changes listed above contribute to improved performance and faster learning of new tasks. The results listed above were observed in EE paradigms with different designs and parameters (total duration, age of initiation, type of EE, etc.). Future studies should carefully parse out the contribution of all changes (adult neurogenesis, synaptogenesis, more integrated cortical networks, increased population sparsity, etc.) observed so far and different types of EE paradigms on schema formation to determine if they all contribute equally or if one factor outweighs the others.

Few-shot learning and forward transfer in artificial neural networks

EE animals clearly show aspects of forward transfer: improved performance and learning speed on new tasks^{40,41}. But can ANNs exhibit similar behavior, and if they can, what factors determine those results? Take an ANN trained on linear algebra and matrices problems and teach it principal component analysis and deep learning until it becomes an expert in all. We can then ask what concepts were learned first, which were easiest to transfer, which led to maximum confusion, etc. Research on ANNs has recently shifted focus from single supervised tasks to the CL domain^{7,74,75}, with increased interest in joint exploration of forward transfer and catastrophic forgetting. A recent set of studies showed that ANNs with highly orthogonal prior knowledge representation show a gradient of forgetting on new task learning, with highly similar and dissimilar old tasks showing maximum and minimum forgetting, respectively^{27,76,77}. For example, a network initially trained to recognize two vehicles (ship, truck) and two animals (deer, dog) showed more forgetting of other vehicle classes like planes and cars because they were similar to the original vehicle classes (Fig. 4a)⁷⁶. To overcome the forgetting of similar classes, the model was trained to jointly replay a higher proportion of similar task exemplars (ship, truck) and novel tasks (car, plane). By exploiting the hierarchical organization of existing knowledge, this joint replay algorithm: similarity-weighted interleaved learning (SWIL), yielded rapid learning of new

tasks while overcoming the catastrophic forgetting problem using substantially less data ^{26,27}. Additionally, the speed of learning novel items increased proportionally to the number of non-overlapping old items. SWIL also helped improve the ability of ANNs to learn new tasks from limited amounts of data and to generalize to new tasks. These results demonstrate the importance of prior knowledge in driving learning accuracy and speed, implying more forward transfer in an agent with multiple orthogonal representations ^{27,76,77}, matching experimental work on schemas in the brain ^{22,23}. In the same vein, ANNs exhibit less forgetting of previously acquired knowledge and increased forward transfer as they learn multiple tasks, i.e., learning the n^{th} thing gets easier than learning the first thing ^{1,78}. For example, learning about principal component analysis (a special case of singular value decomposition) is easier than learning about singular value decomposition for the first time. Also, relearning a previously learned task is faster than learning it for the first time (relearning savings), another characteristic of human learning and demonstrated in EE animals as well ¹⁷.

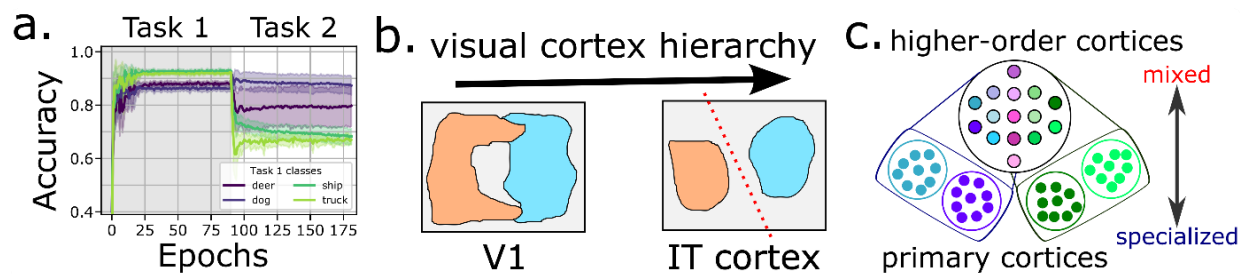


Figure 4: a) Performance of ANN trained on two tasks simultaneously. Task 1 involved training to learn two vehicles (ship (green), truck (lime green)) and two animals (deer (indigo), dog (blue)) classes. Training the same network on Task 2 (car, plane class) leads to more forgetting of similar old classes (ship, truck). Figure modified from ⁷⁶. b) Increasing dimensionality across visual cortex hierarchy. In V1, representations for two classes (orange and sky blue) are intermixed and low-dimensional. The exact representations become high-dimensional and linearly separable in the Inferotemporal (IT) cortex. c) Bottom: representation expected in primary-sensory cortices, showing functionally specialized representations (priors) that can be reused across tasks. Top: higher-order brain regions show high-dimensional mixed selectivity abstract representations that allow for better coding, reliability, and readout by a linear classifier. Mixed-selectivity coding has been shown to provide better flexibility and linear decoding of task variables, allowing for faster and more efficient new item learning ⁷⁹.

What drives this rapid learning? Is there something different about the representations learned by the ANNs? A recent study carefully examined the relationship between the geometry of learned representations and learning speed and found that schema-like learning (prototype learning) led to the best forward transfer in ANNs on learning novel naturalistic concepts, much

better compared to exemplar learning⁸⁰. The study also reported a consistent increase in few-shot learning capability along the layers of the network, with the improvement driven by a sharp rise in the separability of class representations in the last few layers, also reported in the visual cortex⁸¹ (Fig. 4b). Across the visual cortex (and other cortices) hierarchy, cell ensembles in individual brain regions exhibit a variety of coding schemes, ranging from functionally specialized to mixed selectivity^{79,82}. This raises the question of whether there is a difference in the forward transfer capability between the two schemes and when one should be favored over the other. An elegant study attempted to address this question by training recurrent neural networks on N=20 orthogonal tasks in an interleaved design⁸³. The outcome was the emergence of a cluster of units that demonstrated functional specialization and could be combined to solve novel tasks. On the other hand, the same network developed mixed selectivity after learning multiple tasks sequentially. Mixed selectivity codes have been shown to make fewer errors than highly specific codes, providing reliable decoding in sensory and high-order cortices⁸⁴. These results indicate that pre-training ANNs on large datasets spanning a mixture of domains (akin to exposure to many experiences and tasks) might significantly improve forward transfer^{80,85,86}. We propose that the higher fraction of specialized units is similar to neuronal ensembles in primary sensory cortices (more evolution-driven), allowing for reuse and sharing across tasks (compositionality), helping in tasks using a single variable. Mixed-selectivity representations, on the other hand, align well with population activity in higher-order brain regions (such as the prefrontal cortex and HC), allowing not only for compositionality and forward transfer but also simple readout by a linear classifier (Fig. 4c).

However, coding for multiple variables always raises the discrimination-generalization tradeoff concern. A modeling study reported that randomly connected neurons with a population sparsity of ~ 0.1 (10% of total neurons are active to each input stimuli) are optimal for mixed selectivity representations, as they can balance discrimination and generalization⁸⁷. The randomly connected neurons increase the dimensionality of representations, similar to the dentate gyrus (DG) in the HC, allowing for better pattern separation⁸⁸⁻⁹⁰. These results hint at what might be expected in representations for a knowledge-rich vs. naïve brain. We think EE pushes the neural representations closer to a regime where generalization and discrimination (orthogonal coding) capabilities are balanced. These balanced representations might enable EE animals to learn different tasks quickly (utilizing sparse coding) and generalize across contexts. However, EE animals might show more non-linear mixed selectivity in higher-order brain regions, as it allows an efficient encoding of a large repertoire of experiences. In this section, we established that pre-training ANNs on large multi-modal datasets would help them learn generalizable representations, improving forward transfer. We pointed out that schema-like learning, exploiting the existing structured knowledge, achieves much better forward transfer compared to exemplar learning. The findings discussed here provide valuable insights into the internal representations and learning dynamics of ANNs trained to achieve rapid learning.

Enhancing forward transfer: the neuroscience-AI nexus loop

So far, we have looked at how EE enhances learning speed and performance on novel tasks in brains, along with results from forward transfer and few-shot learning in ANNs. Now, we want to bring the two together and discuss different mechanisms where both fields can interact and benefit from each other.

1. Neurogenesis:

New neurons are produced in the brain from early development to adulthood⁹¹, most notably in DG and olfactory bulb⁹². Adult-born immature DG cells have been theorized to promote pattern separation (remapping)⁹³ and are essential for remote memory reconsolidation⁹⁴. Reduced adult neurogenesis in DG extends hippocampal dependence on memory⁹⁵, indicating its role in faster integration of new information into the existing knowledge. The rate of neurogenesis is higher in EE animals, presumably underlying faster discrimination and schema formation^{58,59}. Future work should focus on developing ANNs with the rate of neurogenesis altered by age and task demands. For example, lifelong learning ANNs can be trained with a larger rate of neurogenesis early on in training (similar to childhood), which can be reduced after initial deployment or have a higher neurogenesis rate throughout, identical to EE animals. Also, the higher rates could be hidden layer-specific, with the last layers exhibiting more neurogenesis than earlier layers (similar to DG) (Fig. 5a)⁹⁶. Introducing new neurons in ANNs at different training time points could reveal its impact on categorical discrimination^{97,98}.

2. Learning rule, cost function, and modular architecture:

Most deep-learning approaches use the same learning rule and cost functions across all layers and nodes. On the other hand, learning rules and cost functions probably differ across brain regions (primary sensory vs. association cortex), and cell types, and may even change over development⁹⁹. Also, different brain regions show different levels of plasticity and representation coding schemes, from highly specialized to mixed selectivity modules, each offering different computational benefits^{79,82}. Internal representation properties of EE vs. naïve animals in different phases of learning might provide hints about the biological learning rules. We envision future ANNs to have a highly modular architecture, with each module having its own set of learning rules and cost functions (different levels of plasticity, sparsity, neuromodulators, etc.), which are configurable based on the relevance of the module to the current task (Fig. 5b)^{100,101}. Based on few-shot learning ANNs (previous section), we predict that EE animals will show a more significant extent of non-linear mixed selectivity in higher-order brain regions such as the prefrontal cortex and HC than naïve animals^{9,79}. Also, highly trained ANNs (knowledge-rich) tend

to show a good transfer of both coarse and fine discrimination abilities, which might be true for EE animals and can be tested ²⁸.

3. Replay and functional connectivity:

An interleaved replay of new and old memories during sleep has been hypothesized to improve memory consolidation and categorical knowledge ^{20,102}. EE animals show enhanced functional connectivity across brain regions during sleep ^{43,74}. A crucial difference between biological and artificial network replay is that replay in ANNs is mainly performed as a single process, whereas biological replay happens across cascaded memory systems, potentially driving changes in synaptic weights and rewiring across regions ^{103–107}. Future studies should add replay with different units/layers of the network involved across epochs and time (Fig. 5c). Incorporating biological network properties like sparse recurrent connectivity ¹⁰⁸ and small-worldness (dense local clustering, short path length between distant pairs of nodes) ¹⁰⁹, scale-freeness (connectivity degree distribution follows power law) ¹¹⁰, etc. might yield interesting results. Replay in ANNs could be used to understand how item representations get orthogonalized. Do coarser features separate first, followed by finer features? ⁸. ANNs show larger distances between individual class representations (deeper hierarchical clustering) in the hidden layers late in training compared to the start ²⁰. Based on this, we predict a much more diverse exploration of network activity subspace during replay in EE animals. Also, since the distance increase is larger in the last few layers, we expect the same to be true for higher-order cortices.

4. Neuromodulation triggered uncertainty:

Neuromodulators such as acetylcholine, dopamine, and noradrenaline are triggered by uncertainty and unexpected rewards ^{101,111,112}. Previous research has hypothesized that neuromodulation assists in adapting to new experiences by initiating various forms of learning based on novelty: unknown category (create new schema) or novel exemplar in a familiar category (consolidate into an existing schema) ^{113,114}. Multiple studies have used neuromodulator-triggered uncertainty in ANNs to overcome catastrophic forgetting and rapid adaptation to tasks ^{115,116}. These attempts could significantly benefit from insights from experiments exposing EE and control animals to multiple tasks with diverse types of novelty. For example, for a new item, the network might show increased pattern separation and a larger extent of new item replay initially followed by an increased replay of previously learned knowledge ^{117,118}. The difference in changes in the neuromodulatory tone and its temporal dynamics across tasks in EE animals ^{67,68} can provide insights into novel learning rules for ANNs (Fig. 5d). ANNs can also be used to model how different uncertainty signals come into play and work together. This can guide future biological experimental designs to see if similar mechanisms are employed in brains.

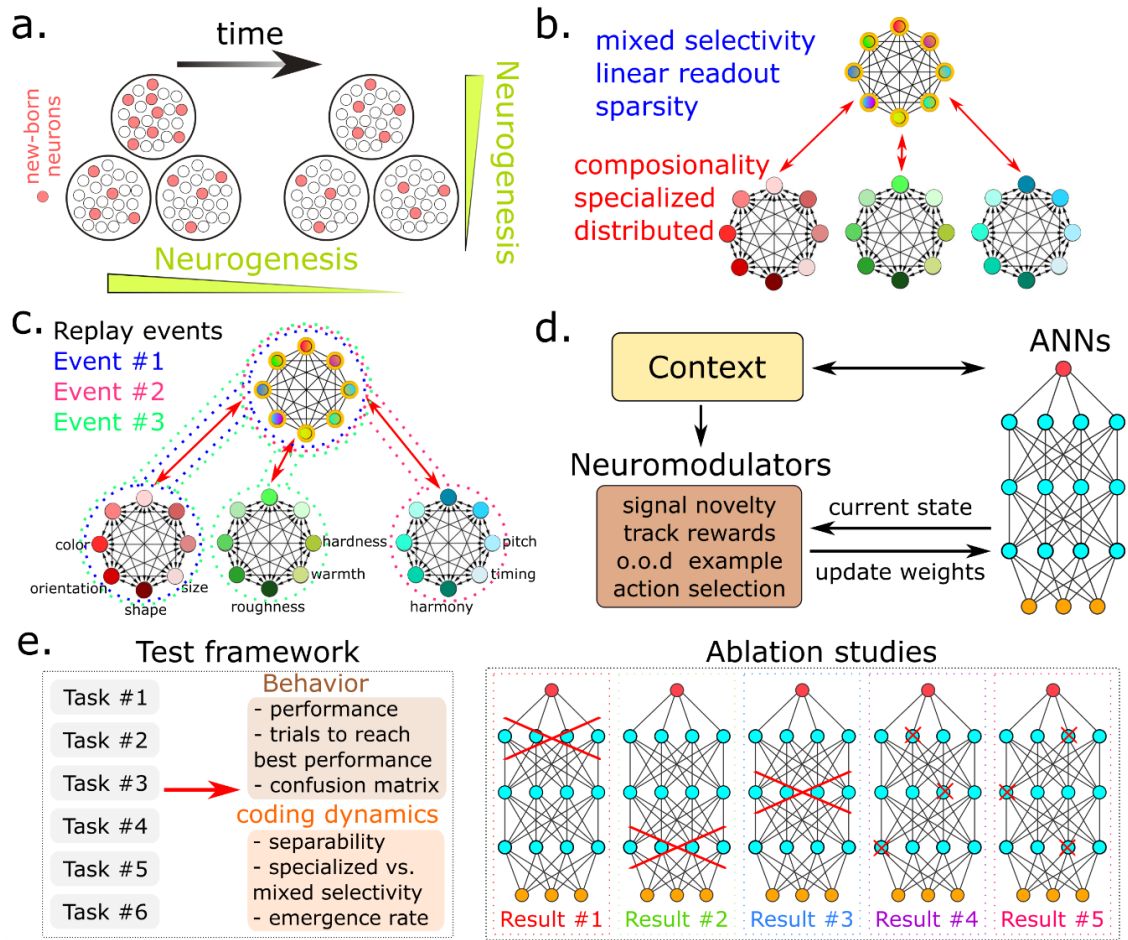


Figure 5: Neuroscience-AI loop. a) illustration of neurogenesis in different modules of an agent across its lifespan. Note that we have shown the same number of total neurons (white + red circles) in all modules to allow better comparison concerning the fraction of newborn neurons (red circles). The agent shows a reduced rate of neurogenesis across time, from initial learning (childhood) to highly learned (adulthood) (left to right). Lower-level modules (primary sensory cortices) tend to typically have a lower rate of neurogenesis throughout than higher-level modules (hippocampus; bottom to top). b) cost functions and learning objectives are different across different modules in the brain (red vs. blue). Rather than a global rule, ANNs can be imbued with different objectives across layers and learning modules similar to our brains. c) biological memory replay involves different brain regions interacting preferentially during different replay events (blue, red, and green). A similar idea can be implemented in ANNs, where different learning modules are engaged across individual joint replay (new + old items) epochs depending on the recently learned item and gradient of forgetting. For example, replay event #1 (blue dotted line) involves replay in the top module with conjunctive coding neurons (similar to HC) and visual feature coding (red color, bottom). Replay event #3 (green dotted line) involves interaction with the top module, visual coding, and tactile coding modules (green color, bottom). We can also

evaluate different rules for replay to see which matches the replay dynamics in biological data. Individual modules are drawn as Hopfield net in b), and c), but that is not a required condition. d) EE animals tend to show higher levels of neuromodulators and better context recognition compared to naïve animals, which might underlie rapid adaptation and forward transfer. A combination of context signal (yellow box) and neuromodulatory inputs (brown box) can be added to ANNs to update hyperparameters and make them work in different learning modes: offline memory replay phase, active learning, high discrimination, or generalization enhancing behavior performance. e) Left: Running biological experiments (Task #1- Task #6) to determine a series of parameters related to behavior and coding dynamics of a knowledge-rich vs. naïve brain. Results from these experiments can then be used to determine objective functions to enforce similar performance in ANNs. Right: Ablation (and manipulation) studies in ANNs to determine the contribution of different modules (hidden layers, first three), a subset of nodes across modules (selective for a given task, last three), and changing learning rules. Such studies can help design precise hypotheses to test in the future.

5. Context gating and pattern separation:

Context signals can help differentiate between similar experiences across different contexts (pattern separation)^{88,89}. Previous studies have shown that context-specific modulation of neurons allows for gating task-relevant cues by increasing attention, thus leading to faster learning of new stimuli^{119,120}. EE animals show increased pattern separation and population sparsity^{17,49,71}. Motivated by this, ANNs can be implemented with a context-gating signal that enables switching between tasks (new and already learned) and separability between internal representations (pattern separation, Fig. 5d)^{121,122}. ANNs can also be programmed to pay attention to and gate a subset of sensory inputs to see how that affects learning. For example, a well-trained network performs well in a high-noise decision-making task by ignoring task-irrelevant features, which might hold true for EE animals.

6. Test frameworks and ablation studies:

We believe EE experiments can aid in developing test frameworks and datasets for ANNs, where we can come up with predictions for behavior (reaction times, time taken to achieve the best performance, number of trials required, etc.) and brain activity during learning (changes in the internal representation, increased sparsity, compositionality, etc.) (Fig. 5e, left). These results can then guide the development of rules to enable brain-like behavior performance and learning in ANNs. Future experiments can control different elements of ANNs through ablations and manipulations, which would be difficult or impossible in biological systems. This makes ANNs a powerful tool for understanding the relative contribution of each input and learning mechanism employed if some mechanisms go haywire (Fig. 5e, right)^{123,124}.

We haven't discussed this, but other interesting approaches such as embodiment or imparting knowledge of intuitive theories of physics can potentially help build enriched ANNs with faster knowledge acquisition and improved problem-solving ^{125,126}. Rapid progress in understanding novel task learning in humans and animals is imminent with the current pace of improvement in recording techniques and a steep rise in interest in neuro-inspired AI. The mechanisms and ideas listed above are just the beginning of this discussion. We believe that EE experiments and deep learning models have much to offer each other, and we hope the current paper has provided some valuable guidelines to harness the strengths of both domains.

Concluding remarks

The unprecedented success of ANNs and significant advances in neural and behavior recording techniques have made researchers appreciate the potential of combining the two fields. In recent decades, there has been remarkable progress in behavior tracking ¹²⁷ and large-scale neural recordings, going from a few hundred neurons to thousands, using Neuropixels probe ¹²⁸, 2-photon mesoscope ¹²⁹, etc. This advancement has enabled experiments that have provided valuable insights into existing questions and generated new hypotheses. However, just recording a large number of neurons will not be sufficient, future questions must be rooted in strong theories. Deep learning neural networks (with careful consideration) can be a powerful tool to analyze neuron representations and uncover connections between brain structure and function ¹³⁰. The aim of this article was to motivate systems neuroscience and AI researchers to interact more closely. In this article, we have discussed how EE experiments can inform new methods to enhance the forward transfer capability of ANNs. We strongly believe that combining several neuro-inspired mechanisms will significantly improve performance and speed up the development of continual learning artificial systems. Similarly, ANNs, especially deep learning models, can provide insights into the emergence of abstract representations from experience and the underlying learning dynamics. Future studies should focus on building models that can explain complex behavior and coding properties without losing interpretability to better guide neuroscience experiments. In closing, we emphasize that understanding continual learning requires a multidisciplinary effort at the intersection of psychology, neuroscience, computer science, and physics. A continuous collaboration between these fields will be critical for developing highly robust ANNs and understanding how learning and memory arise in the mammalian brain.

References

1. Benavides-Prado, D. & Riddle, P. A Theory for Knowledge Transfer in Continual Learning. Preprint at <https://doi.org/10.48550/arXiv.2208.06931> (2022).
2. Ke, Z., Liu, B., Ma, N., Xu, H. & Shu, L. Achieving Forgetting Prevention and Knowledge Transfer in Continual Learning. in *Advances in Neural Information Processing Systems* vol. 34 22443–22456 (Curran Associates, Inc., 2021).
3. Grover, S., Pea, R. & Cooper, S. Factors Influencing Computer Science Learning in Middle School. in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education* 552–557 (Association for Computing Machinery, New York, NY, USA, 2016). doi:10.1145/2839509.2844564.
4. Cresswell, C. & Speelman, C. P. Does mathematics training lead to better logical thinking and reasoning? A cross-sectional assessment from students to professors. *PLOS ONE* **15**, e0236153 (2020).
5. McCloskey, M. & Cohen, N. J. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. in *Psychology of Learning and Motivation* (ed. Bower, G. H.) vol. 24 109–165 (Academic Press, 1989).
6. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C. & Wermter, S. Continual Lifelong Learning with Neural Networks: A Review. *Neural Netw.* **113**, 54–71 (2019).
7. Aljundi, R. Continual Learning in Neural Networks. Preprint at <https://doi.org/10.48550/arXiv.1910.02718> (2019).
8. Saxe, A., Nelli, S. & Summerfield, C. If deep learning is the answer, what is the question? *Nat. Rev. Neurosci.* **22**, 55–67 (2021).
9. Richards, B. A. *et al.* A deep learning framework for neuroscience. *Nat. Neurosci.* **22**, 1761–1770 (2019).
10. Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick, M. Neuroscience-Inspired Artificial Intelligence. *Neuron* **95**, 245–258 (2017).
11. Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
12. Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
13. Kempermann, G. Environmental enrichment, new neurons and the neurobiology of individuality. *Nat. Rev. Neurosci.* **20**, 235–245 (2019).
14. van Praag, H., Kempermann, G. & Gage, F. H. Neural consequences of environmental enrichment. *Nat. Rev. Neurosci.* **1**, 191–198 (2000).
15. Gattas, S. *et al.* A rodent obstacle course procedure controls delivery of enrichment and enhances complex cognitive functions. *NPJ Sci. Learn.* **7**, 21 (2022).
16. Hebb, D. O. *The Organization of Behavior; a Neuropsychological Theory*. xix, 335 (Wiley, Oxford, England, 1949).
17. Zentall, T. R. Effect of Environmental Enrichment on the Brain and on Learning and Cognition by Animals. *Anim. Open Access J. MDPI* **11**, 973 (2021).
18. Emery, B. A., Hu, X., Khanzada, S., Kempermann, G. & Amin, H. Rich Experience Boosts Functional Connectome and High-Dimensional Coding in Hippocampal Network. 2022.02.23.480123 Preprint at <https://doi.org/10.1101/2022.02.23.480123> (2022).

19. Clemenson, G. D. & Stark, C. E. L. Virtual Environmental Enrichment through Video Games Improves Hippocampal-Associated Memory. *J. Neurosci.* **35**, 16116–16125 (2015).
20. McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
21. Marr, D. Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **262**, 23–81 (1971).
22. Tse, D. *et al.* Schemas and memory consolidation. *Science* **316**, 76–82 (2007).
23. Fernández, G. & Morris, R. G. M. Memory, Novelty and Prior Knowledge. *Trends Neurosci.* **41**, 654–659 (2018).
24. McClelland, J. L. Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *J. Exp. Psychol. Gen.* **142**, 1190–1210 (2013).
25. van Kesteren, M. T. R., Ruiters, D. J., Fernández, G. & Henson, R. N. How schema and novelty augment memory formation. *Trends Neurosci.* **35**, 211–219 (2012).
26. McClelland, J. L., McNaughton, B. L. & Lampinen, A. K. Integration of new information in memory: new insights from a complementary learning systems perspective. *Philos. Trans. R. Soc. B Biol. Sci.* **375**, 20190637 (2020).
27. Saxena, R., Shobe, J. L. & McNaughton, B. L. Learning in deep neural networks and brains with similarity-weighted interleaved learning. *Proc. Natl. Acad. Sci.* **119**, e2115229119 (2022).
28. Saxe, A. M., McClelland, J. L. & Ganguli, S. A mathematical theory of semantic development in deep neural networks. *Proc. Natl. Acad. Sci.* **116**, 11537–11546 (2019).
29. Goudar, V., Peysakhovich, B., Freedman, D. J., Buffalo, E. A. & Wang, X.-J. Schema formation in a neural population subspace underlies learning-to-learn in flexible sensorimotor problem-solving. *Nat. Neurosci.* (2023) doi:10.1038/s41593-023-01293-9.
30. Hwu, T. & Krichmar, J. L. A neural model of schemas and memory encoding. *Biol. Cybern.* **114**, 169–186 (2020).
31. Davis, G. P., Katz, G. E., Gentili, R. J. & Reggia, J. A. Compositional memory in attractor neural networks with one-step learning. *Neural Netw.* **138**, 78–97 (2021).
32. Reverberi, C., Görden, K. & Haynes, J.-D. Compositionality of Rule Representations in Human Prefrontal Cortex. *Cereb. Cortex* **22**, 1237–1246 (2012).
33. Frankland, S. M. & Greene, J. D. Concepts and Compositionality: In Search of the Brain's Language of Thought. *Annu. Rev. Psychol.* **71**, 273–303 (2020).
34. Ito, T. *et al.* Compositional generalization through abstract representations in human and artificial neural networks. Preprint at <https://doi.org/10.48550/arXiv.2209.07431> (2022).
35. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).
36. Wang, J. X. Meta-learning in natural and artificial intelligence. *Curr. Opin. Behav. Sci.* **38**, 90–95 (2021).
37. Janssen, H. *et al.* Environmental enrichment: neurophysiological responses and consequences for health. in *Oxford Textbook of Nature and Public Health: The role of nature in improving the health of a population* (eds. van den Bosch, M., Bird, W., van den

- Bosch, M. & Bird, W.) 0 (Oxford University Press, 2018).
doi:10.1093/med/9780198725916.003.0039.
38. Balianetti, M. & Conti, F. Environmental enrichment and the aging brain: is it time for standardization? *Neurosci. Biobehav. Rev.* **139**, 104728 (2022).
 39. Ratuski, A. S. & Weary, D. M. Environmental Enrichment for Rats and Mice Housed in Laboratories: A Metareview. *Anim. Open Access J. MDPI* **12**, 414 (2022).
 40. Cheng, S.-T., Liu, S., Ou-Yang, B., Dai, X.-Y. & Cheng, L. Specific Effects of Characteristics of Enriched Environment on Innovative Problem Solving by Animals. *Psychol. Sci.* **33**, 1097–1111 (2022).
 41. Rountree-Harrison, D., Burton, T. J., Leamey, C. A. & Sawatari, A. Environmental Enrichment Expedites Acquisition and Improves Flexibility on a Temporal Sequencing Task in Mice. *Front. Behav. Neurosci.* **12**, (2018).
 42. Kempermann, G., Kuhn, H. G. & Gage, F. H. More hippocampal neurons in adult mice living in an enriched environment. *Nature* **386**, 493–495 (1997).
 43. Manno, F. A. M. *et al.* Environmental enrichment leads to behavioral circadian shifts enhancing brain-wide functional connectivity between sensory cortices and eliciting increased hippocampal spiking. *NeuroImage* **252**, 119016 (2022).
 44. Kobayashi, S., Ohashi, Y. & Ando, S. Effects of enriched environments with different durations and starting times on learning capacity during aging in rats assessed by a refined procedure of the Hebb-Williams maze task. *J. Neurosci. Res.* **70**, 340–346 (2002).
 45. Bennett, J. C., McRae, P. A., Levy, L. J. & Frick, K. M. Long-term continuous, but not daily, environmental enrichment reduces spatial memory decline in aged male mice. *Neurobiol. Learn. Mem.* **85**, 139–152 (2006).
 46. Kramer, A. F., Bherer, L., Colcombe, S. J., Dong, W. & Greenough, W. T. Environmental influences on cognitive and brain plasticity during aging. *J. Gerontol. A. Biol. Sci. Med. Sci.* **59**, M940-957 (2004).
 47. Williams, B. M. *et al.* Environmental enrichment: effects on spatial memory and hippocampal CREB immunoreactivity. *Physiol. Behav.* **73**, 649–658 (2001).
 48. Nithianantharajah, J. & Hannan, A. J. Enriched environments, experience-dependent plasticity and disorders of the nervous system. *Nat. Rev. Neurosci.* **7**, 697–709 (2006).
 49. Woodcock, E. A. & Richardson, R. Effects of Environmental Enrichment on Rate of Contextual Processing and Discriminative Ability in Adult Rats. *Neurobiol. Learn. Mem.* **73**, 1–10 (2000).
 50. Zeleznikow-Johnston, A., Burrows, E. L., Renoir, T. & Hannan, A. J. Environmental enrichment enhances cognitive flexibility in C57BL/6 mice on a touchscreen reversal learning task. *Neuropharmacology* **117**, 219–226 (2017).
 51. Babaei, P. & Azari, H. B. Exercise Training Improves Memory Performance in Older Adults: A Narrative Review of Evidence and Possible Mechanisms. *Front. Hum. Neurosci.* **15**, 771553 (2022).
 52. Clouston, S. A. P. *et al.* Education and Cognitive Decline: An Integrative Analysis of Global Longitudinal Studies of Cognitive Aging. *J. Gerontol. Ser. B* **75**, e151–e160 (2020).
 53. Xu, W., Yu, J.-T., Tan, M.-S. & Tan, L. Cognitive Reserve and Alzheimer’s Disease. *Mol. Neurobiol.* **51**, 187–208 (2015).

54. Stern, Y. What is cognitive reserve? Theory and research application of the reserve concept. *J. Int. Neuropsychol. Soc. JINS* **8**, 448–460 (2002).
55. Scarmeas, N. & Stern, Y. Cognitive Reserve and Lifestyle. *J. Clin. Exp. Neuropsychol.* **25**, 625–633 (2003).
56. Baldivia, B., Andrade, V. M. & Bueno, O. F. A. Contribution of education, occupation and cognitively stimulating activities to the formation of cognitive reserve. *Dement. Neuropsychol.* **2**, 173–182 (2008).
57. Jirout, J. *et al.* How Lifestyle Factors Affect Cognitive and Executive Function and the Ability to Learn in Children. *Nutrients* **11**, 1953 (2019).
58. Speisman, R. B. *et al.* Environmental enrichment restores neurogenesis and rapid acquisition in aged rats. *Neurobiol. Aging* **34**, 263–274 (2013).
59. Clemenson, G. D., Deng, W. & Gage, F. H. Environmental enrichment and neurogenesis: from mice to humans. *Curr. Opin. Behav. Sci.* **4**, 56–62 (2015).
60. Cortese, G. P., Olin, A., O’Riordan, K., Hullinger, R. & Burger, C. Environmental enrichment improves hippocampal function in aged rats by enhancing learning and memory, LTP, and mGluR5-Homer1c activity. *Neurobiol. Aging* **63**, 1–11 (2018).
61. Artola, A. *et al.* Long-lasting modulation of the induction of LTD and LTP in rat hippocampal CA1 by behavioural stress and environmental enrichment. *Eur. J. Neurosci.* **23**, 261–272 (2006).
62. Schapiro, S. & Vukovich, K. R. Early experience effects upon cortical dendrites: a proposed model for development. *Science* **167**, 292–294 (1970).
63. Jung, C. K. E. & Herms, J. Structural Dynamics of Dendritic Spines are Influenced by an Environmental Enrichment: An In Vivo Imaging Study. *Cereb. Cortex* **24**, 377–384 (2014).
64. Connor, J. R., Wang, E. C. & Diamond, M. C. Increased length of terminal dendritic segments in old adult rats’ somatosensory cortex: an environmentally induced response. *Exp. Neurol.* **78**, 466–470 (1982).
65. Sun, H. *et al.* Environmental enrichment influences BDNF and NR1 levels in the hippocampus and restores cognitive impairment in chronic cerebral hypoperfused rats. *Curr. Neurovasc. Res.* **7**, 268–280 (2010).
66. Ciucci, F. *et al.* Insulin-like growth factor 1 (IGF-1) mediates the effects of enriched environment (EE) on visual cortical development. *PLoS One* **2**, e475 (2007).
67. Rosenzweig, M. R., Krech, D., Bennett, E. L. & Diamond, M. C. Effects of environmental complexity and training on brain chemistry and anatomy: a replication and extension. *J. Comp. Physiol. Psychol.* **55**, 429–437 (1962).
68. Brenes, J. C., Rodríguez, O. & Fornaguera, J. Differential effect of environment enrichment and social isolation on depressive-like behavior, spontaneous activity and serotonin and norepinephrine concentration in prefrontal cortex and ventral striatum. *Pharmacol. Biochem. Behav.* **89**, 85–93 (2008).
69. Engineer, N. D. *et al.* Environmental Enrichment Improves Response Strength, Threshold, Selectivity, and Latency of Auditory Cortex Neurons. *J. Neurophysiol.* **92**, 73–82 (2004).
70. LeMessurier, A. M. *et al.* Enrichment drives emergence of functional columns and improves sensory coding in the whisker map in L2/3 of mouse S1. *eLife* **8**, e46321 (2019).
71. Bilkey, D. K. *et al.* Exposure to complex environments results in more sparse representations of space in the hippocampus. *Hippocampus* **27**, 1178–1191 (2017).

72. Sampedro-Piquero, P. *et al.* Environmental Enrichment Results in Both Brain Connectivity Efficiency and Selective Improvement in Different Behavioral Tasks. *Neuroscience* **388**, 374–383 (2018).
73. Bogado Lopes, J. *et al.* Individual behavioral trajectories shape whole-brain connectivity in mice. *eLife* **12**, e80379 (2023).
74. Aljundi, R. *et al.* Online Continual Learning with Maximally Interfered Retrieval. *ArXiv190804742 Cs Stat* (2019).
75. Cai, Z., Sener, O. & Koltun, V. Online Continual Learning with Natural Distribution Shifts: An Empirical Study with Visual Data. Preprint at <https://doi.org/10.48550/arXiv.2108.09020> (2021).
76. Ramasesh, V. V., Dyer, E. & Raghu, M. Anatomy of Catastrophic Forgetting: Hidden Representations and Task Semantics. in (2021).
77. Ramasesh, V. V., Lewkowycz, A. & Dyer, E. Effect of scale on catastrophic forgetting in neural networks. in (2022).
78. Davidson, G. & Mozer, M. C. Sequential mastery of multiple visual tasks: Networks naturally learn to learn and forget to forget. *ArXiv190510837 Cs Stat* (2020).
79. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016).
80. Sorscher, B., Ganguli, S. & Sompolinsky, H. Neural representational geometry underlies few-shot concept learning. *Proc. Natl. Acad. Sci.* **119**, e2200800119 (2022).
81. Froudarakis, E. *et al.* Object manifold geometry across the mouse cortical visual hierarchy. 2020.08.20.258798 Preprint at <https://doi.org/10.1101/2020.08.20.258798> (2021).
82. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
83. Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X.-J. Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* **22**, 297–306 (2019).
84. Johnston, W. J., Palmer, S. E. & Freedman, D. J. Nonlinear mixed selectivity supports reliable neural computation. *PLOS Comput. Biol.* **16**, e1007544 (2020).
85. Mensink, T., Uijlings, J., Kuznetsova, A., Gygli, M. & Ferrari, V. Factors of Influence for Transfer Learning across Diverse Appearance Domains and Task Types. Preprint at <https://doi.org/10.48550/arXiv.2103.13318> (2021).
86. Cole, M. W., Laurent, P. & Stocco, A. Rapid instructed task learning: A new window into the human brain's unique capacity for flexible cognitive control. *Cogn. Affect. Behav. Neurosci.* **13**, 1–22 (2013).
87. Barak, O., Rigotti, M. & Fusi, S. The Sparseness of Mixed Selectivity Neurons Controls the Generalization–Discrimination Trade-Off. *J. Neurosci.* **33**, 3844–3856 (2013).
88. Yassa, M. A. & Stark, C. E. L. Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–525 (2011).
89. McNaughton, B. L. & Nadel, L. Hebb-Marr networks and the neurobiological representation of action in space. in *Neuroscience and connectionist theory* 1–63 (Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1990).
90. Marr, D. A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470 (1969).

91. Eriksson, P. S. *et al.* Neurogenesis in the adult human hippocampus. *Nat. Med.* **4**, 1313–1317 (1998).
92. Lledo, P.-M. & Valley, M. Adult Olfactory Bulb Neurogenesis. *Cold Spring Harb. Perspect. Biol.* **8**, a018945 (2016).
93. Tuncdemir, S. N. *et al.* Adult-born granule cells facilitate remapping of spatial and non-spatial representations in the dentate gyrus. *Neuron* S0896-6273(23)00703–1 (2023) doi:10.1016/j.neuron.2023.09.016.
94. Lods, M. *et al.* Adult-born neurons immature during learning are necessary for remote memory reconsolidation in rats. *Nat. Commun.* **12**, 1778 (2021).
95. Kitamura, T. *et al.* Adult Neurogenesis Modulates the Hippocampus-Dependent Period of Associative Fear Memory. *Cell* **139**, 814–827 (2009).
96. Miller, S. M. & Sahay, A. Functions of adult-born neurons in hippocampal memory interference and indexing. *Nat. Neurosci.* **22**, 1565–1575 (2019).
97. Aimone, J. B., Wiles, J. & Gage, F. H. Computational Influence of Adult Neurogenesis on Memory Encoding. *Neuron* **61**, 187–202 (2009).
98. Aimone, J. B., Deng, W. & Gage, F. H. Resolving new memories: a critical look at the dentate gyrus, adult neurogenesis, and pattern separation. *Neuron* **70**, 589–596 (2011).
99. Marblestone, A. H., Wayne, G. & Kording, K. P. Toward an Integration of Deep Learning and Neuroscience. *Front. Comput. Neurosci.* **10**, (2016).
100. Ostapenko, O., Rodriguez, P., Caccia, M. & Charlin, L. Continual Learning via Local Module Composition. Preprint at <https://doi.org/10.48550/arXiv.2111.07736> (2021).
101. Nazari, M. *et al.* Regional variation in cholinergic terminal activity determines the non-uniform occurrence of cortical slow waves during REM sleep in mice. *Cell Rep.* **42**, 112450 (2023).
102. Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994).
103. Hayes, T. L. *et al.* Replay in Deep Learning: Current Approaches and Missing Biological Elements. Preprint at <https://doi.org/10.48550/arXiv.2104.04132> (2021).
104. Abadchi, J. K. *et al.* Spatiotemporal patterns of neocortical activity around hippocampal sharp-wave ripples. *eLife* <https://elifesciences.org/articles/51972> (2020) doi:10.7554/eLife.51972.
105. Dahal, P., Rauhala, O. J., Khodagholy, D. & Gelineas, J. N. Hippocampal–cortical coupling differentiates long-term memory processes. *Proc. Natl. Acad. Sci.* **120**, e2207909120 (2023).
106. Jiang, X., Gonzalez-Martinez, J. & Halgren, E. Coordination of Human Hippocampal Sharpwave Ripples during NREM Sleep with Cortical Theta Bursts, Spindles, Downstates, and Upstates. *J. Neurosci. Off. J. Soc. Neurosci.* **39**, 8744–8761 (2019).
107. Kaefer, K., Stella, F., McNaughton, B. L. & Battaglia, F. P. Replay, the default mode network and the cascaded memory systems model. *Nat. Rev. Neurosci.* **23**, 628–640 (2022).
108. Seeman, S. C. *et al.* Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *eLife* **7**, e37349 (2018).
109. Bassett, D. S. & Bullmore, E. Small-world brain networks. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **12**, 512–523 (2006).

110. Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
111. Hasselmo, M. E. & McGaughy, J. High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation. in *Progress in Brain Research* vol. 145 207–231 (Elsevier, 2004).
112. Avery, M. C. & Krichmar, J. L. Neuromodulatory Systems and Their Interactions: A Review of Models, Theories, and Experiments. *Front. Neural Circuits* **11**, 108 (2017).
113. Yu, A. J. & Dayan, P. Uncertainty, Neuromodulation, and Attention. *Neuron* **46**, 681–692 (2005).
114. Hasselmo, M. E. Neuromodulation: acetylcholine and memory consolidation. *Trends Cogn Sci EDAT- 19990826 MHDA- 19990826 AID - S1364661399013650 Pii PST - Ppublish* **3**, 351–359 (1999).
115. Brna, A. P. *et al.* Uncertainty-based modulation for lifelong learning. *Neural Netw.* **120**, 129–142 (2019).
116. Sun, S., Calandriello, D., Hu, H., Li, A. & Titsias, M. Information-theoretic Online Memory Selection for Continual Learning. Preprint at <https://doi.org/10.48550/arXiv.2204.04763> (2022).
117. Atherton, L. A., Dupret, D. & Mellor, J. R. Memory trace replay: the shaping of memory consolidation by neuromodulation. *Trends Neurosci.* **38**, 560–570 (2015).
118. Rouhani, N., Norman, K. A., Niv, Y. & Bornstein, A. M. Reward prediction errors create event boundaries in memory. *Cognition* **203**, 104269 (2020).
119. Rikhye, R. V., Gilra, A. & Halassa, M. M. Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nat. Neurosci.* **21**, 1753–1763 (2018).
120. Johnston, K., Levin, H. M., Koval, M. J. & Everling, S. Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron* **53**, 453–462 (2007).
121. Masse, N. Y., Grant, G. D. & Freedman, D. J. Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *Proc. Natl. Acad. Sci.* **115**, E10467–E10475 (2018).
122. Flesch, T., Nagy, D. G., Saxe, A. & Summerfield, C. Modelling continual learning in humans with Hebbian context gating and exponentially decaying task signals. Preprint at <https://doi.org/10.48550/arXiv.2203.11560> (2022).
123. van de Ven, G. M., Siegelmann, H. T. & Tolias, A. S. Brain-inspired replay for continual learning with artificial neural networks. *Nat. Commun.* **11**, 4069 (2020).
124. Meyes, R., Lu, M., de Puiseau, C. W. & Meisen, T. Ablation Studies in Artificial Neural Networks. Preprint at <https://doi.org/10.48550/arXiv.1901.08644> (2019).
125. Krichmar, J. L. & Hwu, T. J. Design Principles for Neurorobotics. *Front. Neurobotics* **16**, (2022).
126. Piloto, L. S., Weinstein, A., Battaglia, P. & Botvinick, M. Intuitive physics learning in a deep-learning model inspired by developmental psychology. *Nat. Hum. Behav.* **6**, 1257–1267 (2022).
127. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).

128. Steinmetz, N. A. *et al.* Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science* **372**, eabf4588 (2021).
129. Sofroniew, N. J., Flickinger, D., King, J. & Svoboda, K. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *eLife* **5**, e14472 (2016).
130. Higgins, I. *et al.* Unsupervised deep learning identifies semantic disentanglement in single inferotemporal face patch neurons. *Nat. Commun.* **12**, 6456 (2021).