



Lessons from infant learning for unsupervised machine learning

Lorijn Zaadnoordijk¹✉, Tarek R. Besold²✉ and Rhodri Cusack¹✉

The desire to reduce the dependence on curated, labeled datasets and to leverage the vast quantities of unlabeled data has triggered renewed interest in unsupervised (or self-supervised) learning algorithms. Despite improved performance due to approaches such as the identification of disentangled latent representations, contrastive learning and clustering optimizations, unsupervised machine learning still falls short of its hypothesized potential as a breakthrough paradigm enabling generally intelligent systems. Inspiration from cognitive (neuro)science has been based mostly on adult learners with access to labels and a vast amount of prior knowledge. To push unsupervised machine learning forward, we argue that developmental science of infant cognition might hold the key to unlocking the next generation of unsupervised learning approaches. We identify three crucial factors enabling infants' quality and speed of learning: (1) babies' information processing is guided and constrained; (2) babies are learning from diverse, multimodal inputs; and (3) babies' input is shaped by development and active learning. We assess the extent to which these insights from infant learning have already been exploited in machine learning, examine how closely these implementations resemble the core insights, and propose how further adoption of these factors can give rise to previously unseen performance levels in unsupervised learning.

Early deep neural network (DNN) architectures commonly implemented a paradigm using unsupervised pre-training followed by supervised fine-tuning^{1,2}. Inclusion of the initial unsupervised stage was the key to unlocking effective training strategies for deep architectures in the mid-2000s (see, for example, refs. 3–5 for details). Since the mid-2010s, unsupervised pre-training has become less popular and — especially in computer vision applications — has to a certain degree been superseded by supervised pre-training of models using large-scale labeled datasets (such as ImageNet⁶). Although supervised learning has shown impressive performance in a variety of applications (for example, refs. 7–9), a major drawback is its dependency on large-scale, well-curated datasets. Moreover, there is reasonable doubt regarding the efficacy¹⁰ and further expandability^{11,12} of the paradigm. These and related considerations have triggered renewed interest in unsupervised (in current literature sometimes also referred to as self-supervised) learning (for example, refs. 13–16).

Unsupervised representation learning is attracting considerable attention from neuroscience and cognitive science: DNNs seem to offer a loosely biologically inspired modelling and analysis tool for the study of the human brain and mind, with representation learning playing a central role^{17,18}. In fact, there is a long-standing tradition of prominent and active exchanges between machine learning (ML) and these fields¹⁹, going back at least to Rosenblatt's perceptron²⁰ and its inspiration by 'nerve net'²¹. Increasingly, there are calls to look again at animals and humans for insights into their biological neural machinery and how these natural-born intelligent systems learn^{17,22–26}. Commonly, these contributions take mature human cognizers as the conceptual starting point. In contrast, we advocate to focus on results from the study of infants and their development. Human infants are in many ways a close counterpart to a computational system learning in an unsupervised manner, as infants too must learn useful representations from unlabeled data. However, infants' representations are learned rapidly, with relatively

few examples from a relatively small number of classes as compared to the vast datasets required for training state-of-the-art DNNs (see, for example, the 15 million high-resolution images with 22,000 label classes used for training ImageNet⁶), and can be used flexibly and efficiently in various different tasks and contexts. Developmental science, the field studying cognitive development in human infants and children, has identified several facilitating factors for this surprising performance, which merit discussion also for ML. Specifically, they may hold the answers to some of the long-standing questions of representation learning (for example, ref. 27): What can a good representation buy us? What is a good representation in the first place? And what training principles can help in discovering such representations?

We are not the first to seek recourse from developmental research in an ML context^{28–34}. However, these earlier efforts remained on a very general level regarding their engagement with the relevant insights from developmental science. In contrast, in this Perspective, we give specific suggestions for three aspects of infant learning that ML researchers should consider, and discuss to what extent current ML research is already — consciously or by coincidence — working towards integrating these insights. We mostly limit our focus to developmental learning until the first year after birth, as this learning period is most similar to the process of unsupervised ML. Although human unsupervised learning occurs later in life as well, the likelihood that this process is influenced by labels or knowledge obtained through supervised learning increases, making it impossible to establish whether that unsupervised learning process could have happened without supervised learning. Accordingly, we outline the core findings from developmental science that have been crucial for understanding learning in early infancy and that offer valuable inspiration for the advancement of unsupervised ML. These findings relate to the way the neural architecture and cognitive biases guide and constrain early information processing (lesson 1), the richness and multimodal diversity of infants' inputs

¹Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. ²Philosophy & Ethics Group, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands. ✉e-mail: L.Zaadnoordijk@tcd.ie; T.R.Besold@tue.nl; CusackRh@tcd.ie

(lesson 2) and the way inputs are shaped by development and active learning (lesson 3). These lessons do not cover everything that is known about infant learning^{35–39}. However, the lessons considered in this Perspective play an important role in infants' cognitive development and have a proven potential to integrate with and improve ML.

Three lessons from infant development for machine learning

We discuss three main lessons synthesized from the current state of knowledge regarding infant learning, and highlight their direct relevance for advancing unsupervised ML. We initially zoom in on the starting conditions for unsupervised learning in infants, tying into recent work on pre-wiring and pre-training of DNNs. Subsequently, we summarize insights about multimodal statistical learning in infants and establish the connection to corresponding efforts in ML. Finally, we engage with the mechanisms that shape infants' inputs. There, we provide an overview of research into how developmental changes affect infants' inputs, which connects directly to concepts such as non-stationarity in continual representation learning, and we discuss how infants and machines process these inputs during active learning.

Babies' information processing is guided and constrained. It is a common belief that infants' behavioural immaturity stems from a highly immature brain architecture^{40,41}. Indeed, grey matter volume and surface area do not peak until around six and eleven years old, respectively, and at six months old they are growing at the fastest rate of the whole lifespan⁴². Furthermore, processes such as synaptic pruning⁴³ and myelination⁴⁴ cause important changes in the young brain. However, recent neuroimaging research shows that, although the infant brain is plastic and still grows and develops, a lot of its structure is already present very early on. By term age, magnetic resonance imaging has shown that the cortex is folded like that of an adult⁴⁵, major sulci are present⁴⁶, and structural connectivity patterns observed in early infancy are very similar to adult structural connectivity⁴⁷. Adult-like functional connectivity patterns can also be observed in infants^{48,49} even for networks that support cognitive capacities that are not yet behaviourally manifested, such as speech⁵⁰. Functional visual cortex activation of abstract categories (faces and scenes) is adult-like in its spatial organization in infants as young as four months of age and is refined in its response pattern through development⁵¹. Even brain regions responsible for higher cognitive functions, like the hippocampus⁵² and frontal lobes^{53,54}, are active in early infancy. Moreover, functional connectivity patterns in neonates have been shown to be predictive of later development⁵⁵. Thus, although, compared to adults, the neural structure in infants is still more plastic and can change dramatically depending on the type of input it receives, it is already fairly determined as a whole. Infant learning is probably made possible by the strong plastic nature of the brain (for more information, see the last paragraph of this section as well as the section on the mechanisms shaping infants' learning input), while being guided by the constrained processing through the early brain structure (Box 1).

In ref. ²⁶ it is argued that creators of neural networks should take into consideration that the highly structured neural connectivity is what allows animals to learn so rapidly. Indeed, ML researchers have been exploring the impact of pre-wiring on network learning. A recent example is work on learning general identity rules. Based on the results of infant experiments using artificial grammar tasks, in ref. ⁵⁶ it was suggested with regard to simple recurrent networks⁵⁷ that 'such mechanisms cannot account for how humans generalize rules to new items that do not overlap with the items that appeared in training'. If correct, this would imply substantial limitations on the generalization capabilities of an entire family of recurrent neural networks. Revisiting the claim and the debate triggered by the

Box 1 | Lesson 1 from infant learning: constraining and guiding information processing

Even when not invoking rich interpretations of newborn cognition, the earliest processing of information is constrained by the neural architecture and guided by biases. Developmental science suggests that it is not only the type of ML architecture selected and the training algorithm or the training data that matter, but that the particular instantiation of the architecture and set-up of the network play an important role as enablers of efficient and effective training. These insights find a direct ML counterpart in the importance of starting conditions and the growing efforts invested in the study of pre-wiring and pre-training of neural networks. Architectural paradigms such as fibred neural networks offer ways to create richer models, which — when combined with novel, more efficient approaches to hyperparameter optimization — may unlock previously unachievable levels of generality for ML systems. In summary, not all inductive biases have been made equal and infants do not have to start their developmental trajectory from a tabula rasa or from an arbitrary starting set-up — and neither should neural networks.

original paper, ref. ⁵⁸ introduced a delay-line memory (a concept first suggested in neuroscience⁵⁹) into an Echo State Network⁶⁰ and showed that — in combination with a training algorithm incrementally challenging the network with novel stimuli — this indeed enables the network to perform a comparable learning task.

Additional constraints may arise from 'pre-programmed' capabilities. Although there is consensus that both nature and nurture play a role in development, the developmental science community is still heavily debating what processes and building blocks of cognition might already be present at birth, including what the properties of these processes and building blocks are, and what has to be learned through experience (for example, refs. ^{61–63}). Thus, many appeals to inborn knowledge as an explanation for empirical observations are heavily contested. However, there are some findings of biases in newborns on which there is general consensus. Rather than being pre-programmed knowledge, these inductive biases shape the process of learning from the environment. For example, infants prefer biological motion⁶⁴ and have a general attentional bias for top-heavy visual information^{64–66}, which develops into a preference for faces within the first months^{67,68}. Early processing biases are found also in non-visual domains, such as in audition and speech^{69–71}. Regarding the implementation of such biases, it has been found, for instance, that the strong human priors towards finding continuous lines and contiguous surfaces probably arise from the visual cortical micro-circuitry through local cortical motifs like feature-based lateral and top-down suppression and facilitation⁷². Initial attempts at integrating these insights into artificial neural networks (ANNs) have given rise to a DNN model⁷³ solving contour detection tasks with better sample efficiency than state-of-the-art feedforward networks.

Going beyond the general notion of inductive bias in ML architectures^{74,75}, biases play an important role in the training of DNNs. Even in the early days, the initial unsupervised pre-training stage — in a way similar to regularization — also served the purpose of introducing a type of 'starting bias' into the architecture, namely a reduction in variance and a shift in the parameter space towards useful configurations for supervised learning⁵. Since then, the study of biases in networks has enjoyed increasing interest, for example, regarding the role of shape bias within networks in performing visual tasks⁷⁶ or in the context of learning identity relationships with different network architectures⁷⁷. In the latter work, the authors show that relation based patterns⁷⁸ can be implemented as a Bayesian prior on the network weights, helping to overcome the

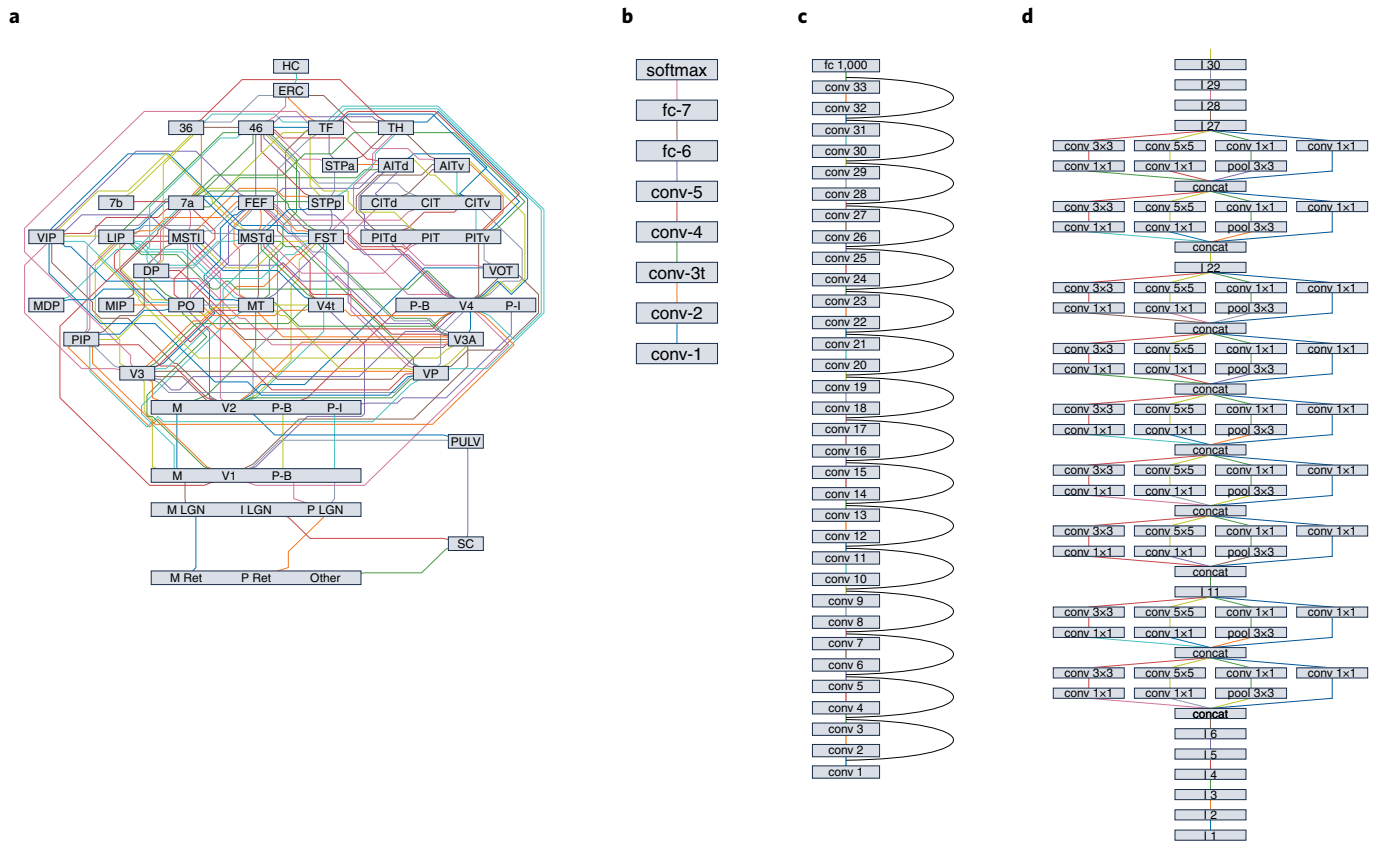


Fig. 1 | Schematic overview of the connectivity within biological and artificial neural networks. **a**, The macaque visual cortex⁸⁰. **b**, AlexNet⁸¹. **c**, ResNet⁸². **d**, Inception⁸³. There is a large difference in the level of complexity between the primate visual cortex and the three ANNs. Figure created with code adapted from <https://github.com/ericmjonas/vanessen> using macaque network rendering from Neurooptikon, Janelia Farm Research Campus Software.

limitations neural networks frequently exhibit related to the learning of identity rules and to generalization beyond the training dataset. In computer vision, the usefulness of combining innate biases and learning mechanisms was demonstrated by the striking performance of a model that was trained to be sensitive to mover events (that is, events in which something changes a previously stationary part of the image, which often found hands grasping objects in the training dataset) — a parallel to infants' bias for biological movement — in recognizing hands and gaze direction from video input⁷⁹.

Given these early successes and the insights from developmental science, we believe that there is overwhelming evidence for the role biases and architectural constraints can play in guiding and augmenting the training of neural networks. Unsupervised ML needs to explore richer architectures that are likely to be pre-wired in some ways and plastic in others. The human neural network has a very complex architecture. If one compares the connectivity of the primate visual cortex (mapped out, for instance, for the macaque⁸⁰) to ANNs typically used for vision (for example, AlexNet⁸¹, ResNet⁸² and Inception⁸³), the difference is unmistakable (Fig. 1). In addition, although the human neural network is strongly recurrent, many ANNs in use today still follow an exclusively feedforward architectural paradigm. Thus far, 'evolution by engineers' has not yielded a change towards a level of network complexity and richness that is comparable to the human brain, by any measure. An example for a promising method to change this comes from the neural-symbolic integration⁸⁴ community: fibring⁸⁵. The fibring paradigm allows the putting together of different types of computational system (for example, different types of neural network, Bayesian network or even logical systems) to work together in a coordinated manner as a 'network of networks', including recursively embedded subnetworks,

subnetworks whose connection weights depend on the activation status of other neurons (that is, subnetworks) in the global network, and so on.

Another avenue worth exploring comprises richer and more complex hyperparameter spaces as a starting point. Following the analogy to infant learning, these could allow us to encode the counterparts of biological evolution or parental instructions and teaching, with the agent subsequently optimizing the model during training. Possible implementation mechanisms can include the optimization of model parameters on the hyperparameter level, altering the topology and/or the size of the network top-down, or the approximation of individual connection weights to 0 over the course of a training cycle, effectively pruning the network's topology bottom-up. Combining these notions with richer architectural paradigms such as fibring networks, the distinction between hyperparameter tuning and parameter training starts to weaken. There, parameter training on the level of the top-level 'network of networks' amounts to hyperparameter tuning of the overall, fully expanded network structure including all subnetworks, effectively reconciling the top-down and bottom-up evolution in a mechanism more akin to the evolution of neural networks in the human brain. Pointing towards the later discussion regarding the impact different learning inputs can have, in a developmentally inspired implementation the hyperparameter space could then be seen as suggesting a learning curriculum, facilitating model fine-tuning via active learning. Having said that, it is unlikely that current hyperparameter optimization methods will in practice be able to deal with the complexity of networks like the ones we proposed earlier. The combinatorial space resulting from millions of possible values for the respective parameters within the neural architecture soon becomes intractable

using current computational resources and brute-force approaches. Deploying methods like fibring will thus also require progress to be made on the efficiency of hyperparameter optimization methods applicable within the respective architectural paradigm. Still, we are convinced that solving these open challenges and subsequently deploying these approaches will prove to be crucial for achieving the complex but plastic nature of the human neural architecture.

Babies learn statistical relations across diverse inputs. There is an abundance of literature studying the capacity of infants for detecting and learning from statistical regularities in their physical and social environment across different domains and from birth^{86,87}. Infants readily learn simple associations, such as sensorimotor contingencies^{88,89}, (in)correct handling of objects⁹⁰ and associations between voices and faces⁹¹. Interest in more complex statistical learning emerged initially in the language domain when researchers showed that infants can use statistical properties in the language input to detect a wide range of phenomena^{92–94}. However, statistical learning is not limited to language learning; infants pick up statistical regularities in, for example, visual stimuli^{95,96}, auditory stimuli⁹⁵ and in action sequences^{97,98} as well. Probabilistic relations have also been shown to affect infants' attention⁹⁹. In addition to cognitive domains, statistical learning has been suggested as a basis for social learning as well¹⁰⁰.

So far, a major difference between statistical learning in infants and in machines resides in the multi- and cross-modal nature of infants' learning processes, as contrasted to the predominantly unimodal computational setting. In everyday life, infants encounter many situations in which they learn and integrate signals from multiple modalities (Box 2). Many studies have shown that infants are sensitive to the statistical relations between, for example, visual and auditory (linguistic and non-linguistic) stimuli^{91,101}, visual and tactile stimuli^{102,103}, and auditory and tactile stimuli¹⁰⁴. These multimodal associations are thought to develop by means of a combination of brain maturation and multimodal experience, which allow for the detection of temporal synchrony (for example, ref. ¹⁰⁵). The importance of early experience was shown in individuals who were temporarily deaf before they received a cochlear implant:

Box 2 | Lesson 2 from infant learning: richer representations through multimodal input

Infants learn statistical relations across diverse multimodal input streams and the resulting representations probably benefit from these richer sources of information. Multimodal approaches have also successfully been pursued in ML for decades. However, these successes have not caused a widespread shift from unimodal to multimodal training of DNNs. The recent surge in interest in multimodal ML, with methods such as contrastive learning in a multiview setting, might trigger the wider adoption of multimodal representation learning more generally, even for unimodal tasks. Unsupervised ML will not only incrementally benefit from the use of richer datasets during training, but we also expect that without them it will not reach sufficient performance. Until now, unsupervised ML has closely followed the developments on the side of supervised ML, using similar benchmarks and relying on similar (usually relatively simple) datasets. In doing so — although, in comparison, already being handicapped by the lack of supervision — unsupervised ML has inherited the (especially data-related) limitations of supervised ML. To develop the next generation of unsupervised ML algorithms, it is time to let go of the supervised ML traditions and make use of the additional possibilities that the rich and varied world of unlabeled data has to offer.

They showed decreased audiotactile¹⁰⁶ and audiovisual¹⁰⁷ integration, even after their hearing had been restored.

Infants' processing of signals from diverse inputs probably leads to improved representations and task performance. Multimodal information can support the disambiguation of conflicting or seemingly incoherent input otherwise obtained from a single sensory stream (for example, ref. ¹⁰⁸). It further enables the performance of tasks for which a single type of input is not sufficient. Even representations that seem related to one sensory domain benefit from input from other modalities. Reference ¹⁰⁹ has shown that spatial representations (for example, distance) for both auditory and proprioceptive stimuli are impaired in congenitally blind children and adults, suggesting that visual input is important for these non-visual representations. This work hints at a much greater need for learning from diverse multimodal inputs than one might intuitively consider necessary for unimodal tasks. Moreover, multimodal inputs are thought to lead to richer conceptual representations¹¹⁰, an important aspect of human intelligence.

Some ML researchers are already actively making use of the advantages that multi- or cross-modal information processing offers. Multimodal processing has been exploited, for instance, in robotics in object categorization tasks^{111,112}, and at the intersection between the vision and the language domain¹¹³, including in emotion recognition systems^{114,115} and in movie summarization tasks¹¹⁶. Regarding unsupervised learning, early successes were achieved in the late 1990s, for example, by performing category learning through multimodal sensing¹¹⁷. Multimodality has also moved into the focus of some researchers in the DNN community, spanning a range of application scenarios from image synthesis¹¹⁸ to unsupervised robot perception¹¹⁹ and image captioning¹²⁰. However, although the advantages of multimodal feature learning were identified almost a decade ago^{121,122}, most current DNNs are still being trained on unimodal data. The recent surge in interest in contrastive learning (for example, refs. ^{15,123}) suggests that this might be about to change. It is worth noting that state-of-the-art contributions such as contrastive learning¹²⁴ make explicit reference to the structure and performance of human multimodal information processing as inspiration and motivation for the approach. Similarly, ref. ¹²⁵ takes inspiration from infant learning when leveraging manifold alignment^{126,127} for building and aligning conceptual systems across modalities. The examples presented in a recent review of (mostly supervised) multimodal ML research¹²⁸ show that progress has been made in audio–visual integration, but that other types of sensory information are still lacking. In contrast, infants receive a continuous stream of multimodal input, which does not stop at pictures or movies and audio or text. Exploration and learning happen through taste, smell, sensations of temperature and hunger, proprioception and motion, touch and so on. Because the infant has the benefit of having a body through which it can sense different types of input and supervised ML algorithms have the benefit of labels, the challenge for unsupervised ML will be to generate the same richness of information. We believe that multimodal ML approaches have achieved good results so far (for example, refs. ^{129,130}) and should become the standard for many tasks that require unsupervised learning solutions. We suggest that next generations of unsupervised ML will need to include richer and different modalities.

Babies' input is shaped by development and active learning.

Infants' input is shaped both by their development (that is, their change over time) as well as by their active learning capacities (Box 3). The type of input infants receive is critically dependent on what they can do at any given point in time. In utero, fetuses have limited visual experience. Notwithstanding, they can already learn from waves spontaneously generated in the retina¹³¹. Sound from the environment can be heard (although being low-pass filtered), and newborns can recognize properties of their native language¹³²

Box 3 | Lesson 3 from infant learning: learning through curricula and curiosity

The infant cortex changes its learning mechanisms and architecture during maturation. Most ANNs are fixed in architecture, and those that change during training to a certain extent (such as refs. ^{1,171}) have far simpler trajectories than the infant cortex. We hypothesize that the changing infant architecture shapes the development of the representations that allow adults to generalize so effectively. Critical periods might be transferred to ANNs by instantiating temporally limited windows of elevated learning rate in targeted parts of a network, freezing aspects of the representation. A bio-mimetic subplate would exist just for the first part of training, and could comprise a self-organizing map that induces ordering in the primary network that would in turn grow as the subplate disappears. Infants' experience also changes dramatically with development due to their changing senses and body, the 'natural curriculum' provided by their caregivers, and the way infants curate their training set through curiosity. Training sets for ANNs are typically stationary, and infant-like changes might alter representations in a way that improves generalization. For example, the bias towards local features rather than global form in existing visual recognition networks²⁰⁸ might be reduced if networks had an initially blurry visual experience, like newborn infants, and make recognition more robust to noise.

and even recognize passages they heard in utero¹³³. These first experiences shape infants' later learning and biases. A newborn infant will primarily see whatever their caregivers bring into their visual field, a crawling infant will get extensive visual input of the floor, and a sitting infant will be able to see as far into the distance as a walking infant, but does not get the experience of optical flow while seated. The result is that the level and possibilities of exploration of the environment change when infants transition from one motor ability to the next. For example, differences in posture affect infants' possibilities for object exploration^{134,135}. Furthermore, the onset of locomotion makes infants aware of height and space¹³⁶. These types of finding show how much a developmental step in one domain can affect other domains of cognition, and how, similar to the influence of embodied/grounded cognition on mental representations in adults¹³⁷, infants' physical developmental stages affect their mental representations. Sensory processing goes through developmental changes too. For example, rather than being born with full visual capacities, newborns' visual acuity is low and gradually increases in the first six months of life^{138,139}. These changes in sensory and motor development not only allow infants to explore different aspects of their environment — they also drive an expansion of the range of obtained inputs in the direction of increasingly varied stimuli occurring in increasingly complex combinations, thereby introducing a phased structure in infants' learning experiences. Recent experimentation with head-mounted cameras (for example, refs. ^{140–143}) provides further evidence for phased input and for the natural curriculum that is generated by child development¹⁴⁴. For example, ref. ¹⁴⁵ showed that, in approximately a quarter of 1- to 24-month-old infants' visual input, there is another person present (as measured by the presence of at least a face or a hand in the infant's visual field). Importantly, the frequency of faces versus hands significantly changed as a function of age: the proportion of faces in view declined with age while the proportion of hands increased.

In addition to this natural curriculum generated by physical and cognitive development, the infants' learning process is guided and supported by other agents (such as their caregivers) in their environment. Decades of research have shown that infants learn from observation and imitation (for example, refs. ^{146–150}).

Furthermore, caregivers facilitate language and action learning through infant-directed speech (for example, ref. ¹⁵¹) and motionese^{152,153}. Caregivers also increase infants' attentiveness to objects during joint play¹⁵⁴ and are responsive to where infants are looking¹⁵⁵. Parental responsiveness and contingent reactions to an infant's actions means that scaffolding and guided play are both interactive as well as dynamic¹⁵⁶, ensuring maximal support to the infant's learning process. Moreover, infants have been shown to use the knowledge and expertise of other agents when they need information or support^{157,158}.

Development thus shapes the inputs that infants might have access to, and caregivers can facilitate this curriculum. At the same time, infant learning does not just happen passively based on whatever input happens to be in their environment. Infants play an active role in directing their attention to stimuli from which they learn. This process is known as curiosity-driven or active learning. Curiosity is taken to be a state of arousal that requires actions to modulate the aroused state, with the degree of novelty determining infants' ability to learn¹⁵⁹. The spectrum of arousal is subdivided into relaxation (insufficient arousal), curiosity (optimal for learning) and anxiety (too much arousal)¹⁶⁰, with relaxation and anxiety considered to create little opportunity for learning. Empirical research shows that infants attend significantly longer to stimuli that are at an intermediate level of complexity, a finding that was dubbed 'the Goldilocks effect'^{161,162}. Moreover, ref. ¹⁶³ showed that infants allocate their attention as a function of the learning progress a stimulus offers. Infants have been shown to use violations of expectations as indicators of learning opportunities and to preferably explore surprising events⁶³. Related to this U-shaped curve of attention, infants vary in their preference for familiar or novel stimuli. This is thought to be dependent on the degree of encoding of the stimuli^{164,165}. If the encoding is not yet complete, infants will show a familiarity preference. Once encoding is complete, they exhibit a novelty preference. These ideas have led to a plethora of looking time studies. Although the richness of the interpretation of looking time studies can be questioned (for example, refs. ^{166,167}), the theories on active learning and methods to investigate stimulus encoding have provided important insights into the nature of infants' learning mechanisms. Infants' drive to explore can be seen in their motor behaviour as well. Infants travel big distances, with ~2,400 steps per waking hour in the first half of their second year¹⁶⁸. Their walking bouts tend to have a distinguishable destination for only a small proportion, giving rise to the idea that movement 'is enough motivation for infants to get up and go'¹⁶⁹. Moreover, active exploration of objects has been shown to increase learning about those objects compared to passive observation¹⁷⁰.

The general idea of applying a structured learning curriculum — as naturally experienced by infants during development — to artificial systems has been put forward by ref. ¹⁷¹ and was prominently addressed in the deep learning literature¹⁷². Improved performance and generalization of CNNs with low initial visual acuity — obtained by starting the network training with blurred rather than with high-resolution images — corroborates the idea that phased sensory input indeed improves learning¹⁷³. Moreover, using rich head-mounted camera data from infants as training data for unsupervised machine learning can lead to the emergence of powerful high-level visual representations¹⁷⁴. One of the main challenges of providing a structured training scheme for neural networks is the high sensitivity of the curriculum's effectiveness to the mode of progression through the learning tasks, that is, the syllabus. Two of the main factors influencing this sensitivity are critical learning periods and catastrophic interference.

Critical periods — which are well documented in biological learners — reflect moments of peak plasticity during a specific developmental state (often early in life) that are followed by reduced plasticity (see, for example, ref. ¹⁷⁵ for review and reflection). Around

the time of birth, infants are undergoing a striking change in brain architecture. In adults, the cortex receives input from the senses and elsewhere via the white matter tracts — the ‘wiring’ of the brain. Early in development, however, a structure called the subplate separates the white matter tracts from the cortex, and mediates all connections. The subplate is thought to play a role in organizing the cortex and is, for example, responsible for the spatial maps in the visual cortex and frequency maps in the auditory cortex¹⁷⁶. In the middle of gestation, the subplate is thicker than the cortex, yet it is gone by one year of age. It evolved recently and is largest in humans, suggesting it may be particularly important for complex brains¹⁷⁷. Critical periods can be contrasted to open-ended learning systems or systems in which plasticity increases with maturation or experience¹⁷⁵. Reference ¹⁷⁸ shows that DNNs also exhibit critical periods during which a temporary stimulus deficit can impair the future performance of the network, even to an unrecoverable degree: the initial learning transient plays a key role in determining the outcome of the training process, and shortcomings or biases, for example, in the variety of input samples, during early training may not be recovered during the remainder of the training process.

As a consequence of the stability/plasticity dilemma¹⁷⁹, ANNs can suffer from catastrophic interference — a process whereby new knowledge overwrites rather than integrates previous knowledge¹⁸⁰. Different solutions for this general problem have been proposed¹⁸¹, including rehearsal and pseudo-rehearsal learning¹⁸², the use of pairs of (slowly changing) plastic network weights storing long(er)-term knowledge and (fast-changing) elastic weights encoding a temporary context that can be used to efficiently ‘deblur’ previous learning outcomes¹⁸³, and brain-inspired approaches suggesting the use of dual-memory architectures¹⁸⁴ or building on synaptic consolidation¹⁸⁵. Regarding unsupervised learning in particular, catastrophic interference has, among others, been addressed in the context of continuous and lifelong learning, including the use of undercomplete autoencoders trained for feature transfer across tasks¹⁸⁶ or approaches motivated by results from neuroscience such as neurogenesis deep learning¹⁸⁷. Relatedly, the explicit (meta-)learning of representations for continual learning that avoid catastrophic interference has been proposed¹⁸⁸. Still, although presenting a major hurdle on the road to further advanced ML systems, as of now the problem of catastrophic interference remains unsolved¹⁸⁹. To overcome this challenge, we suggest another closer look at the brain and its development during infancy, taking inspiration in the temporal evolution of network characteristics. In today’s practice, much focus is put on exploring different learning rates, and less effort is dedicated to better understanding and eventually exploiting the impact of altering the structure of neural networks (for example, the number and type of neurons) over time.

Turning to active learning, the recognition that an algorithm may learn better and more efficiently if it is allowed to select the data from which it learns (for example, refs. ^{190,191}) is not new. Over the years, different types of curiosity mechanism have been proposed for artificial systems. Some researchers suggest that curiosity could be prediction-based, causing agents to attend to input for which predictability is minimal¹⁹² or maximal¹⁹³. In the context of curriculum learning, ref. ¹⁹⁴ proposed a multi-armed bandit-based approach to finding progress-maximizing stochastic policies over different learning tasks. More closely related to the findings in developmental science, ref. ¹⁹⁵ argued that curiosity-driven learning occurs most optimally when the agent seeks out information as a function of its compressibility. Furthermore, it has been suggested that active learning is driven by a goal to maximize learning progress by interacting with the environment in a novel manner^{196,197}. Supporting this line of thought, computational modelling approaches that compared presenting stimuli in a fixed order or allowing the model to choose its own input showed that maximal learning happens when the model can maximize stimulus novelty relative to its internal

states¹⁹⁸. This work emphasized the importance of the interaction between the structure of the environment and the previously acquired knowledge of the learner. Similarly, ref. ¹⁹⁹ created an agent with a world model that learned to predict the consequences of the agent’s actions, and a meta-cognitive self-model that tracks and adversarially challenges the performance of the world model. This caused the agent to autonomously explore novel interactions with the environment, leading to new behaviours and improved learning.

Discussion

We have presented three insights from developmental science that have the potential to make a fundamental difference to ML through the learning of more flexible and efficient representations. We have focused especially on infant learning in the first year of life as their learning process also requires them to learn useful representations from unlabeled data. Throughout the Perspective, we have spelled out to what extent these three components of infant learning are already mirrored in ML algorithms, and where further steps can reasonably be made. Improving the quality, flexibility and efficiency of learned representations will directly translate to improved ML performance.

Our review has highlighted the considerable differences between the prevailing practices in ML and infant learning. In ML, it is common to remove targeted interference in the learning process as much as possible and leave everything to be learned to the data itself¹, which stands in stark contrast to infant learning. Infants’ input has been found to be optimized for learning about specific features of the input. When comparing the three lessons on infant learning to current approaches in ML, the following overarching insights can be extracted (Fig. 2):

- (1) There is more initial structure to constrain and guide infants’ learning processes.
- (2) Infants’ learning opportunities are richer and more flexible.
- (3) Infant learning is shaped by an interaction of environmental, developmental and intrinsic factors.

Factors such as innate biases, saliency, curiosity and development over time all play an important role in shaping infants’ learning curriculum and contribute to the speed and flexibility with which infants learn. Reflecting these insights back into ML, they cast substantial doubt on the assumption that ‘the data will fix it’ is indeed the most efficient and effective approach to training neural networks.

Regarding the implications for current architectural paradigms in ML, we are convinced that complex cognition arises from a ‘system of systems’ consisting of interacting but heterogeneous components. Although monolithic models might in principle be able to learn any computable function and subsequently exhibit the corresponding behaviour, and although further increases in network depth, quantity of training data and available computing power might gradually boost the performance of such models further, we believe that, to reach complex behaviour, a qualitatively different approach is required. Thinking about the search for the right architecture, we want to explicitly mention three — and, in our opinion, very promising — directions.

First, the discussed ‘innate architectures’ could themselves be the result of self-organization throughout development. Regarding biological inspiration for this line of thought, in our discussion of different learning inputs and the potential role(s) these play, we mention, for example, the spontaneous waves of activity across the retina in utero that propagate to the primary visual cortex V1 and help wire up a retinotopic visual cortex long before visual stimulation from the outside world is ever received. Although there has been some work on evolved plastic ANNs²⁰⁰, more generally the idea of multiple timescales of self-organization constraining first the connectivity and then the ‘weights’ is a concept that in our

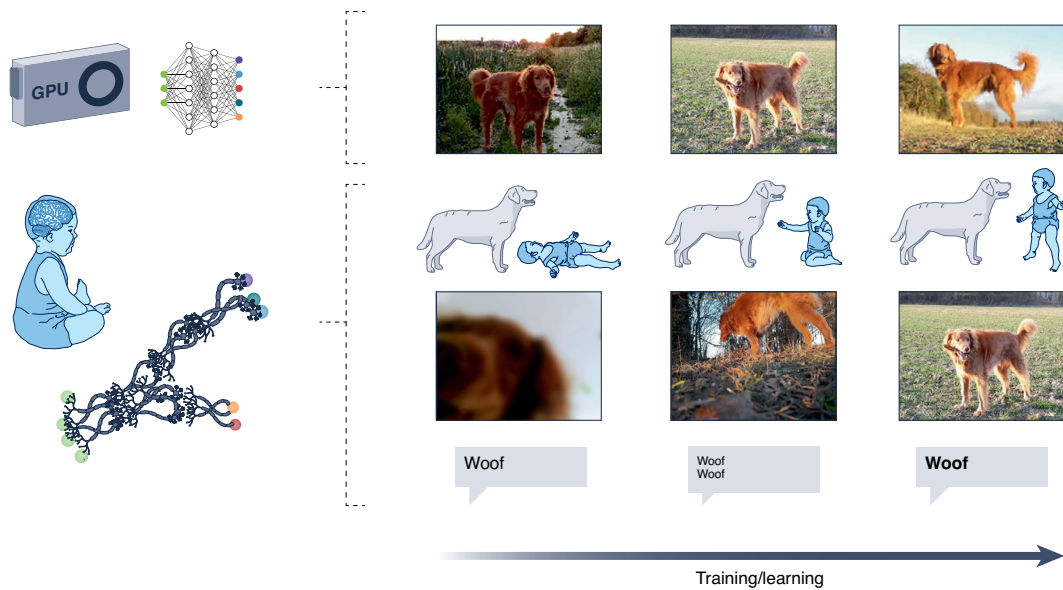


Fig. 2 | Typical ANN training versus infant development. Compared to the ANN, the neural architecture of the infant has a more complex initial structure, which provides an inductive bias that guides learning. The infant's inputs are multimodal (only visual and auditory are depicted here, but infants could touch and smell the dog too) and variable (not every dog looks or moves the same, not every woof sounds the same), allowing for a richer representation and generalization, while the ANN learns from pictures only, which are frequently of a stereotypical composition. The infant's inputs are structured by their embodied development: they see things from different perspectives depending on their motor development (lying down, sitting up, walking) and their bodily changes (their vision becomes gradually less blurry, for example). The ANN's input does not have this variety or the non-stationarity of experiences seen in a human developmental trajectory.

opinion has not yet received sufficient consideration in ML research (as also supported by our discussions relating to hyperparameters and networks-of-networks). The potential implications are far-reaching, though, as this general approach would alleviate the need to figure out and hand-build the whole complex interconnected architecture; instead, it would suffice to identify the correct rules and stimuli for self-organization.

Second, multimodality in infant learning occurs in at least two ways. As discussed, infants experience a continuous stream of multimodal input, allowing them to learn rich representations. At the same time, presumably they are also often learning by several objectives at once, for example, aiming to predict their visual input at the next moment from their current visual experience combined with their motor actions, tactile sensations and auditory data. The usefulness of this multi-objective approach varies, however. Several sensory modalities probably help with the prediction of each individual one, but the actual contributions of each to the others are in most cases unequal and may vary greatly by situation. For example, when we imagine a child shaking a rattle, in that case motor signals, tactile sensation, audition and vision all are informative for predicting one another. This is probably different when watching a leaf fall silently to the ground — then vision, touch and audition are likely to be uncorrelated. Additionally, the specifics of the input change with development (both in terms of the type of input as well as the features of that input (for example, there are big differences between visual input in the womb, visual input in the first postnatal months and visual input later in development), which will affect in what ways multimodality influences individual modalities. The ML community has started to look into multimodal, multi-objective learning (see, for example, refs. ^{201–203} for recent contributions). Zooming in on a particular example, ref. ¹²⁹ takes inspiration in research into learning by expectation²⁰⁴ and presents an unsupervised learning model for stimulus prediction across modalities using multimodal bindings to enhance unimodal perception. (In light of our suggestion regarding systems-of-systems, it is worth noting that the applied

model in ref. ¹²⁹ implements a hybrid architecture combining neural autoencoder networks for each of the unimodal channels with an extended recurrent, self-organizing Gamma-GWR network²⁰⁵). Still, as was the case with the previous point, we believe that this topic deserves a marked increase in attention and might hold the key to several critical advances, especially regarding unsupervised ML.

Third, when considering neural adaptation and comparing between biological and artificial neural networks, it has to be noted that although ML has studied notions of plasticity under the conceptual headline of 'lifelong learning', little attention has been given to short- and mid-term adaptations. In the brain, neurons change their response properties constantly to keep their regions of maximal response sensitivity matched to the properties of the sorts of stimuli they recently encountered. In what precise way an ANN would profit from similar features is an open research question, but we believe that understanding the consequences of this salient difference between biological and artificial networks will be a worthwhile undertaking.

Neither the developmental nor the ML research presented in this Perspective is exhaustive. For example, one important aspect that we have only touched briefly upon is the role of other agents. Other agents in infants' everyday life provide essential scaffolds to their learning process. In unsupervised ML, parental scaffolding has not been widely adopted, but the successes gained by training networks first on auxiliary tasks before moving onto the target task suggests that this is a space worth exploring. As concluded in a review on computationally modelling infant learning: "[...] to understand [the] power of children's learning, it is important to investigate it in a social context" (p. 98 in ref. ²⁰⁶). This does not have to take the form of a teacher–student set-up, which would introduce some levels of supervision. Instead, one could consider an approach similar to an expanded OpenAI Gym²⁰⁷. Such a framework could simultaneously incorporate aspects of curriculum learning (for example, by enabling different syllabi composed of sequences of benchmarks) and aspects of active learning (for example, by allowing agents to

optimize their path through learning space by choosing from what they are learning). It is outside of the scope of this Perspective to go into these topics more deeply. Nonetheless, we hope to have provided a representative sample of previous work on the topics we have addressed. Importantly, the three lessons presented here were chosen based on their potential to qualitatively improve the next generation of unsupervised ML algorithms as well as on their integration with current ML implementations. By focusing on this intersection, we aim to increase the likelihood that these lessons can be meaningfully considered in theory and implementations.

The argument to take inspiration from human (infant) learning that we and others have made rests on the observation that human learning leads to robust representations that can be flexibly used in various tasks with an acceptable to excellent level of performance across the board. Clearly, some scepticism is warranted as to whether adopting insights from infant learning will be equally valuable for all ML purposes. It is possible that some of the lessons provide an advantage across domains, whereas others might turn out to be particularly beneficial for specific tasks. Taking this Perspective as a conceptual anchor, future research will explore the exact interactions of each of the given insights from infant learning with their counterpart(s) in ML.

Received: 3 June 2021; Accepted: 8 April 2022;

Published online: 22 June 2022

References

- Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. Greedy layer-wise training of deep networks. In *Proc. Advances in Neural Information Processing Systems* Vol. 19 (eds. Schölkopf, B., Platt, J. & Hoffman, T.) 153–160 (NIPS, 2006).
- Hinton, G. E., Osindero, S. & Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **18**, 1527–1554 (2006).
- Baldi, P. Autoencoders, unsupervised learning and deep architectures. In *Proc. ICML Workshop on Unsupervised and Transfer Learning* (eds. Guyon, I., Dror, G., Lemaire, V., Taylor, G. & Silver, D.) 37–49 (JMLR, 2012).
- Bengio, Y., Courville, A. & Vincent, P. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798–1828 (2013).
- Erhan, D. et al. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.* **11**, 625–660 (2010).
- Russakovsky, O. et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vision* **115**, 211–252 (2015).
- Carreira, J. & Zisserman, A. Quo vadis, action recognition? A new model and the kinetics dataset. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 6299–6308 (IEEE, 2017).
- Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 3431–3440 (IEEE, 2015).
- Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems* Vol. 28 (eds. Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. & Garnett, R.) 91–99 (NIPS, 2015).
- He, K., Girshick, R. & Dollár, P. Rethinking ImageNet pre-training. In *Proc. IEEE International Conference on Computer Vision* 4918–4927 (IEEE, 2019).
- Huh, M., Agrawal, P. & Efros, A. A. What makes ImageNet good for transfer learning? Preprint at <https://arxiv.org/abs/1608.08614> (2016).
- Recht, B., Roelofs, R., Schmidt, L. & Shankar, V. Do ImageNet classifiers generalize to imagenet? In *Proc. 36th International Conference on Machine Learning* (eds. Chaudhuri, K. & Salakhutdinov, R.) 5389–5400 (PMLR, 2019).
- Burgess, C. P. et al. Understanding disentangling in β -VAE. Preprint at <https://arxiv.org/abs/1804.03599> (2018).
- Caron, M., Bojanowski, P., Joulin, A. & Douze, M. Deep clustering for unsupervised learning of visual features. In *Proc. European Conference on Computer Vision* (eds. Ferrari, V., Hebert, M. I., Sminchisescu, C. & Weiss, Y.) 132–149 (Springer, 2018).
- Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. In *Proc. 37th International Conference on Machine Learning* (eds. Daumé, H. III & Singh, A.) 1597–1607 (PMLR, 2020).
- Zbontar, J., Jing, L., Misra, I., LeCun, Y. & Deny, S. Barlow twins: self-supervised learning via redundancy reduction. In *Proc. 38th International Conference on Machine Learning* (eds. Meila, M. & Zhang, T.) 12310–12320 (PMLR, 2021).
- Ma, W. J. & Peters, B. A neural network walks into a lab: towards using deep nets as models for human behavior. Preprint at <https://arxiv.org/abs/2005.02181> (2020).
- Yamins, D. L. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19**, 356–365 (2016).
- Macpherson, T. et al. Natural and artificial intelligence: a brief introduction to the interplay between AI and neuroscience research. *Neural Networks* **144**, 603–613 (2021).
- Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386–408 (1958).
- McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).
- Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick, M. Neuroscience-inspired artificial intelligence. *Neuron* **95**, 245–258 (2017).
- 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).
- Saxe, A., Nelli, S. & Summerfield, C. If deep learning is the answer, then what is the question? *Nat. Rev. Neurosci.* **22**, 55–67 (2021).
- Sinz, F. H., Pitkow, X., Reimer, J., Bethge, M. & Tolias, A. S. Engineering a less artificial intelligence. *Neuron* **103**, 967–979 (2019).
- Zador, A. M. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* **10**, 3770 (2019).
- Bengio, Y. Deep learning of representations for unsupervised and transfer learning. In *Proc. ICML Workshop on Unsupervised and Transfer Learning* (eds. Guyon, I., Dror, G., Lemaire, V., Taylor, G. & Silver, D.) 17–36 (JMLR, 2012).
- Cangelosi, A. & Schlesinger, M. *Developmental Robotics: From Babies to Robots* (MIT Press, 2015).
- Kidd, C. How to know. In *Proc. 33rd Conference on Neural Information Processing Systems* (NIPS, 2019).
- Gopnik, A. An AI that knows the world like children do. *Sci. Am. Mind* **28**, 21–28 (2017).
- Kosoy, E. et al. Exploring exploration: comparing children with RL agents in unified environments. Preprint at <https://arxiv.org/abs/2005.02880> (2020).
- Smith, L. B. & Breazeal, C. The dynamic lift of developmental process. *Dev. Sci.* **10**, 61–68 (2007).
- Smith, L. B. & Slone, L. K. A developmental approach to machine learning? *Front. Psychol.* **8**, 2124 (2017).
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. How to grow a mind: statistics, structure and abstraction. *Science* **331**, 1279–1285 (2011).
- Adolph, K. E., Hoch, J. E. & Cole, W. G. Development (of walking): 15 suggestions. *Trends Cogn. Sci.* **22**, 699–711 (2018).
- Byrge, L., Sporns, O. & Smith, L. B. Developmental process emerges from extended brain-body-behavior networks. *Trends Cogn. Sci.* **18**, 395–403 (2014).
- Hunnus, S. Early cognitive development: five lessons from infant learning. In *Oxford Research Encyclopedia of Psychology* (ed. Braddick, O.) (Oxford Univ. Press, in the press).
- Karmiloff-Smith, A. An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis. *AIMS Neurosci.* **2**, 91–104 (2015).
- von Hofsten, C. & Rosander, K. The development of sensorimotor intelligence in infants. *Adv. Child Dev. Behav.* **55**, 73–106 (2018).
- Dunsworth, H. M., Warrener, A. G., Deacon, T., Ellison, P. T. & Pontzer, H. Metabolic hypothesis for human altriciality. *Proc. Natl Acad. Sci. USA* **109**, 15212–15216 (2012).
- Haeusler, M. et al. The obstetrical dilemma hypothesis: there's life in the old dog yet. *Biol. Rev.* **96**, 2031–2057 (2021).
- Bethlehem, R. A. et al. Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
- Huttenlocher, P. R. et al. Synaptic density in human frontal cortex—developmental changes and effects of aging. *Brain Res.* **163**, 195–205 (1979).
- Deoni, S. C. et al. Mapping infant brain myelination with magnetic resonance imaging. *J. Neurosci.* **31**, 784–791 (2011).
- Hill, J. et al. A surface-based analysis of hemispheric asymmetries and folding of cerebral cortex in term-born human infants. *J. Neurosci.* **30**, 2268–2276 (2010).
- Clouchoux, C. et al. Quantitative in vivo MRI measurement of cortical development in the fetus. *Brain Struct. Funct.* **217**, 127–139 (2012).
- Cabral, L., Zubiaurre, L., Wild, C., Linke, A. & Cusack, R. Category-selective visual regions have distinctive signatures of connectivity in early infancy. Preprint at <https://www.biorxiv.org/content/10.1101/675421v2.full> (2019).
- Doria, V. et al. Emergence of resting state networks in the preterm human brain. *Proc. Natl Acad. Sci. USA* **107**, 20015–20020 (2010).
- Kamps, F. S., Hendrix, C. L., Brennan, P. A. & Dilks, D. D. Connectivity at the origins of domain specificity in the cortical face and place networks. *Proc. Natl Acad. Sci. USA* **117**, 6163–6169 (2020).

50. Cusack, R., Wild, C. J., Zubiurre-Elorza, L. & Linke, A. C. Why does language not emerge until the second year? *Hearing Res.* **366**, 75–81 (2018).
51. Deen, B. et al. Organization of high-level visual cortex in human infants. *Nat. Commun.* **8**, 13995 (2017).
52. Ellis, C. T. et al. Evidence of hippocampal learning in human infants. *Curr. Biol.* **31**, 3358–3364 (2021).
53. Ellis, C. T., Skalaban, L. J., Yates, T. S. & Turk-Browne, N. B. Attention recruits frontal cortex in human infants. *Proc. Natl Acad. Sci. USA* **118**, e2021474118 (2021).
54. Raz, G. & Saxe, R. Learning in infancy is active, endogenously motivated, and depends on the prefrontal cortices. *Annu. Rev. Dev. Psychol.* **2**, 247–268 (2020).
55. Linke, A. C. et al. Disruption to functional networks in neonates with perinatal brain injury predicts motor skills at 8 months. *NeuroImage Clin.* **18**, 399–406 (2018).
56. Marcus, G. F., Vijayan, S., Rao, S. B. & Vishton, P. M. Rule learning by seven-month-old infants. *Science* **283**, 77–80 (1999).
57. Elman, J. L. Finding structure in time. *Cogn. Sci.* **14**, 179–211 (1990).
58. Alhama, R. G. & Zuidema, W. Pre-wiring and pre-training: what does a neural network need to learn truly general identity rules? *J. Artif. Intell. Res.* **61**, 927–946 (2018).
59. Jeffress, L. A. A place theory of sound localization. *J. Comp. Physiol. Psychol.* **41**, 35–39 (1948).
60. Jaeger, H. *The ‘Echo State’ Approach to Analysing and Training Recurrent Neural Networks—With an Erratum Note*. Technical Report 148, 13 (German National Research Center for Information Technology (GMD), 2001).
61. Smith, L. B. Do infants possess innate knowledge structures? The con side. *Dev. Sci.* **2**, 133–144 (1999).
62. Spelke, E. Initial knowledge: six suggestions. *Cognition* **50**, 431–445 (1995).
63. Stahl, A. E. & Feigenson, L. Observing the unexpected enhances infants’ learning and exploration. *Science* **348**, 91–94 (2015).
64. Simion, F., Di Giorgio, E., Leo, I. & Bardi, L. The processing of social stimuli in early infancy: from faces to biological motion perception. In *Progress in Brain Research* Vol. 189 (eds. Braddick, O., Atkinson, J. & Innocenti, G. M.) 173–193 (Elsevier, 2011).
65. Reynolds, G. D. & Roth, K. C. The development of attentional biases for faces in infancy: a developmental systems perspective. *Front. Psychol.* **9**, 222 (2018).
66. Viola Macchi, C., Turati, C. & Simion, F. Can a nonspecific bias toward top-heavy patterns explain newborns’ face preference? *Psychol. Sci.* **15**, 379–383 (2004).
67. Chien, S. H.-L. No more top-heavy bias: Infants and adults prefer upright faces but not top-heavy geometric or face-like patterns. *J. Vision* **11**, 13 (2011).
68. Ichikawa, H., Tsuruhara, A., Kanazawa, S. & Yamaguchi, M. K. Two- to three-month-old infants prefer moving face patterns to moving top-heavy patterns. *Jap. Psychol. Res.* **55**, 254–263 (2013).
69. Cooper, R. P. & Aslin, R. N. Preference for infant-directed speech in the first month after birth. *Child Dev.* **61**, 1584–1595 (1990).
70. Peña, M. et al. Sounds and silence: an optical topography study of language recognition at birth. *Proc. Natl Acad. Sci. USA* **100**, 11702–11705 (2003).
71. Vouloumanos, A. & Werker, J. F. Listening to language at birth: evidence for a bias for speech in neonates. *Dev. Sci.* **10**, 159–164 (2007).
72. Mély, D. A., Linsley, D. & Serre, T. Complementary surrounds explain diverse contextual phenomena across visual modalities. *Psychol. Rev.* **125**, 769 (2018).
73. Linsley, D., Kim, J., Ashok, A. & Serre, T. Recurrent neural circuits for contour detection. In *Proc. 8th International Conference on Learning Representations (ICLR, 2020)*.
74. Michalski, R. S. in *Machine Learning* 83–134 (Morgan Kaufmann, 1983).
75. Mitchell, T. *The Need for Biases in Learning Generalizations*. Rutgers Computer Science Technical Report cbm-tr-117 (Rutgers University, 1980).
76. Feinman, R. & Lake, B. M. Learning inductive biases with simple neural networks. In *Proc. 40th Annual Meeting of the Cognitive Science Society* (eds. Kalish, C., Rau, M. A., Zhu, X. & Rogers, T. T.) (CSS, 2018).
77. Koppert, R. M. & Weyde, T. Weight priors for learning identity relations. In *Advances in Neural Information Processing Systems* Vol. 33 (eds. Wallach, H. M., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F. E. Fox, A. & Garnett, R.) (NIPS, 2020).
78. Weyde, T. & Koppert, R. M. Modelling identity rules with neural networks. *J. Appl. Logics* **6**, 745–769 (2019).
79. Ullman, S., Harari, D. & Dorfman, N. From simple innate biases to complex visual concepts. *Proc. Natl Acad. Sci. USA* **109**, 18215–18220 (2012).
80. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
81. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012).
82. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
83. Szegedy, C. et al. Going deeper with convolutions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 1–9 (IEEE, 2015).
84. Besold, T. R. et al. Neural-symbolic learning and reasoning: a survey and interpretation. In *Neuro-Symbolic Artificial Intelligence: The State of the Art* (eds. Hitzler, P. & Sarker, M. K.) 1–51 (IOS Press, 2021).
85. d’Avila Garcez, A. S. & Gabbay, D. M. Fibring neural networks. In *Proc. Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence* (eds. McGuinness, D. L. & Ferguson, G.) 342–347 (AAAI Press/MIT Press, 2004).
86. Saffran, J. R. & Kirkham, N. Z. Infant statistical learning. *Annu. Rev. Psychol.* **69**, 181–203 (2018).
87. Teinonen, T., Fellman, V., Näätänen, R., Alku, P. & Huotilainen, M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* **10**, 21 (2009).
88. Jacquey, L., Fagard, J., Esseily, R. & O’Regan, J. K. Detection of sensorimotor contingencies in infants before the age of one year: a comprehensive review. *Dev. Psychol.* **56**, 1233–1251 (2020).
89. Zaadnoordijk, L. et al. From movement to action: an EEG study into the emerging sense of agency in early infancy. *Dev. Cogn. Neurosci.* **42**, 100760 (2020).
90. Hunnius, S. & Bekkering, H. The early development of object knowledge: a study of infants’ visual anticipations during action observation. *Dev. Psychol.* **46**, 446–454 (2010).
91. Brookes, H. et al. Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. *Infant Child Dev.* **10**, 75–82 (2001).
92. Gómez, R. & Maye, J. The developmental trajectory of nonadjacent dependency learning. *Infancy* **7**, 183–206 (2005).
93. Maye, J., Werker, J. F. & Gerken, L. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82**, B101–B111 (2002).
94. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
95. Emberson, L. L., Misyak, J. B., Schwade, J. A., Christiansen, M. H. & Goldstein, M. H. Comparing statistical learning across perceptual modalities in infancy: an investigation of underlying learning mechanism(s). *Dev. Sci.* **22**, e12847 (2019).
96. Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* **83**, B35–B42 (2002).
97. Monroy, C. D. et al. Sensitivity to structure in action sequences: an infant event-related potential study. *Neuropsychologia* **126**, 92–101 (2019).
98. Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M. & Hirsh-Pasek, K. Infants segment continuous events using transitional probabilities. *Child Dev.* **85**, 1821–1826 (2014).
99. Tummelshammer, K. S. & Kirkham, N. Z. Learning to look: probabilistic variation and noise guide infants’ eye movements. *Dev. Sci.* **16**, 760–771 (2013).
100. Ruffman, T., Taumoepeau, M. & Perkins, C. Statistical learning as a basis for social understanding in children. *Br. J. Dev. Psychol.* **30**, 87–104 (2012).
101. Bristow, D. et al. Hearing faces: how the infant brain matches the face it sees with the speech it hears. *J. Cogn. Neurosci.* **21**, 905–921 (2008).
102. Bremner, A. J., Mareschal, D., Lloyd-Fox, S. & Spence, C. Spatial localization of touch in the first year of life: early influence of a visual spatial code and the development of remapping across changes in limb position. *J. Exp. Psychol. Gen.* **137**, 149–162 (2008).
103. Zmyj, N., Jank, J., Schütz-Bosbach, S. & Daum, M. M. Detection of visual-tactile contingency in the first year after birth. *Cognition* **120**, 82–89 (2011).
104. Tanaka, Y., Kanakogi, Y., Kawasaki, M. & Myowa, M. The integration of audio-tactile information is modulated by multimodal social interaction with physical contact in infancy. *Dev. Cogn. Neurosci.* **30**, 31–40 (2018).
105. Lewkowicz, D. J. The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychol. Bull.* **126**, 281–308 (2000).
106. Landry, S. P., Guillemot, J.-P. & Champoux, F. Temporary deafness can impair multisensory integration: a study of cochlear-implant users. *Psychol. Sci.* **24**, 1260–1268 (2013).
107. Stevenson, R., Sheffield, S. W., Butera, I. M., Gifford, R. H. & Wallace, M. Multisensory integration in cochlear implant recipients. *Ear Hearing* **38**, 521–538 (2017).
108. Weatherhead, D. & White, K. S. Read my lips: visual speech influences word processing in infants. *Cognition* **160**, 103–109 (2017).
109. Cappagli, G., Cocchi, E. & Gori, M. Auditory and proprioceptive spatial impairments in blind children and adults. *Dev. Sci.* **20**, e12374 (2017).
110. Bruni, E., Tran, N.-K. & Baroni, M. Multimodal distributional semantics. *J. Artif. Intell. Res.* **49**, 1–47 (2014).

111. Marton, Z.-C., Pangercic, D., Blodow, N. & Beetz, M. Combined 2D-3D categorization and classification for multimodal perception systems. *Int. J. Robot. Res.* **30**, 1378–1402 (2011).
112. Nakamura, T., Nagai, T. & Iwahashi, N. Multimodal object categorization by a robot. In *Proc. 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems* 2415–2420 (IEEE, 2007).
113. Mogadala, A., Kalimuthu, M. & Klakow, D. Trends in integration of vision and language research: a survey of tasks, datasets, and methods. *J. Artif. Intell. Res.* **71**, 1183–1317 (2021).
114. Barbieri, F. et al. Towards a multimodal time-based empathy prediction system. In *Proc. 2019 14th IEEE International Conference on Automatic Face and Gesture Recognition* 1–5 (IEEE, 2019).
115. Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B. W. & Zafeiriou, S. End-to-end multimodal emotion recognition using deep neural networks. *IEEE J. Select. Top. Signal Process.* **11**, 1301–1309 (2017).
116. Evangelopoulos, G. et al. Multimodal saliency and fusion for movie summarization based on aural, visual and textual attention. *IEEE Trans. Multimedia* **15**, 1553–1568 (2013).
117. de Sa, V. R. & Ballard, D. H. Category learning through multimodality sensing. *Neural Comput.* **10**, 1097–1117 (1998).
118. Radford, A., Metz, L. & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In *Proc. 4th International Conference on Learning Representations* (eds. Bengio, Y. & LeCun, Y.) (ICLR, 2016).
119. Droniou, A., Ivaldi, S. & Sigaud, O. Deep unsupervised network for multimodal perception, representation and classification. *Robot. Auton. Syst.* **71**, 83–98 (2015).
120. Feng, Y., Ma, L., Liu, W. & Luo, J. Unsupervised image captioning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 4125–4134 (IEEE, 2019).
121. Ngiam, J. et al. Multimodal deep learning. In *Proc. 28th International Conference on International Conference on Machine Learning* (eds. Getoor, L. & Scheffer, T.) 689–696 (OmniPress, 2011).
122. Srivastava, N. & Salakhutdinov, R. R. Multimodal learning with deep Boltzmann machines. In *Advances in Neural Information Processing Systems* (Pereira, F., Burges, C. J., Bottou, L. & Weinberger, K. Q.) 2222–2230 (NIPS, 2012).
123. Bachman, P., Hjelm, R. D. & Buchwalter, W. Learning representations by maximizing mutual information across views. In *Advances in Neural Information Processing Systems* (eds. Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. & Garnett, R.) 15509–15519 (NIPS, 2019).
124. Tian, Y., Krishnan, D. & Isola, P. Contrastive multiview coding. In *Proc. Computer Vision—ECCV 2020: 16th European Conference Part XI* 16 (Vedaldi, A., Bischof, H., Brox, T. & Frahm, J.-M.) 776–794 (Springer, 2020).
125. Roads, B. D. & Love, B. C. Learning as the unsupervised alignment of conceptual systems. *Nat. Mach. Intell.* **2**, 76–82 (2020).
126. Wang, C. & Mahadevan, S. Manifold alignment without correspondence. In *Proc. 21st International Joint Conference on Artificial Intelligence* (ed. Boutilier, C.) 1273–1278 (Morgan Kaufmann, 2009).
127. Wang, C. & Mahadevan, S. Heterogeneous domain adaptation using manifold alignment. In *Proc. Twenty-Second International Joint Conference on Artificial Intelligence* Vol. 2 (ed. Walsh, T.) 1541–1546 (AAAI Press, 2011).
128. Baltrušaitis, T., Ahuja, C. & Morency, L.-P. Multimodal machine learning: a survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 423–443 (2018).
129. Barros, P., Eppe, M., Parisi, G. I., Liu, X. & Wermter, S. Expectation learning for stimulus prediction across modalities improves unsensory classification. *Front. Robot. AI* **6**, 137 (2019).
130. Peterson, S. M., Rao, R. P. & Brunton, B. W. Learning neural decoders without labels using multiple data streams. Preprint at <https://www.biorxiv.org/content/10.1101/2021.09.10.459775v1.full> (2021).
131. Ackman, J. B., Burbridge, T. J. & Crair, M. C. Retinal waves coordinate patterned activity throughout the developing visual system. *Nature* **490**, 219–225 (2012).
132. Moon, C., Lagercrantz, H. & Kuhl, P. K. Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica* **102**, 156–160 (2013).
133. DeCasper, A. J. & Spence, M. J. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behav. Dev.* **9**, 133–150 (1986).
134. Lobo, M. A., Kokkoni, E., de Campos, A. C. & Galloway, J. C. Not just playing around: infants' behaviors with objects reflect ability, constraints and object properties. *Infant Behav. Dev.* **37**, 334–351 (2014).
135. Soska, K. C. & Adolph, K. E. Postural position constrains multimodal object exploration in infants. *Infancy* **19**, 138–161 (2014).
136. Campos, J. J. et al. Travel broadens the mind. *Infancy* **1**, 149–219 (2000).
137. Barsalou, L. W. Grounded cognition. *Annu. Rev. Psychol.* **59**, 617–645 (2008).
138. Dobson, V. & Teller, D. Y. Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies. *Vision Res.* **18**, 1469–1483 (1978).
139. Sokol, S. Measurement of infant visual acuity from pattern reversal evoked potentials. *Vision Res.* **18**, 33–39 (1978).
140. Fiser, J., Aslin, R., Lathrop, A., Rothkopf, C. & Markant, J. An infants' eye view of the world: implications for learning in natural contexts. In *Proc. International Conference on Infant Studies* (2006).
141. Franchak, J. M., Kretch, K. S., Soska, K. C. & Adolph, K. E. Head-mounted eye tracking: a new method to describe infant looking. *Child Dev.* **82**, 1738–1750 (2011).
142. Smith, L. B., Yu, C., Yoshida, H. & Fausey, C. M. Contributions of head-mounted cameras to studying the visual environments of infants and young children. *J. Cogn. Dev.* **16**, 407–419 (2015).
143. Yoshida, H. & Smith, L. B. What's in view for toddlers? Using a head camera to study visual experience. *Infancy* **13**, 229–248 (2008).
144. Smith, L. B., Jayaraman, S., Clerkin, E. & Yu, C. The developing infant creates a curriculum for statistical learning. *Trends Cogn. Sci.* **22**, 325–336 (2018).
145. Fausey, C. M., Jayaraman, S. & Smith, L. B. From faces to hands: changing visual input in the first two years. *Cognition* **152**, 101–107 (2016).
146. Davis, J. et al. Does neonatal imitation exist? Insights from a meta-analysis of 336 effect sizes. *Perspect. Psychol. Sci.* <https://doi.org/10.1177/1745691620959834> (2021).
147. Hunnius, S. & Bekkering, H. What are you doing? How active and observational experience shape infants' action understanding. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130490 (2014).
148. Meltzoff, A. N. & Moore, M. K. Explaining facial imitation: a theoretical model. *Infant Child Dev.* **6**, 179–192 (1997).
149. Meltzoff, A. N. & Marshall, P. J. Human infant imitation as a social survival circuit. *Curr. Opin. Behav. Sci.* **24**, 130–136 (2018).
150. Ray, E. & Heyes, C. Imitation in infancy: the wealth of the stimulus. *Dev. Sci.* **14**, 92–105 (2011).
151. Soderstrom, M. Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev.* **27**, 501–532 (2007).
152. Brand, R. J., Baldwin, D. A. & Ashburn, L. A. Evidence for 'motionese': modifications in mothers' infant-directed action. *Dev. Sci.* **5**, 72–83 (2002).
153. van Schaik, J. E., Meyer, M., van Ham, C. R. & Hunnius, S. Motion tracking of parents' infant-versus adult-directed actions reveals general and action-specific modulations. *Dev. Sci.* **23**, e12869 (2020).
154. Wass, S. V. et al. Infants' visual sustained attention is higher during joint play than solo play: is this due to increased endogenous attention control or exogenous stimulus capture? *Dev. Sci.* **21**, e12667 (2018).
155. Yu, C. & Smith, L. B. The social origins of sustained attention in one-year-old human infants. *Curr. Biol.* **26**, 1235–1240 (2016).
156. Yu, Y. et al. The theoretical and methodological opportunities afforded by guided play with young children. *Front. Psychol.* **9**, 1152 (2018).
157. Bazhydai, M., Westermann, G. & Parise, E. 'I don't know but I know who to ask': 12-month-olds actively seek information from knowledgeable adults. *Dev. Sci.* **23**, e12938 (2020).
158. Poulin-Dubois, D. & Brosseau-Liard, P. The developmental origins of selective social learning. *Curr. Directions Psychol. Sci.* **25**, 60–64 (2016).
159. Berlyne, D. E. *Conflict, Arousal and Curiosity* (McGraw-Hill, 1960).
160. Day, H. I. Curiosity and the interested explorer. *Performance & Instruction* **21**, 19–22 (1982).
161. Kidd, C., Piantadosi, S. T. & Aslin, R. N. The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* **7**, e36399 (2012).
162. Kidd, C., Piantadosi, S. T. & Aslin, R. N. The Goldilocks effect in infant auditory attention. *Child Dev.* **85**, 1795–1804 (2014).
163. Poli, F., Serino, G., Mars, R. & Hunnius, S. Infants tailor their attention to maximize learning. *Sci. Adv.* **6**, eabb5053 (2020).
164. Cohen, L. B. Uses and misuses of habituation and related preference paradigms. *Infant Child Dev.* **13**, 349–352 (2004).
165. Hunter, M. A. & Ames, E. W. A multifactor model of infant preferences for novel and familiar stimuli. *Adv. Infancy Res* **5**, 69–95 (1988).
166. Aslin, R. N. What's in a look? *Dev. Sci.* **10**, 48–53 (2007).
167. Haith, M. M. Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behav. Dev.* **21**, 167–179 (1998).
168. Adolph, K. E. et al. How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychol. Sci.* **23**, 1387–1394 (2012).
169. Hoch, J. E., O'Grady, S. M. & Adolph, K. E. It's the journey, not the destination: locomotor exploration in infants. *Dev. Sci.* **22**, e12740 (2019).
170. Oakes, L. M. & Baumgartner, H. A. Manual object exploration and learning about object features in human infants. In *Proc. 2012 IEEE International Conference on Development and Learning and Epigenetic Robotics* 1–6 (IEEE, 2012).

171. Elman, J. L. Learning and development in neural networks: the importance of starting small. *Cognition* **48**, 71–99 (1993).
172. Bengio, Y., Louradour, J., Collobert, R. & Weston, J. Curriculum learning. In *Proc. 26th Annual International Conference on Machine Learning* (eds. Pothoreckyj, A., Danyluk, L., Bottou, M. & Littman, L.) 41–48 (ACM, 2009).
173. Vogelsang, L. et al. Potential downside of high initial visual acuity. *Proc. Natl Acad. Sci. USA* **115**, 11333–11338 (2018).
174. Orhan, A. E., Gupta, V. V. & Lake, B. M. Self-supervised learning through the eyes of a child. In *Advances in Neural Information Processing Systems* Vol. 33 (eds. Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. A. & Garnett, R.) 9960–9971 (NIPS, 2020).
175. Newport, E. L., Bavelier, D. & Neville, H. J. Critical thinking about critical periods: perspectives on a critical period for language acquisition. In *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (ed. Dupoux, E.) 481–502 (MIT Press, 2001).
176. Molnár, Z., Luhmann, H. J. & Kanold, P. O. Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science* **370**, eabb2153 (2020).
177. Kostovic, I. & Rakic, P. Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain. *J. Comp. Neurol.* **297**, 441–470 (1990).
178. Achille, A., Rovere, M. & Soatto, S. Critical learning periods in deep neural networks. In *Proc. 7th International Conference on Learning Representations* (ICLR, 2019).
179. Carpenter, G. A. & Grossberg, S. The art of adaptive pattern recognition by a self-organizing neural network. *Computer* **21**, 77–88 (1988).
180. French, R. M. Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.* **3**, 128–135 (1999).
181. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C. & Wermter, S. Continual lifelong learning with neural networks: a review. *Neural Networks* **113**, 54–71 (2019).
182. Robins, A. Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Sci.* **7**, 123–146 (1995).
183. Hinton, G. E. & Plaut, D. C. Using fast weights to deblur old memories. In *Proc. 9th Annual Conference of the Cognitive Science Society* 177–186 (Erlbaum, 1987).
184. Kemker, R. & Kanan, C. FearNet: brain-inspired model for incremental learning. In *Proc. 6th International Conference on Learning Representations* (ICLR, 2018).
185. Kirkpatrick, J. et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl Acad. Sci. USA* **114**, 3521–3526 (2017).
186. Rannen, A., Aljundi, R., Blaschko, M. B. & Tuytelaars, T. Encoder based lifelong learning. In *Proc. IEEE International Conference on Computer Vision* 1320–1328 (IEEE, 2017).
187. Draelos, T. J. et al. Neurogenesis deep learning: extending deep networks to accommodate new classes. In *Proc. 2017 International Joint Conference on Neural Networks* 526–533 (IEEE, 2017).
188. Javed, K. & White, M. Meta-learning representations for continual learning. In *Advances in Neural Information Processing Systems* (eds. Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. & Garnett, R.) 1818–1828 (NIPS, 2019).
189. Kemker, R., McClure, M., Abitino, A., Hayes, T. L. & Kanan, C. Measuring catastrophic forgetting in neural networks. In *Proc. Thirty-Second AAAI Conference on Artificial Intelligence* Vol. 415 (eds. McIlraith, S. A. & Weinberger, K. Q.) 3390–3398 (AAAI, 2018).
190. Settles, B. *Active Learning Literature Survey* (Univ. Wisconsin-Madison Department of Computer Sciences, 2009).
191. Settles, B. From theories to queries: active learning in practice. In *Proc. Active Learning and Experimental Design Workshop in Conjunction with AISTATS 2010* Vol. 16 (eds. Guyon, I., Cawley, G., Dror, G., Lemaire, V. & Statnikov, A.) 1–18 (MLR, 2011).
192. Botvinick, M. M., Niv, Y. & Barto, A. G. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* **113**, 262–280 (2009).
193. Lefort, M. & Gepperth, A. Active learning of local predictable representations with artificial curiosity. In *Proc. 2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics* 228–233 (IEEE, 2015).
194. Graves, A., Bellemare, M. G., Menick, J., Munos, R. & Kavukcuoglu, K. Automated curriculum learning for neural networks. In *Proc. 34th International Conference on Machine Learning* Vol. 70 (eds. Precup, D. & Teh, Y. W.) 1311–1320 (JMLR, 2017).
195. Schmidhuber, J. Driven by compression progress: a simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In *Proc. Workshop on Anticipatory Behavior in Adaptive Learning Systems* (eds. Pezzulo, G., Butz, M. V., Sigaud, O. & Baldassarre, G.) 48–76 (Springer, 2008).
196. Oudeyer, P.-Y. Computational theories of curiosity-driven learning. In *The New Science of Curiosity* (ed. Gordon, G.) 43–72 (Nova Science, 2018).
197. Oudeyer, P.-Y., Kaplan, F. & Hafner, V. V. Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evolution. Comput.* **11**, 265–286 (2007).
198. Twomey, K. E. & Westermann, G. Curiosity-based learning in infants: a neurocomputational approach. *Dev. Sci.* **21**, e12629 (2018).
199. Haber, N., Mrowca, D., Fei-Fei, L. & Yamins, D. L. Emergence of structured behaviors from curiosity-based intrinsic motivation. Preprint at <https://arxiv.org/abs/1802.07461> (2018).
200. Soltoggio, A., Stanley, K. O. & Risi, S. Born to learn: the inspiration, progress, and future of evolved plastic artificial neural networks. *Neural Networks* **108**, 48–67 (2018).
201. Arandjelovic, R. & Zisserman, A. Look, listen and learn. In *Proc. IEEE International Conference on Computer Vision* 609–617 (IEEE, 2017).
202. Barros, P., Parisi, G. I., Weber, C. & Wermter, S. Emotion-modulated attention improves expression recognition: a deep learning model. *Neurocomputing* **253**, 104–114 (2017).
203. Senocak, A., Oh, T.-H., Kim, J., Yang, M.-H. & Kweon, I. S. Learning to localize sound source in visual scenes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 4358–4366 (IEEE, 2018).
204. Ashby, F. G. & Vucoovich, L. E. The role of feedback contingency in perceptual category learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **42**, 1731–1746 (2016).
205. Parisi, G. I., Tani, J., Weber, C. & Wermter, S. Lifelong learning of human actions with deep neural network self-organization. *Neural Networks* **96**, 137–149 (2017).
206. Bonawitz, E. & Shafto, P. Computational models of development, social influences. *Curr. Opin. Behav. Sci.* **7**, 95–100 (2016).
207. Brockman, G. et al. OpenAI gym. Preprint at <https://arxiv.org/abs/1606.01540> (2016).
208. Geirhos, R. et al. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *Proc. 7th International Conference on Learning Representations* (2019).

Acknowledgements

This work was supported by an ERC Advanced Grant (FOUNDCOG, #787981) awarded to R.C. and an MSCA Individual Fellowship (InterPlay, #891535) awarded to L.Z. We thank K. Storrs and K. Körding as well as the *Nature Machine Intelligence* editorial team for their helpful feedback.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Lorian Zaadnoordijk, Tarek R. Besold or Rhodri Cusack.

Peer review information *Nature Machine Intelligence* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2022