Symposium

Recent Advances at the Interface of Neuroscience and Artificial Neural Networks

Yarden Cohen,[1] Tatiana A. Engel,[2] Christopher Langdon,[2] Grace W. Lindsay,[3] Torben Ott,[4] Megan A. K. Peters,[5] James M. Shine,[6] Vincent Breton-Provencher,[7] and Srikanth Ramaswamy[8]

[1]Department of Brain Sciences, Weizmann Institute of Science, Rehovot, 76100, Israel, [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY 11724, [3]Department of Psychology, Center for Data Science, New York University, New York, NY 10003, [4]Bernstein Center for Computational Neuroscience Berlin, Institute of Biology, Humboldt University of Berlin, 10117, Berlin, Germany, [5]Department of Cognitive Sciences, University of California–Irvine, Irvine, CA 92697, [6]Brain and Mind Centre, University of Sydney, Sydney, NSW 2006, Australia, [7]Département de psychiatrie et neurosciences, Université Laval, Quebec City, Québec, G1J 2G3, Canada, and [8]Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, United Kingdom

Biological neural networks adapt and learn in diverse behavioral contexts. Artificial neural networks (ANNs) have exploited biological properties to solve complex problems. However, despite their effectiveness for specific tasks, ANNs are yet to realize the flexibility and adaptability of biological cognition. This review highlights recent advances in computational and experimental research to advance our understanding of biological and artificial intelligence. In particular, we discuss critical mechanisms from the cellular, systems, and cognitive neuroscience fields that have contributed to refining the architecture and training algorithms of ANNs. Additionally, we discuss how recent work used ANNs to understand complex neuronal correlates of cognition and to process high throughput behavioral data.

Key words: plasticity; artificial neural networks; neuromodulators; behavior; vision; cognition

Introduction

Recent technological advances have transformed our access to the fine-grain spatiotemporal organization of the anatomy and physiology of biological neural networks. Over the years, big data on an astounding diversity of genes, proteins, neurons and glia, dendrites, synapses, and neural network functions have transformed our understanding of the brain (Sejnowski et al., 2014). On the other hand, brain-inspired implementations of artificial neural networks (ANNs), the perceptron model (McCulloch and Pitts, 1943; Rosenblatt, 1958), Boltzmann machines (Ackley et al., 1985), and Hopfield networks (Hopfield, 1982), have had profound implications for biological research and computational problems. These ANN architectures have been foundational for applications in pattern completion, attractor networks, dynamical systems, and diverse algorithmic capabilities in modern convolutional, multilayer, and recurrent neural network (RNN) models.

Although biological neural networks have continued to guide the development of their artificial counterparts, the beacon has been held by mathematics and statistical physics to develop efficient models of optimization functions (Sutskever et al., 2011; Cox and Dean, 2014). ANNs have leapfrogged from nonlinear systems and networks (Minsky and Papert, 1972; Haykin, 1994) to deep and recurrent networks (LeCun et al., 2015; Schmidhuber, 2015). More recently, backpropagation of error (Werbos, 1974, 1982; Rumelhart et al., 1986) has enabled the efficient training of neural networks, by computing gradients with respect to the weights of a multilayer network. Although methods to train ANNs have evolved to include improved weight initializations, optimization, and gradient descent algorithms, they do not appear to have any analogous neurobiological principles (Marblestone et al., 2016).

Here, we review the current state of the art of ANN models in terms of "biological realism," their applications and limitations, with the ultimate aim of identifying the operational principles and functional settings through which biological neural networks and ANNs can inform each other toward synergistic development. With this background, the review focuses on four distinct aspects.

1. How can biological intelligence guide the refinement of ANN architectures?
2. How can ANNs drive a better understanding of cognition?
3. What are the limitations of ANNs with respect to modeling human cognition?
4. What are the recent advances in applying ANNs to quantify complex behavior?
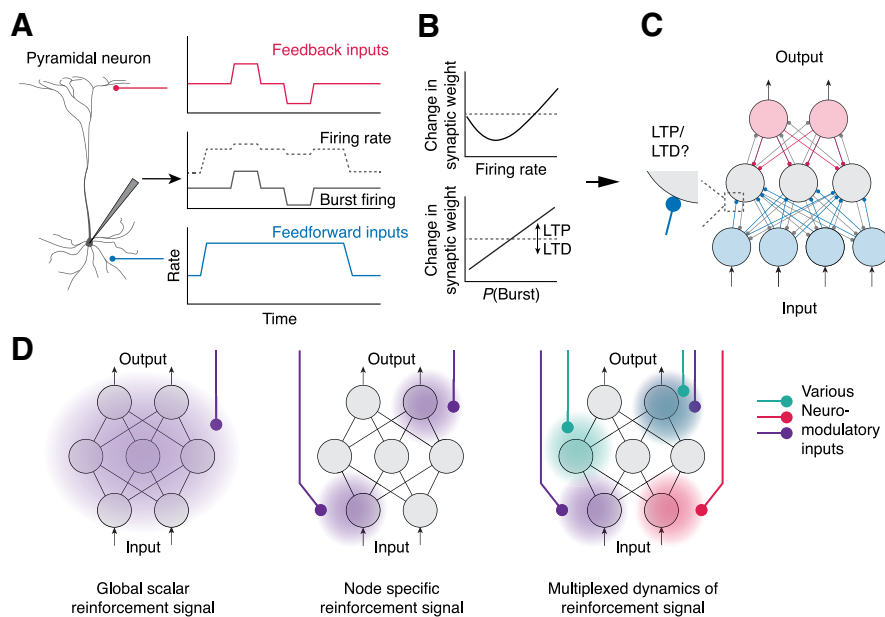
Figure 1. Dendritic integration of inputs and neuromodulation-aware deep learning. **A**, How a pyramidal neuron responds to an input depends on dendritic location. Feedforward inputs located near the soma directly drive the firing rate of the neuron, whereas feedback inputs on apical dendrites affect burst firing (P(Burst)). **B**, The firing rate of presynaptic and postsynaptic neurons and P(Burst) control plasticity long-term potentiation - LTP; long-term depression - LTD. **C**, The dendritic integration of feedback and feedforward inputs by cortical neurons could solve the credit assignment problem in hierarchical ANNs. **D**, Diagram of how neuromodulation can be integrated by ANNs. Left, Error signal of a network perturbation is carried through a global neuromodulatory influence. Middle, Error signals are carried through node-specific neuromodulatory inputs. Right, Various neuromodulatory inputs could take part in signaling distinct error functions. **A–C**, Adapted from Payeur et al. (2021).

Refining ANNs with biological precision

ANNs share many interesting features in common with biological neural networks. This is, of course, no accident, as the original ANN algorithms were in part inspired by the anatomy of the cerebral cortex (Sejnowski, 2020). The successful use of ANNs to model computations has forced a recalibration of our working models of the nervous system, leading to the embrace of dynamical models of computation that incorporate distributed computations across widespread, ever-changing networks (Sohn et al., 2019). Advances in cellular neuroscience, neuroimaging, and computational modeling are enabling the integration of new details into advanced versions of ANNs that will, hopefully, bring us closer to the goal of understanding how the brain works, while simultaneously refining artificial intelligence (AI).

A prime example of an emerging tension between neuroscience and AI is the recognition that pyramidal neurons, the workhorse of the cerebral cortex and the primary feature mimicked by ANNs, have highly nonlinear operating modes (Larkum, 2013). Traditional models of pyramidal neurons assumed that the dendrites of pyramidal neurons linearly summed their action potentials within a given window, and only spiked when the inputs exceeded a certain threshold (Larkum, 2022). In stark contrast, recent work has clearly demonstrated that many pyramidal neurons in the cerebral cortex have distinct modes of operation, sometimes firing linearly with inputs and other times ignoring inputs altogether (Fig. 1A,B) (Ramaswamy and Markram, 2015; Roelfsema and Holtmaat, 2018; Richards et al., 2019). Rather than reflecting passive integrative inputs, the active dendrites of pyramidal neurons have been shown to underpin striking computational complexity (Johnston and Narayanan, 2008; Sruston, 2008; Poirazi and Papoutsi, 2020; Larkum, 2022). Indeed, deep neural networks with at least 5-8 layers are needed to model the complex input/output functions of pyramidal cells (Beniaguev et al., 2021). The ability to convert the presumed integrator-like dynamics of neurons and their dendrites to coincident detectors (or resonators) is an important function that specific ion channels perform (Rudolph and Destexhe, 2003; Ratté et al., 2013). The identity of individual neurons (integrators vs resonators) has important implications for connectivity, computation, and information coding (Rudolph and Destexhe, 2003; Ratté et al., 2013). Such features have been recently incorporated into ANNs, toward solving the so-called credit-assignment problem (Payeur et al., 2021) using single-phase learning (Fig. 1C,D) (Greedy et al., 2022).

Nonetheless, plasticity, as a biological mechanism, is not limited to synaptic contacts. With emerging roles attributed to entire neurons (engram cells) in the physiology of memory, learning theories that focus exclusively on synaptic plasticity appear to be inadequate as a premise for models of ANNs (Titley et al., 2017; Lisman et al., 2018; Josselyn and Tonegawa, 2020). Plasticity is a ubiquitous phenomenon, which spans multiple scales of organization in biological neural networks: from synapses and dendritic branches to neurons and microcircuits (Le Bé and Markram, 2006; Branco and Häusser, 2010; Titley et al., 2017; Mishra and Narayanan, 2021). Incorporating a broad repertoire of plasticity mechanisms, such as those available to biological neural networks, is an essential step in refining ANN architectures and extending their utility. We are only just scratching the surface of the potential for biological insights to suggest novel algorithmic solutions to problems that have been trained on classical network architectures, such as RNNs.

The cerebral cortex is also deeply embedded within a web of dense interconnections with a number of highly conserved subcortical structures whose functional importance to the working of the nervous system should not be understated. One particular structure that is often overlooked in ANNs is the thalamus, a bilateral structure in the diencephalon that is densely (and heterogeneously) interconnected with the cerebral cortex (Jones, 2001). Although the functional benefits of one class of corticothalamic cell is relatively well understood (the so-called "core" regions) (Crandall et al., 2015), the more diffusely projecting "matrix" cells remain more enigmatic. Recently, a neural mass model of the corticothalamic system was created to investigate the impact of this topological projection on emergent whole-brain dynamics (Shine, 2021). In brief, the model found that the matrix cells tuned the functional repertoire of the brain, providing a flexible, yet robust, platform for instantiating an array of novel combinations of cortical coalitions. Others have shown that these same cells can alter information flow in neural circuits (Anastasiades et al., 2021; Mukherjee et al., 2021) and are crucial sites for behaviorally relevant plasticity (Williams and Holtmaat, 2019). It would be interesting to note how these circuit-level features could inform future implementations of ANNs, such as

models that mimic the interactions between the cerebellum and cortex (Pemberton et al., 2021; Boven et al., 2022).

The operating mode of the cerebral cortex (along with the rest of the brain) is also fundamentally altered by the presence (or absence) of neuromodulatory chemicals, such as noradrenaline, acetylcholine, and serotonin. By interacting with GPCRs on target neurons and glia, these ligands can alter the excitability and receptivity of the network (Shine et al., 2021), facilitating different information processing regimens that shift neural populations between information storage and information transfer (Li et al., 2019). These changes in gain, while relatively low-dimensional, can substantially impact the functional outputs of ANNs (Stroud et al., 2018), suggesting that their incorporation into modern deep learning architectures could be quite informative (Mei et al., 2022). In addition, by combining these approaches with sophisticated, high-resolution recordings of the neuromodulatory system *in vivo* (Breton-Provencher et al., 2022), we can also simultaneously test hypotheses regarding the functional operation of the brain as well.

A prominent example of biologically inspired ANNs that has gained considerable interest in machine vision is the convolutional neural network (CNN). CNNs are extensions of ANNs with an architecture inspired by that of the mammalian visual system, with convolutions representing the function of simple cells and the pooling operations of complex cells (Lindsay, 2021). When trained appropriately, these models can produce representations that match those of biological visual systems better than previous models (Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014). Traditionally CNNs are strictly feedforward; that is, they do not include lateral or feedback recurrent connections (Fig. 2). Yet, it is known that visual systems of humans contain many such connections, and these connections are implicated in important computations, such as object recognition (Wyatte et al., 2012). Previous work has shown how these connections can make models better at visual tasks and better match biological processing (Fig. 2) (Spoerer et al., 2017; Linsley et al., 2018; Kubilius et al., 2019; Nayebi et al., 2021). An unmet potential of these models, however, is to use them as an idealized experimental setup to analyze the computational role that recurrence plays. Promising work in this direction has shown that recurrence can help object classification by carrying information about unrelated, auxiliary variables (Thorat et al., 2021).

In a recent study (Lindsay et al., 2022), four different kinds of recurrence were added to a feedforward CNN: feedback connections that implement predictive processing to one network, lateral connections that implement surround suppression to another, and two more networks with feedback and lateral connections trained directly to classify degraded images. This choice of task, wherein the network must classify images of digits that are degraded by one of several types of noise, such as occlusion and blur, was chosen to capture some of the functions believed to be performed by recurrence. Counterintuitively, recurrence added to the CNN was not related to its function in these models: both forms of task-trained recurrence (feedback and
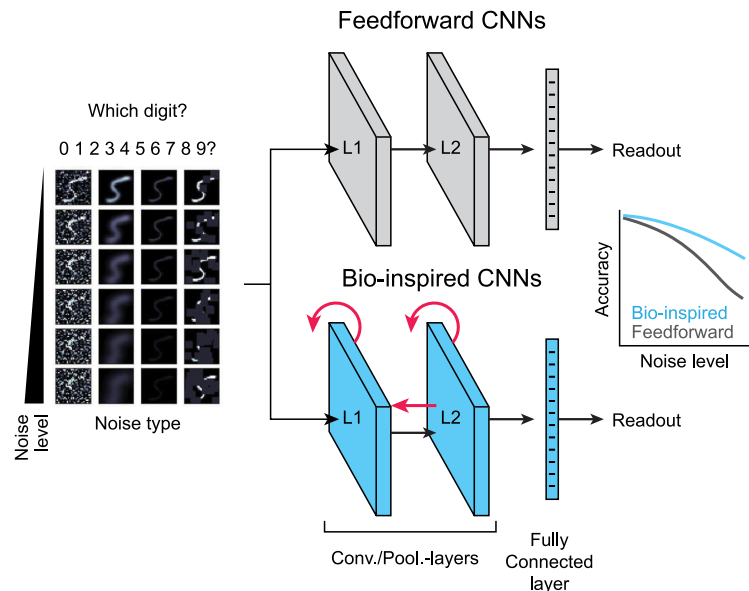

Figure 2. Feedforward versus bio-inspired CNNs. By adding connections inspired by the anatomy and physiology of the visual system, such as lateral (e.g., center surround suppression) or feedback (e.g., top-down predictions), CNNs with recurrent connections show improved accuracy. Black and red arrows represent feedforward and recurrent connections, respectively. Adapted from Lindsay et al. (2022).

lateral connections) change neural activity and behavior similarly to each other and differently from their bio-inspired anatomic counterparts. Specifically, in the case of feedback, predictive feedback denoises the representation of noisy images at the first layer of the network, leading to an expected increase in classification performance. In the task-trained networks, representations are not denoised over time at the first layer (indeed, they become "noisier"), yet these dynamics do lead to denoising at later layers and increased performance. We analyzed an open fMRI dataset (Abdelhack and Kamitani, 2018) using the same tools, such as dimensionality reduction, activity correlations, and representational geometry analysis, applied to the models and found weak support for the predictive feedback model. Such analysis of artificial networks provides an opportunity to test the tools of systems neuroscience (Lindsay, 2022).

Decisions, artificial RNNs, and functional neuron types
Many ingredients make up our decisions, a rich stream of sensory information, a lifetime of memories, long-term goals, and current mood or emotions. This poses a challenge in identifying the neural processes of decision formation: The activity of cortical neurons, for instance, reflects an equally large complexity of decision-related features, from sensory and spatial information (Rao et al., 1997), to short-term memory (Funahashi et al., 1989), economic value (Padoa-Schioppa and Assad, 2006), risk (Ogawa et al., 2013) and confidence (Kepecs et al., 2008), or abstract rules (Wallis et al., 2001). Furthermore, single neurons often demonstrate mixtures of these features (Mante et al., 2013; Rigotti et al., 2013; Fusi et al., 2016), precluding straightforward functional interpretations of the signal they carry (Fig. 3). How can we identify any organizational principles by which cortical neurons or neural networks take part in decision-making?

Recent approaches have focused on neural population as the primary computational units for cognition (Pandarinath et al., 2018; Saxena and Cunningham, 2019; Vyas et al., 2020; Barack and Krakauer, 2021; Duncker and Sahani, 2021; Ebitz and Hayden, 2021; Jazayeri and Ostojic, 2021). Population
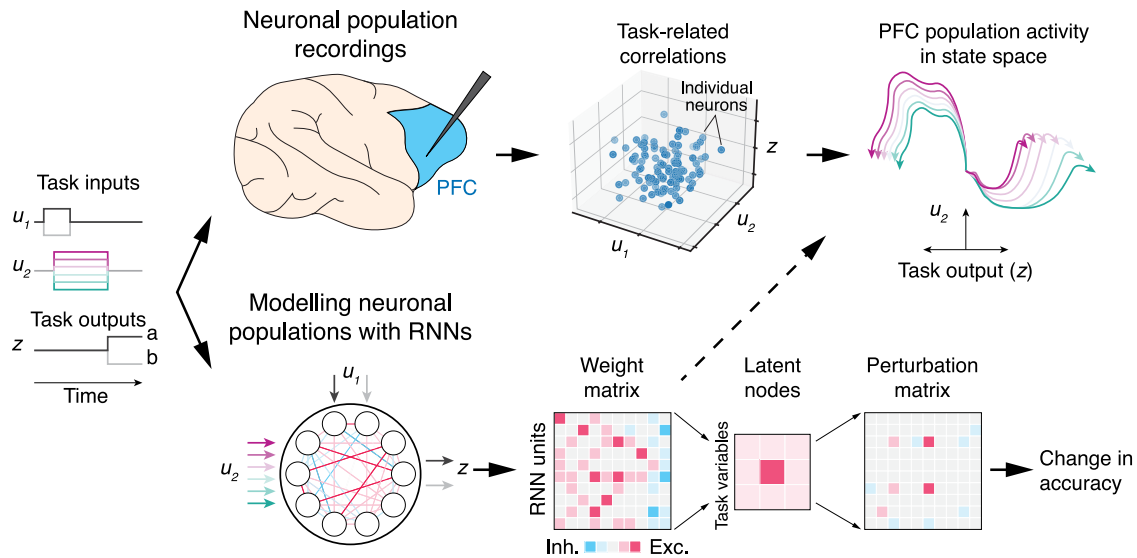
Figure 3. Using RNNs to study neuronal correlates of complex tasks. In an example task where various contexts ($u1$) and sensory cues ($u2$) guide task outputs (z) or decisions, recordings of neurons from associative brain areas (e.g., pre-frontal cortex - PFC) show multidimensional encoding of task variables by individual neurons. Representing population dynamics in neural state space where each point in space represents a unique pattern of neuronal activity that is useful to dissect how the correlated activity of a large number of neurons represents task variables. To model physiological dynamics, RNNs are trained to perform a similar task. Key features of physiological dynamics of neuronal populations are reproduced by RNNs. Complex perturbation studies can thus be performed with these trained RNNs to test causality. In a recent study, Langdon and Engel (2022) found that low-dimensional latent circuits can be extracted from high-dimensional RNN dynamics and used to perform patterned connectivity perturbations. Adapted from Mante et al. (2013) and Langdon and Engel (2022).

approaches identify low-dimensional patterns in neural population data, describing the subspaces or manifolds in which neural trajectories move (Cunningham and Yu, 2014). Applied to neural population recordings during flexible decision-making as a prime example, information about sensory information, choice, and rules can be reliably separated at the level of neural populations (Mante et al., 2013; Malagon-Vina et al., 2018; Sohn et al., 2019; Aoi et al., 2020; Ebitz and Hayden, 2021).

RNNs, which are able to mimic the complexity of real cortical responses, have served as a valuable model for understanding computation in large heterogeneous neural populations. For instance, RNNs suggest dynamic network mechanisms by which decision rules are flexibly applied to determine a decision (Mante et al., 2013). Hopfield networks and restricted Boltzmann machines provide valuable insight into the storage and retrieval of associative memories via unsupervised learning rules (Marullo and Agliari, 2020). Recently, supervised learning approaches have been used to train RNNs to perform the same cognitive tasks as behaving animals. This approach provides a powerful alternative for studying how neural computations underlying cognitive tasks are distributed across heterogeneous populations (Fig. 3) (Mante et al., 2013; Song et al., 2016; Wang et al., 2018; Yang et al., 2019) and how networks leverage previously learned tasks for continual learning (Driscoll et al., 2022). Because of the complexity of their connectivity and dynamics, reverse engineering trained RNNs mimic the challenges faced when analyzing real neural data. This observation has motivated their use as a testbed for candidate dimensionality reduction methods aimed at uncovering low-dimensional latent dynamics. Such methods model heterogeneous neural responses as linear mixing of task-relevant variables and can uncover neural mechanisms which exist only at the population level (Cunningham and Yu, 2014; Kobak et al., 2016). The ability to perform precise perturbation tests in RNNs (Yang et al., 2019) offers the possibility of validating the causal role of neural representations revealed by

candidate dimensionality reduction strategies (Mante et al., 2013; Song et al., 2016; Wang et al., 2018; Yang et al., 2019).

In a recent study, RNNs were trained on cognitive tasks to develop and validate latent circuit models of heterogeneous neural responses (Langdon and Engel, 2022) (Fig. 3). The latent circuit model uncovers low-dimensional task-relevant representations together with recurrent circuit interactions between these representations. To validate this method, RNNs were trained on a motion-color discrimination task in which the subject must flexibly discriminate either the motion or color of random dot stimulus depending on a contextual cue (Langdon and Engel, 2022) (Fig. 3). Fitting a latent circuit model to the responses of this RNN revealed a latent inhibitory mechanism in which contextual representations inhibit irrelevant stimulus representations, allowing the network to flexibly select the correct stimulus–response association (Langdon and Engel, 2022). This inhibitory mechanism is mirrored in dynamics as a suppression of irrelevant stimulus representations.

Despite the success of population-centered analysis, recent studies have discovered groups of single neurons with prototypical dynamic activity and encoding of decision variables. For instance, Hirokawa et al. (2019) started from neural population activity but considered the possibility that the neurons' dynamic activity as well as its tuning to decision variables in a combined sensory and value-based decision task clustered into distinct groups of neurons with distinct dynamic and tuning profiles. Unsupervised clustering revealed dedicated groups of single neurons in rat orbitofrontal cortex that were tuned to canonical decision variables, that is, combinations of task features that explained the animals' decision behavior, such as reward likelihood, integrated value, and choice outcome. A dedicated group of neurons in the orbitofrontal cortex carried information about the certainty that a decision was correct (i.e., decision confidence) (Masset et al., 2020). These neurons predicted subsequent confidence-guided behavior: the variable time rats invested into their decision to obtain an uncertain, delayed reward before abandoning their investment (Lak

et al., 2014; Ott et al., 2018). These groups of neurons might constitute functional neuron types, characterized by assuming specific algorithmic roles to realize decision computations (Christensen et al., 2022). Similarly, functional clusters were found in the orbitofrontal cortex during value-based decision tasks in rats (Hocker et al., 2021) and primates (Onken et al., 2019), and in mice using calcium imaging during associative learning (Namboodiri et al., 2019).

When and why might we expect to find functional neuron types? Recent computational studies using RNNs suggest that neural subpopulations with distinct dynamics or categorical representations arise in trained networks that are required for flexible decision-making, such as context-dependent decision tasks (Dubreuil et al., 2022; Flesch et al., 2022; Langdon and Engel, 2022). Functional neuron types might thus be a feature shared by biological neural networks and ANNs to provide a robust computational solution for flexible decision-making. On the other hand, these interpretations are limited, since it is unclear what the biological counterpart to an ANN unit might precisely be. While many approaches interpret RNN units as candidates for single neurons (Barrett et al., 2019), the complex computations performed by single neurons outperform simple RNN units and can only be described by deep networks themselves (Beniaguev et al., 2021). Specifically, the functional coupling between neuronal compartments (dendrites and soma, compare with previous section) can be controlled by thalamic input (Aru et al., 2020), and depends on learning (d'Aquin et al., 2022) further suggesting that RNN units might correspond to computations of neuronal compartments or biophysical processes. Categorical representations in RNNs might thus shed light onto the functions performed by biophysical elements of neurons.

Functional neuron types might emerge as a result of the cortical microcircuit structure. Emerging evidence suggests that cortical cell types, defined by distinct gene expression or connectivity patterns (Tasic et al., 2018; Winnubst et al., 2019), assume specialized functions during decision-making. For example, orbitofrontal cortex neurons that project to the striatum predominantly carry sustained task-related signals (Bari et al., 2019; Terra et al., 2020), such as information about choice outcome (Hirokawa et al., 2019) (whether the animal was rewarded or not), and projection-defined neurons in motor cortex signal movement onset or choice signals, respectively (Economo et al., 2018). Cell type identity might thus be a structural constraint on the dynamic decision algorithms in biological neural networks that could inform the design of ANNs (Sacramento et al., 2018; Greedy et al., 2022).

Uncertainty and decisions: can insights from human (and animal) cognition contribute to AI development?
Humans make decisions based on perceptual, value-based, or other information. Such decisions are accompanied by a sense of confidence. That is, our brains seem to compute not only the best decisional outcome, but also estimates related to the probability that the decision is correct (Pouget et al., 2016; Mamassian, 2022; Peters, 2022). This sense of uncertainty accompanies the moment-to-moment information processing across many perceptual and cognitive domains, and can help any organism decide whether to update their internal models of the world, how to allocate resources, or how to sample new information. Importantly, computing decisional confidence can also help us to better learn from erroneous predictions (Guggenmos et al., 2016; Stolyarova et al., 2019; Ptasczynski et al., 2022).

An argument can thus be made that getting artificial systems to also compute such a confidence judgment could lead not only to better decision-making under uncertainty, but also better and more self-directed learning. A foundational goal of AI research is to build systems that not only behave adaptively in the world, but which "know" when they have made correct or erroneous decisions, or when they have such a high level of uncertainty that they should sample more information before committing to a decision at all. Thus far, most "confidence" type signals in artificial systems typically compute uncertainty estimates according to probabilistic inference: for example, the variance of a (posterior) probability distribution, or entropy of an outcome distribution, can potentially be reasonable proxies for confidence in biological systems (Li and Ma, 2020). This is because these quantities reflect the relative evidence in favor of multiple possible decisional outcomes. However, there are a number of problems with these approaches for uncertainty estimation in artificial systems.

First, it is not clear that humans and other animals rate confidence according to optimal inference, as implemented in ANNs or similar; instead, a large body of work suggests that other influences on decisional confidence are likely, ranging from motor preparation and execution (Fleming et al., 2015) to detectability heuristics (Maniscalco et al., 2016, 2021; Rausch et al., 2018). While these contributions to uncertainty/confidence estimates may seem suboptimal or even random, we also must note that our systems have been optimized through millennia of evolution, such that apparent "biases" in confidence judgments may actually reflect some optimal behavior where the cost function remains unknown to us as researchers (Michel and Peters, 2021).

Second, current implementations of confidence in multidecision alternatives often do not have an option for artificial systems to "opt out" of the decision, and instead decide to sample more information. Current AI does not have agency in such a fashion. However, we know that biological observers use confidence to guide their decision-making behavior, including decisions about whether and how to continue sampling their environments (Kepecs and Mainen, 2012; Guggenmos et al., 2016; Stolyarova et al., 2019; Ptasczynski et al., 2022). One area in which uncertainty-based self-directed information sampling is likely to be of utility is in meta-learning, wherein an artificial system must learn which weights to update based on an inferred context (and thus may avoid catastrophic forgetting when trained on multiple tasks). Several architectures implementing explicit metacognition or confidence have been proposed (Griffiths et al., 2019). For example, in rats trained to report confidence by placing a wager on difficult decisions, single neurons in the frontal cortex encode uncertainty information across multiple sensory modalities, and predict both confidence-scaled time investment in learning (Lak et al., 2014; Masset et al., 2020). These results suggest a generalized representation of confidence as a "summary scalar" that could provide a robust uncertainty signal used for subsequent decisions or learning processes and therefore constitute a precursor of metacognitive signals. The field is ripe for more exploration whether such biological implementations could inform uncertainty predictions in ANNs (Griffiths et al., 2019; Gawlikowski et al., 2021).

Here, we have discussed one among many examples of how artificial system development may benefit from the study of metacognition and confidence in biological systems, and vice versa. Future work may also examine how cooperation between humans and artificial systems may be optimized through confidence-weighted communication, as it is between dyads or small

groups of human deciders (Bahrami et al., 2010). Efforts to align vocabularies, literatures, and concepts across the fields of cognitive science and AI will assuredly benefit both fields.

Development of deep learning algorithms for high throughput processing of complex behavior

Studying natural behaviors affords new understanding of how the brain controls motion (Krakauer et al., 2017) and processes sensory inputs (Hayhoe, 2018). But two major characteristics of natural behaviors challenge their use in neuroscience experiments: dynamic properties that are often difficult to quantify and rich repertoires that require processing datasets much larger than tractable with traditional manual or semimanual methods. Modern machine learning offers both unsupervised and supervised approaches to meet these challenges.

Unsupervised algorithms help researchers identify structure in high-dimensional behavior data (Gilpin et al., 2020; McCullough and Goodhill, 2021). For example, researchers applied unsupervised dimensionality reduction, linear projections, or nonlinear embeddings (Tenenbaum et al., 2000; van der Maaten and Hinton, 2008; McInnes et al., 2018) to videos of freely moving worms (Stephens et al., 2008), fish (Mearns et al., 2020), flies (Berman et al., 2014), rodents (Stringer et al., 2019), and primates (Yan et al., 2020). Then, unsupervised clustering could discover robust behavior states in the resulting low-dimensional representations. Additionally, clustering was replaced by hidden Markov models (HMMs), capturing sequences of behavior states (Wiltschko et al., 2015).

Successful approaches using ANNs to process textual, auditory, and visual data, not as models in neuroscience, were recently harnessed and applied to quantify complex behavior in tandem or replacing these machine learning methods. Unsupervised ANNs were used because they better capture various data distributions (Graving and Couzin, 2020); convolutional and variational autoencoders (Batty et al., 2019; Graving and Couzin, 2020; Luxem et al., 2022) performed dimensionality reduction before clustering or fitting HMMs and generative adversarial networks (Goodfellow et al., 2014; Radford et al., 2015) improved the interpolation between low-dimensional representations (Sainburg et al., 2018).

Supervised ANNs are extremely useful in automating manual processing where humans can identify which data features to track and provide labeled training examples. Many striking examples come from pose estimation in movies of freely behaving animals. These algorithms learn to track human-defined anatomic features, such as joints and locations on the body of single or multiple animals, in videos captured from single (Mathis et al., 2018; Graving et al., 2019; Pereira et al., 2019) or multiple (Marshall et al., 2021) cameras. Mostly using deep convolutional ANNs, these supervised methods, like the unsupervised ones, extended and improved previous methods based on supervised classifier models (Dankert et al., 2009; Kabra et al., 2013; Machado et al., 2015).

Together, these ANN-based algorithms ushered the field of computational neuroethology (Anderson and Perona, 2014; Datta et al., 2019; Pereira et al., 2020). However, while many species naturally vocalize and offer a rich window onto complex social interactions, fewer works developed audio analysis methods comparable with those created for video analysis (Sainburg and Gentner, 2021).

Audio analyses predominantly start by converting sound signals to spectrograms, a two-dimensional representation in the time and frequency domains. This "image of sound," like visual data, was used to extract low-dimensional representations. For example, human-defined features, such as pitch, entropy, and amplitude, were continuously extracted from spectrograms and automated measuring similarities between juvenile and tutor zebra finch songs (Tchernichovski et al., 2000). Unsupervised variational autoencoders were also used for continuous low-dimensional embedding of spectrograms (Goffinet et al., 2021). Still, rather than working on continuous signals, most machine learning tools for bioacoustics were developed for analyzing audio segments, thought to represent basic vocal units or syllables.

Segmenting vocal communication allows creating models of syntax (Berwick et al., 2011; Jin and Kozhevnikov, 2011; Markowitz et al., 2013; Hedley, 2016) and motor learning (Sober and Brainard, 2009, 2012), and to relate syllable acoustics and sequence to neural activity (Leonardo and Fee, 2005; Sober et al., 2008; Wohlgemuth et al., 2010). Researchers used unsupervised similarity metrics (Mets and Brainard, 2018), clustering (Daou et al., 2012; Burkett et al., 2015), embedding (Morfi et al., 2021; Sainburg et al., 2021), variational autoencoders (Kohlsdorf et al., 2020), and other generative deep networks (Pagliarini et al., 2021) to assist human identification of vocal units, visualize repertoire structures (Sainburg et al., 2020), and study their dynamics (Mets and Brainard, 2018; Kollmorgen et al., 2020). When human annotators created training sets of labeled audio segments, those segments were used to train supervised algorithms (Nicholson, 2016), support vector machines (Tachibana et al., 2014), template matching (Anderson et al., 1996), HMMs (Kogan and Margoliash, 1998), and k-nearest neighbors (Nicholson, 2016, 2021), that allowed scaling up analyses on annotated syllables.

Still, these methods require the audio to be *a priori* well segmented. Traditional segmentation techniques hence create a bottleneck, limiting the questions researchers can answer. Using supervised deep ANNs introduces various solutions to this problem in rodents (Coffey et al., 2019) and songbirds (Koumura and Okanoya, 2016; Steinfath et al., 2021; Cohen et al., 2022). For example, TweetyNet (Cohen et al., 2022) is a supervised deep ANN that leverages temporal vocal dependencies to achieve high-precision annotation of multiple species. TweetyNet offers a powerful tool to study the neuronal encoding of bird song syntax (Cohen et al., 2020) and demonstrates how development in modern machine learning opens new boundaries in the study of natural behavior.

Finally, as different research laboratories, developing and using various ANNs to analyze behavior, also develop their own data formats and algorithms, it is of uttermost importance for our community to develop and foster an ecosystem of interoperable methods to increase reproducibility and access.

In conclusion, we have briefly reviewed the state of the art of ANNs and how their development has been inspired by biological neural networks. Although ANNs are remarkably effective at solving specific tasks, they lack the ability of biological neural networks to generalize robustly across tasks (but see Reed et al., 2022). We suggest that future implementations of ANNs should incorporate some of the intricate multiscale organizing features of biological neural networks to generalize as well as they do and learn continually over a lifetime of experience.

Neuromodulatory systems endow biological neural networks with the ability to learn and adapt to constantly changing behavioral demands. Neuromodulators, such as dopamine, serotonin, noradrenaline, and acetylcholine, play crucial roles in modulating a repertoire of brain states from reward assessment,

motivation, patience, arousal, and attention. The diverse phenomenology of neuromodulatory function is yet to be fully explored in ANNs, and their implementation has so far been mostly restricted to models of reinforcement learning (Shine et al., 2021; Mei et al., 2022).

Neurotransmitter and neuromodulator receptors are thought to modulate perceptual processes by activating receptor "hotspots" on the distal apical dendrites of neocortical layer 5 pyramidal cells (Takahashi et al., 2016). Recent implementations of ANNs have incorporated dendritic mechanisms to address the credit-assignment problem (Sacramento et al., 2018). Incorporating neurotransmitter receptor clusters on dendrites in ANNs could help unravel their role in gating perceptual processes, for example, NMDA receptors that are distributed nonlinearly on the dendrites of most neuron types (Chavlis and Poirazi, 2021).

Biological neural networks promote renormalization and homeostasis of synaptic strength during different states of sleep, and facilitate learning and memory through replay. Replay enables the brain to consolidate memory and overcome forgetting of acquired knowledge, also referred to as "catastrophic forgetting" in machine learning. Implementing "sleep-like states" in deep neural networks could mimic biological replay mechanisms and prevent catastrophic forgetting (Roscow et al., 2021; Kudithipudi et al., 2022; Mei et al., 2022; Tsuda et al., 2022).

Recent findings demonstrate that metabolic state dynamically governs coding precision in biological neural networks. Metabolic scarcity in the brain inactivates biological neural networks required for long-term memory to preserve energy (Padamsey et al., 2022). Biological neural networks regulate energy use through intrinsic mechanisms that determine the degree of energy consumption by reducing the impact of subthreshold variability on information coding. Therefore, biological neural networks dynamically adapt their coding precision and energy expenditure in a context-dependent manner (Padamsey et al., 2022). Most ANN architectures are energetically expensive. How could metabolic principles controlling coding precision inform implementations of energy efficient ANNs?

Neuroscience is witnessing significant advances in our understanding of biological learning mechanisms that can continue to inform new avenues for ANNs. We suggest that the machine learning community could adopt these ideas and integrate them into standard ANN frameworks to build a solid foundation, and develop the next generation of ANNs informed by the multiscale organizing features of their biological analogs.

References

Abdelhack M, Kamitani Y (2018) Sharpening of hierarchical visual feature representations of blurred images. eNeuro 5:ENEURO.0443-17.2018.

Ackley DH, Hinton GE, Sejnowski TJ (1985) A learning algorithm for Boltzmann machines. Cogn Sci 9:147–169.

Anastasiades PG, Collins DP, Carter AG (2021) Mediodorsal and ventromedial thalamus engage distinct L1 circuits in the prefrontal cortex. Neuron 109:314–330.e5.

Anderson DJ, Perona P (2014) Toward a science of computational ethology. Neuron 84:18–31.

Anderson SE, Dave AS, Margoliash D (1996) Template-based automatic recognition of birdsong syllables from continuous recordings. J Acoust Soc Am 100:1209–1219.

Aoi MC, Mante V, Pillow JW (2020) Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. Nat Neurosci 23:1410–1420.

Aru J, Suzuki M, Larkum ME (2020) Cellular mechanisms of conscious processing. Trends Cogn Sci 24:814–825.

Bahrami B, Olsen K, Latham PE, Roepstorff A, Rees G, Frith CD (2010) Optimally interacting minds. Science 329:1081–1085.

Barack DL, Krakauer JW (2021) Two views on the cognitive brain. Nat Rev Neurosci 22:359–371.

Bari BA, Grossman CD, Lubin EE, Rajagopalan AE, Cressy JI, Cohen JY (2019) Stable representations of decision variables for flexible behavior. Neuron 103:922–933.e7.

Barrett DG, Morcos AS, Macke JH (2019) Analyzing biological and artificial neural networks: challenges with opportunities for synergy? Curr Opin Neurobiol 55:55–64.

Batty E, Whiteway M, Saxena S, Biderman D, Abe T, Musall S, Gillis W, Markowitz J, Churchland A, Cunningham JP, Datta SR, Linderman S, Paninski L (2019) BehaveNet: nonlinear embedding and Bayesian neural decoding of behavioral videos. Advances in Neural Information Processing Systems 32.

Beniaguev D, Segev I, London M (2021) Single cortical neurons as deep artificial neural networks. Neuron 109:2727–2739.e3.

Berman GJ, Choi DM, Bialek W, Shaevitz JW (2014) Mapping the stereotyped behaviour of freely moving fruit flies. J R Soc Interface 11:20140672.

Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ (2011) Songs to syntax: the linguistics of birdsong. Trends Cogn Sci 15:113–121.

Boven E, Pemberton J, Chadderton P, Apps R, Costa RP (2022) Cerebro-cerebellar networks facilitate learning through feedback decoupling. bioRxiv 2022.01.28.477827.

Branco T, Häusser M (2010) The single dendritic branch as a fundamental functional unit in the nervous system. Curr Opin Neurobiol 20:494–502.

Breton-Provencher V, Drummond GT, Feng J, Li Y, Sur M (2022) Spatiotemporal dynamics of noradrenaline during learned behaviour. Nature 606:732–738.

Burkett ZD, Day NF, Peñagarikano O, Geschwind DH, White SA (2015) VoICE: a semi-automated pipeline for standardizing vocal analysis across models. Sci Rep 5:10237.

Chavlis S, Poirazi P (2021) Drawing inspiration from biological dendrites to empower artificial neural networks. Curr Opin Neurobiol 70:1–10.

Christensen AJ, Ott T, Kepecs A (2022) Cognition and the single neuron: How cell types construct the dynamic computations of frontal cortex. Curr Opin Neurobiol 77:102630.

Coffey KR, Marx RG, Neumaier JF (2019) DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacology 44:859–868.

Cohen Y, Nicholson DA, Sanchioni A, Mallaber EK, Skidanova V, Gardner TJ (2022) Automated annotation of birdsong with a neural network that segments spectrograms. Elife 11:e63853.

Cohen Y, Shen J, Semu D, Leman DP, Liberti WA, Perkins LN, Liberti DC, Kotton DN, Gardner TJ (2020) Hidden neural states underlie canary song syntax. Nature 582:539–544.

Cox DD, Dean T (2014) Neural networks and neuroscience-inspired computer vision. Curr Biol 24:R921–R929.

Crandall SR, Cruikshank SJ, Connors BW (2015) A corticothalamic switch: controlling the thalamus with dynamic synapses. Neuron 86:768–782.

Cunningham JP, Yu BM (2014) Dimensionality reduction for large-scale neural recordings. Nat Neurosci 17:1500–1509.

d'Aquin S, Szonyi A, Mahn M, Krabbe S, Gründemann J, Lüthi A (2022) Compartmentalized dendritic plasticity during associative learning. Science 376:eabf7052.

Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P (2009) Automated monitoring and analysis of social behavior in *Drosophila*. Nat Methods 6:297–303.

Daou A, Johnson F, Wu W, Bertram R (2012) A computational tool for automated large-scale analysis and measurement of bird-song syntax. J Neurosci Methods 210:147–160.

Datta SR, Anderson DJ, Branson K, Perona P, Leifer A (2019) Computational neuroethology: a call to action. Neuron 104:11–24.

Driscoll L, Shenoy K, Sussillo D (2022) Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. bioRxiv 2022.08.15.503870.

Dubreuil A, Valente A, Beiran M, Mastrogiuseppe F, Ostojic S (2022) The role of population structure in computations through neural dynamics. Nat Neurosci 25:783–794.

Duncker L, Sahani M (2021) Dynamics on the manifold: identifying computational dynamical activity from neural population recordings. Curr Opin Neurobiol 70:163–170.

Ebitz RB, Hayden BY (2021) The population doctrine in cognitive neuroscience. Neuron 109:3055–3068.

Economo MN, Viswanathan S, Tasic B, Bas E, Winnubst J, Menon V, Graybuck LT, Nguyen TN, Smith KA, Yao Z, Wang L, Gerfen CR, Chandrashekar J, Zeng H, Looger LL, Svoboda K (2018) Distinct descending motor cortex pathways and their roles in movement. Nature 563:79–84.

Fleming SM, Maniscalco B, Ko Y, Amendi N, Ro T, Lau H (2015) Action-specific disruption of perceptual confidence. Psychol Sci 26:89–98.

Flesch T, Juechems K, Dumbalska T, Saxe A, Summerfield C (2022) Orthogonal representations for robust context-dependent task performance in brains and neural networks. Neuron 110:1258–1270.e11.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349.

Fusi S, Miller EK, Rigotti M (2016) Why neurons mix: high dimensionality for higher cognition. Curr Opin Neurobiol 37:66–74.

Gawlikowski J, Tassi CR, Ali M, Lee J, Humt M, Feng J, Kruspe AM, Triebel R, Jung P, Roscher R, Shahzad M, Yang W, Bamler R, Zhu XX (2021) A survey of uncertainty in deep neural networks. arXiv:2107.03342.

Gilpin W, Huang Y, Forger DB (2020) Learning dynamics from large biological data sets: machine learning meets systems biology. Curr Opin Syst Biol 22:1–7.

Goffinet J, Brudner S, Mooney R, Pearson J (2021) Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. Elife 10:e67855.

Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial networks. arXiv: 2107.03342.

Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, Couzin ID (2019) DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. Elife 8:e47994.

Graving JM, Couzin ID (2020) VAE-SNE: a deep generative model for simultaneous dimensionality reduction and clustering. bioRxiv 207993. https://doi.org/10.1101/2020.07.17.207993.

Greedy W, Zhu HW, Pemberton J, Mellor J, Costa RP (2022) Single-phase deep learning in cortico-cortical networks. arXiv:2206.11769.

Griffiths TL, Callaway F, Chang MB, Grant E, Krueger PM, Lieder F (2019) Doing more with less: meta-reasoning and meta-learning in humans and machines. Curr Opin Behav Sci 29:24–30.

Guggenmos M, Wilbertz G, Hebart MN, Sterzer P (2016) Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. Elife 5:e13388.

Hayhoe MM (2018) Davida Teller Award Lecture 2017: what can be learned from natural behavior? J Vis 18:10.

Haykin SS (1994) Neural networks: a comprehensive foundation. New York: Maxwell Macmillan.

Hedley RW (2016) Complexity, predictability and time homogeneity of syntax in the songs of Cassin's vireo (*Vireo cassinii*). PLoS One 11:e0150822.

Hirokawa J, Vaughan A, Masset P, Ott T, Kepecs A (2019) Frontal cortex neuron types categorically encode single decision variables. Nature 576:446–451.

Hocker DL, Brody CD, Savin C, Constantinople CM (2021) Subpopulations of neurons in lOFC encode previous and current rewards at time of choice. Elife 10:e70129.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79:2554–2558.

Jazayeri M, Ostojic S (2021) Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. Curr Opin Neurobiol 70:113–120.

Jin DZ, Kozhevnikov AA (2011) A compact statistical model of the song syntax in Bengalese finch. PLoS Comput Biol 7:e1001108.

Johnston D, Narayanan R (2008) Active dendrites: colorful wings of the mysterious butterflies. Trends Neurosci 31:309–316.

Jones EG (2001) The thalamic matrix and thalamocortical synchrony. Trends Neurosci 24:595–601.

Josselyn SA, Tonegawa S (2020) Memory engrams: recalling the past and imagining the future. Science 367:eaaw4325.

Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K (2013) JAABA: interactive machine learning for automatic annotation of animal behavior. Nat Methods 10:64–67.

Kepecs A, Mainen ZF (2012) A computational framework for the study of confidence in humans and animals. Philos Trans R Soc Lond B Biol Sci 367:1322–1337.

Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455:227–231.

Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput Biol 10:e1003915.

Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi XL, Romo R, Uchida N, Machens CK (2016) Demixed principal component analysis of neural population data. Elife 5:e10989.

Kogan JA, Margoliash D (1998) Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. J Acoust Soc Am 103:2185–2196.

Kohlsdorf D, Herzing D, Starner T (2020) An auto encoder for audio dolphin communication. International Joint Conference on Neural Networks IJCNN, pp 1–7. IEEE.

Kollmorgen S, Hahnloser RH, Mante V (2020) Nearest neighbours reveal fast and slow components of motor learning. Nature 577:526–530.

Koumura T, Okanoya K (2016) Automatic recognition of element classes and boundaries in the birdsong with variable sequences. PLoS One 11: e0159188.

Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA, Poeppel D (2017) Neuroscience needs behavior: correcting a reductionist bias. Neuron 93:480–490.

Kubilius J, Schrimpf M, Kar K, Hong H, Majaj NJ, Rajalingham R, Issa EB, Bashivan P, Prescott-Roy J, Schmidt K, Nayebi A, Bear D, Yamins DL, DiCarlo JJ (2019) Brain-like object recognition with high-performing shallow recurrent ANNs. Advances in neural information processing systems, 32.

Kudithipudi D, et al. (2022) Biological underpinnings for lifelong learning machines. Nat Mach Intell 4:196–210.

Lak A, Costa GM, Romberg E, Koulakov AA, Mainen ZF, Kepecs A (2014) Orbitofrontal cortex is required for optimal waiting based on decision confidence. Neuron 84:190–201.

Langdon C, Engel TA (2022) Latent circuit inference from heterogeneous neural responses during cognitive tasks. bioRxiv 2022.01.23.477431.

Larkum M (2013) A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. Trends Neurosci 36:141–151.

Larkum ME (2022) Are dendrites conceptually useful? Dendritic Contrib Biol Artif Comput 489:4–14.

Le Bé JV, Markram H (2006) Spontaneous and evoked synaptic rewiring in the neonatal neocortex. Proc Natl Acad Sci USA 103:13214–13219.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444.

Leonardo A, Fee MS (2005) Ensemble coding of vocal control in birdsong. J Neurosci 25:652–661.

Li HH, Ma WJ (2020) Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. Nat Commun 11:2004.

Li M, Han Y, Aburn MJ, Breakspear M, Poldrack RA, Shine JM, Lizier JT (2019) Transitions in information processing dynamics at the whole-brain network level are driven by alterations in neural gain. PLoS Comput Biol 15:e1006957.

Lindsay GW (2021) Convolutional neural networks as a model of the visual system: past, present, and future. J Cogn Neurosci 33:2017–2031.

Lindsay GW (2022) Testing the tools of systems neuroscience on artificial neural networks. arXiv:2202.07035.

Lindsay GW, Mrsic-Flogel TD, Sahani M (2022) Bio-inspired neural networks implement different recurrent visual processing strategies than task-trained ones do. bioRxiv 483196. https://doi.org/10.1101/2022.03.07.483196.

Linsley D, Kim J, Veerabadran V, Windolf C, Serre T (2018) Learning long-range spatial dependencies with horizontal gated recurrent units. Advances in neural information processing systems, 31.

Lisman J, Cooper K, Sehgal M, Silva AJ (2018) Memory formation depends on both synapse-specific modifications of synaptic strength and cell-specific increases in excitability. Nat Neurosci 21:309–314.

Luxem K, Mocellin P, Fuhrmann F, Kürsch J, Remy S, Bauer P (2022) Identifying behavioral structure from deep variational embeddings of animal motion. bioRxiv 095430. https://doi.org/10.1101/2020.05.14.095430.

Machado AS, Darmohray DM, Fayad J, Marques HG, Carey MR (2015) A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. Elife 4:e07892.

Malagon-Vina H, Ciocchi S, Passecker J, Dorffner G, Klausberger T (2018) Fluid network dynamics in the prefrontal cortex during multiple strategy switching. Nat Commun 9:309.

Mamassian P (2022) Uncertain perceptual confidence. Nat Hum Behav 6:179–180.

Maniscalco B, Peters MA, Lau H (2016) Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. Atten Percept Psychophys 78:923–937.

Maniscalco B, Odegaard B, Grimaldi P, Cho SH, Basso MA, Lau H, Peters MA (2021) Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. PLoS Comput Biol 17:e1008779.

Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503:78–84.

Marblestone AH, Wayne G, Kording KP (2016) Toward an integration of deep learning and neuroscience. Front Comput Neurosci 10:94.

Markowitz JE, Ivie E, Kligler L, Gardner TJ (2013) Long-range order in canary song. PLoS Comput Biol 9:e1003052.

Marshall JD, Aldarondo DE, Dunn TW, Wang WL, Berman GJ, Ölveczky BP (2021) Continuous whole-body 3D kinematic recordings across the rodent behavioral repertoire. Neuron 109:420–437.e8.

Marullo C, Agliari E (2020) Boltzmann machines as generalized Hopfield networks: a review of recent results and outlooks. Entropy Basel Switz 23:34.

Masset P, Ott T, Lak A, Hirokawa J, Kepecs A (2020) Behavior- and modality-general representation of confidence in orbitofrontal cortex. Cell 182:112–126.e18.

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M (2018) DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 21:1281–1289.

McCullough MH, Goodhill GJ (2021) Unsupervised quantification of naturalistic animal behaviors for gaining insight into the brain. Curr Opin Neurobiol 70:89–100.

McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133.

McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426.

Mearns DS, Donovan JC, Fernandes AM, Semmelhack JL, Baier H (2020) Deconstructing hunting behavior reveals a tightly coupled stimulus-response loop. Curr Biol 30:54–69.e9.

Mei J, Muller E, Ramaswamy S (2022) Informing deep neural networks by multiscale principles of neuromodulatory systems. Trends Neurosci 45:237–250.

Mets DG, Brainard MS (2018) An automated approach to the quantitation of vocalizations and vocal learning in the songbird. PLoS Comput Biol 14:e1006437.

Michel M, Peters MA (2021) Confirmation bias without rhyme or reason. Synthese 199:2757–2772.

Minsky M, Papert S (1972) Artificial Intelligence Progress Report. AI Memo 252. Cambridge, MA.

Mishra P, Narayanan R (2021) Stable continual learning through structured multiscale plasticity manifolds. Curr Opin Neurobiol 70:51–63.

Morfi V, Lachlan RF, Stowell D (2021) Deep perceptual embeddings for unlabelled animal sound events. J Acoust Soc Am 150:2–11.

Mukherjee A, Lam NH, Wimmer RD, Halassa MM (2021) Thalamic circuits for independent control of prefrontal signal and noise. Nature 600:100–104.

Namboodiri VM, Otis JM, van Heeswijk K, Voets ES, Alghorazi RA, Rodriguez-Romaguera J, Mihalas S, Stuber GD (2019) Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. Nat Neurosci 22:1110–1121.

Nayebi A, Sagastuy-Brena J, Bear DM, Kar K, Kubilius J, Ganguli S, Sussillo D, DiCarlo JJ, Yamins DL (2021) Goal-driven recurrent neural network models of the ventral visual stream. bioRxiv 431717. https://doi.org/10.1101/2021.02.17.431717.

Nicholson D (2016) Comparison of machine learning methods applied to birdsong element classification. In: Proceedings of the 15th Python in Science Conference (Benthall S, Rostrup S, eds), pp 57–61.

Nicholson D (2021) NickleDave/hybrid-vocal-classifier. 0.3.0Github. Available at https://github.com/NickleDave/hybrid-vocal-classifier.git.

Ogawa M, van der Meer MA, Esber GR, Cerri DH, Stalnaker TA, Schoenbaum G (2013) Risk-responsive orbitofrontal neurons track acquired salience. Neuron 77:251–258.

Onken A, Xie J, Panzeri S, Padoa-Schioppa C (2019) Categorical encoding of decision variables in orbitofrontal cortex. PLoS Comput Biol 15:e1006667.

Ott T, Masset P, Kepecs A (2018) The neurobiology of confidence: from beliefs to neurons. Cold Spring Harb Symp Quant Biol 83:9–16.

Padamsey Z, Katsanevaki D, Dupuy N, Rochefort NL (2022) Neocortex saves energy by reducing coding precision during food scarcity. Neuron 110:280–296.e10.

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223–226.

Pagliarini S, Trouvain N, Leblois A, Hinaut X (2021) What does the canary say? Low-dimensional GAN applied to birdsong. hal-03244723v2.

Pandarinath C, Ames KC, Russo AA, Farshchian A, Miller LE, Dyer EL, Kao JC (2018) Latent factors and dynamics in motor cortex and their application to brain–machine interfaces. J Neurosci 38:9390–9401.

Payeur A, Guerguiev J, Zenke F, Richards BA, Naud R (2021) Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. Nat Neurosci 24:1010–1019.

Pemberton JO, Boven E, Apps R, Costa RP (2021) Cortico-cerebellar networks as decoupling neural interfaces. Advances in neural information processing systems 34:7745–7759.

Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SS, Murthy M, Shaevitz JW (2019) Fast animal pose estimation using deep neural networks. Nat Methods 16:117–125.

Pereira TD, Shaevitz JW, Murthy M (2020) Quantifying behavior to understand the brain. Nat Neurosci 23:1537–1549.

Peters MA (2022) Confidence in Decision-Making. Oxford Research Encyclopedia of Neuroscience.

Poirazi P, Papoutsi A (2020) Illuminating dendritic function with computational models. Nat Rev Neurosci 21:303–321.

Pouget A, Drugowitsch J, Kepecs A (2016) Confidence and certainty: distinct probabilistic quantities for different goals. Nat Neurosci 19:366–374.

Ptasczynski LE, Steinecker I, Sterzer P, Guggenmos M (2022) The value of confidence: confidence prediction errors drive value-based learning in the absence of external feedback. PLOS Comput Biol 18:e1010580.

Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434.

Ramaswamy S, Markram H (2015) Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron. Front Cell Neurosci 9:233.

Rao SC, Rainer G, Miller EK (1997) Integration of what and where in the primate prefrontal cortex. Science 276:821–824.

Ratté S, Hong S, De Schutter E, Prescott SA (2013) Impact of neuronal properties on network coding: roles of spike initiation dynamics and robust synchrony transfer. Neuron 78:758–772.

Rausch M, Hellmann S, Zehetleitner M (2018) Confidence in masked orientation judgments is informed by both evidence and visibility. Atten Percept Psychophys 80:134–154.

Reed S, et al. (2022) A generalist agent. arXiv:2205.06175.

Richards BA, et al. (2019) A deep learning framework for neuroscience. Nat Neurosci 22:1761–1770.

Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. Nature 497:585–590.

Roelfsema PR, Holtmaat A (2018) Control of synaptic plasticity in deep cortical networks. Nat Rev Neurosci 19:166–180.

Roscow EL, Chua R, Costa RP, Jones MW, Lepora N (2021) Learning offline: memory replay in biological and artificial reinforcement learning. Trends Neurosci 44:808–821.

Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65:386–408.

Rudolph M, Destexhe A (2003) Tuning neocortical pyramidal neurons between integrators and coincidence detectors. J Comput Neurosci 14:239–251.

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536.

Sacramento J, Ponte Costa R, Bengio Y, Senn W (2018) Dendritic cortical microcircuits approximate the backpropagation algorithm. Advances in neural information processing systems, 31.

Sainburg T, Gentner TQ (2021) Toward a computational neuroethology of vocal communication: from bioacoustics to neurophysiology, emerging tools and future directions. Front Behav Neurosci 15:811737.

Sainburg T, Thielk M, Theilman B, Migliori B, Gentner T (2018) Generative adversarial interpolative autoencoding: adversarial training on latent space interpolations encourage convex latent distributions. arXiv:1807.06650.

Sainburg T, Thielk M, Gentner TQ (2020) Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. PLoS Comput Biol 16:e1008228.

Sainburg T, McInnes L, Gentner TQ (2021) Parametric UMAP embeddings for representation and semisupervised learning. Neural Comput 33:2881–2907.

Saxena S, Cunningham JP (2019) Towards the neural population doctrine. Curr Opin Neurobiol 55:103–111.

Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117.

Sejnowski TJ (2020) The unreasonable effectiveness of deep learning in artificial intelligence. Proc Natl Acad Sci USA 117:30033–30038.

Sejnowski TJ, Churchland PS, Movshon JA (2014) Putting big data to good use in neuroscience. Nat Neurosci 17:1440–1441.

Shine JM (2021) The thalamus integrates the macrosystems of the brain to facilitate complex, adaptive brain network dynamics. Prog Neurobiol 199:101951.

Shine JM, Müller EJ, Munn B, Cabral J, Moran RJ, Breakspear M (2021) Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics. Nat Neurosci 24:765–776.

Sober SJ, Brainard MS (2009) Adult birdsong is actively maintained by error correction. Nat Neurosci 12:927–931.

Sober SJ, Brainard MS (2012) Vocal learning is constrained by the statistics of sensorimotor experience. Proc Natl Acad Sci USA 109:21099–21103.

Sober SJ, Wohlgemuth MJ, Brainard MS (2008) Central contributions to acoustic variation in birdsong. J Neurosci 28:10370–10379.

Sohn H, Narain D, Meirhaeghe N, Jazayeri M (2019) Bayesian computation through cortical latent dynamics. Neuron 103:934–947.e5.

Song HF, Yang GR, Wang XJ (2016) Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework. PLoS Comput Biol 12:e1004792.

Spoerer CJ, McClure P, Kriegeskorte N (2017) Recurrent convolutional neural networks: a better model of biological object recognition. Front Psychol 8:1551.

Spruston N (2008) Pyramidal neurons: dendritic structure and synaptic integration. Nat Rev Neurosci 9:206–221.

Steinfath E, Palacios-Muñoz A, Rottschäfer JR, Yuezak D, Clemens J (2021) Fast and accurate annotation of acoustic signals with deep neural networks. Elife 10:e68837.

Stephens GJ, Johnson-Kerner B, Bialek W, Ryu WS (2008) Dimensionality and dynamics in the behavior of C. elegans. PLoS Comput Biol 4: e1000028.

Stolyarova A, Rakhshan M, Hart EE, O'Dell TJ, Peters MA, Lau H, Soltani A, Izquierdo A (2019) Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. Nat Commun 10:4704.

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) Spontaneous behaviors drive multidimensional, brainwide activity. Science 364:eaav7893.

Stroud JP, Porter MA, Hennequin G, Vogels TP (2018) Motor primitives in space and time via targeted gain modulation in cortical networks. Nat Neurosci 21:1774–1783.

Sutskever I, Martens J, Hinton G (2011) Generating text with recurrent neural networks. In: Proceedings of the 28th International Conference on International Conference on Machine Learning, pp 1017–1024. Madison, WI: Omnipress.

Tachibana RO, Oosugi N, Okanoya K (2014) Semi-automatic classification of birdsong elements using a linear support vector machine. PLoS One 9: e92584.

Takahashi N, Oertner TG, Hegemann P, Larkum ME (2016) Active cortical dendrites modulate perception. Science 354:1587–1590.

Tasic B, et al. (2018) Shared and distinct transcriptomic cell types across neocortical areas. Nature 563:72–78.

Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP (2000) A procedure for an automated measurement of song similarity. Anim Behav 59:1167–1176.

Tenenbaum JB, Silva VD, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323.

Terra H, Bruinsma B, de Kloet SF, van der Roest M, Pattij T, Mansvelder HD (2020) Prefrontal cortical projection neurons targeting dorsomedial striatum control behavioral inhibition. Curr Biol 30:4188–4200.e5.

Thorat S, Aldegheri G, Kietzmann TC (2021) Category-orthogonal object features guide information processing in recurrent neural networks trained for object categorization. arXiv:2111.07898.

Titley HK, Brunel N, Hansel C (2017) Toward a neurocentric view of learning. Neuron 95:19–32.

Tsuda B, Pate SC, Tye KM, Siegelmann HT, Sejnowski TJ (2022) Neuromodulators generate multiple context-relevant behaviors in a recurrent neural network by shifting activity hypertubes. bioRxiv 446462. https://doi.org/10.1101/2021.05.31.446462.

van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605.

Vyas S, Golub MD, Sussillo D, Shenoy KV (2020) Computation through neural population dynamics. Annu Rev Neurosci 43:249–275.

Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. Nature 411:953–956.

Wang J, Narain D, Hosseini EA, Jazayeri M (2018) Flexible timing by temporal scaling of cortical responses. Nat Neurosci 21:102–110.

Werbos PJ (1974) Beyond regression: new tools for prediction and analysis in the behavioral sciences. PhD dissertation, Harvard University.

Werbos PJ (1982) Applications of advances in nonlinear sensitivity analysis. In: System modeling and optimization (Drenick RF, Kozin F, eds), pp 762–770. Berlin: Springer.

Williams LE, Holtmaat A (2019) Higher-order thalamocortical inputs gate synaptic long-term potentiation via disinhibition. Neuron 101:91–102.e4.

Wiltschko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, Abraira VE, Adams RP, Datta SR (2015) Mapping sub-second structure in mouse behavior. Neuron 88:1121–1135.

Winnubst J, et al. (2019) Reconstruction of 1,000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain. Cell 179:268–281.e13.

Wohlgemuth MJ, Sober SJ, Brainard MS (2010) Linked control of syllable sequence and phonology in birdsong. J Neurosci 30:12936–12949.

Wyatte D, Curran T, O'Reilly R (2012) The limits of feedforward vision: recurrent processing promotes robust object recognition when objects are degraded. J Cogn Neurosci 24:2248–2261.

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci USA 111:8619–8624.

Yan Y, Goodman JM, Moore DD, Solla SA, Bensmaia SJ (2020) Unexpected complexity of everyday manual behaviors. Nat Commun 11:3564.

Yang GR, Joglekar MR, Song HF, Newsome WT, Wang XJ (2019) Task representations in neural networks trained to perform many cognitive tasks. Nat Neurosci 22:297–306.