Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Modeling Normal and Abnormal Circuit Development with Recurrent Neural Networks

Daniel Zavitz,[1] ShiNung Ching,[2] and Geoffrey Goodhill[1]

[1]Departments of Developmental Biology and Neuroscience, Washington University in St. Louis School of Medicine, St. Louis, Missouri 63110, USA

[2]Departments of Electrical and Systems Engineering and Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA

Correspondence: g.goodhill@wustl.edu

Neural development must construct neural circuits that can perform the computations necessary for survival. However, many theoretical models of development do not explicitly address the computational goals of the resulting networks, or computations that evolve in time. Recurrent neural networks (RNNs) have recently come to prominence as both models of neural circuit computation and building blocks of powerful artificial intelligence systems. Here, we review progress in using RNNs for understanding how developmental processes lead to effective computations, and how abnormal development disrupts these computations.

The development of a functioning nervous system proceeds through multiple complex stages, including neurulation, regional specification, neurogenesis, cell fate determination, cell migration, axon guidance and dendritic development, and synapse formation and pruning, followed by ongoing plasticity and refinement. Across different species, the timescales involved in these processes range from hours to years. Interesting theoretical problems abound at all these stages (van Ooyen 2011; Goodhill 2018). For instance, understanding how molecular and/or mechanical cues promote tissue folding, regional identity, and axon guidance, especially given the presence of unavoidable noise in the measurement of concentration (e.g., Gregor et al. 2007; Bicknell et al. 2015; Tkačik et al. 2015). On the other hand, research in the neural network community has focused mostly on the formation of connections between neurons, and in particular on how connection strengths are coded/learned within a fixed architecture so that the network performs specific computations.

Mathematical instantiations of Hebb's rule (Hebb 1949) have proven effective at reproducing receptive field structures found in early sensory areas such as the visual system in response to appropriate activity patterns (von der Malsburg 1973; Linsker 1986; Goodhill 1993), including the effects of altered sensory input such as monocular deprivation (Miller et al. 1989). These approaches are unsupervised, and attempt to address developmental processes fairly explicitly. In contrast in supervised approaches, learning takes place by giving the network a target output for every input pattern and adjusting the weights

Editors: Laura C. Andreae, Justus M. Kebschull, and Anthony M. Zador
Additional Perspectives on Evolution and Development of Neural Circuits available at www.cshperspectives.org

Advanced Online Article. Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a041507

in the network to achieve this outcome, usually via backpropagation (Rumelhart et al. 1986). An early example of the application of backpropagation to understand neural circuits was the reproduction of neural response properties in area 7a of the posterior parietal cortex of monkeys (Zipser and Andersen 1988). A more recent example is Yamins et al. (2014), which matched response properties through many layers of a deep network with those in the primate visual system. Supervised approaches show how learning a particular computational task can lead to structure matching that found biologically, but do not claim to reproduce actual developmental processes.

Until recently supervised learning approaches to understanding neural computation were primarily focused on layered, feedforward networks, where inputs and outputs do not have a time dimension (e.g., the examples above). However, in the past few years, recurrent neural networks (RNNs) have gained prominence through their ability to accept time-varying inputs and produce time-varying outputs. Such networks have proved very useful for understanding computations involving the accrual of information over time, such as when animals are trained to respond in specific ways to cues occurring at specific times. This opens up a much more complex repertoire of task possibilities. While the focus in this regard has most commonly been on adult learning occurring on relatively rapid timescales, the question naturally arises of whether such networks can also be applied to understand the emergence of network computations on developmental timescales. We argue that this provides a promising direction for understanding both normal and abnormal development, but that several important issues remain to addressed.

RECURRENT NEURAL NETWORKS

Basic Neurobiological Components of RNNs

RNNs can be interpreted as firing rate models of neural circuits (Fig. 1; Sussillo 2014). This approach allows RNNs to approximate the differ-

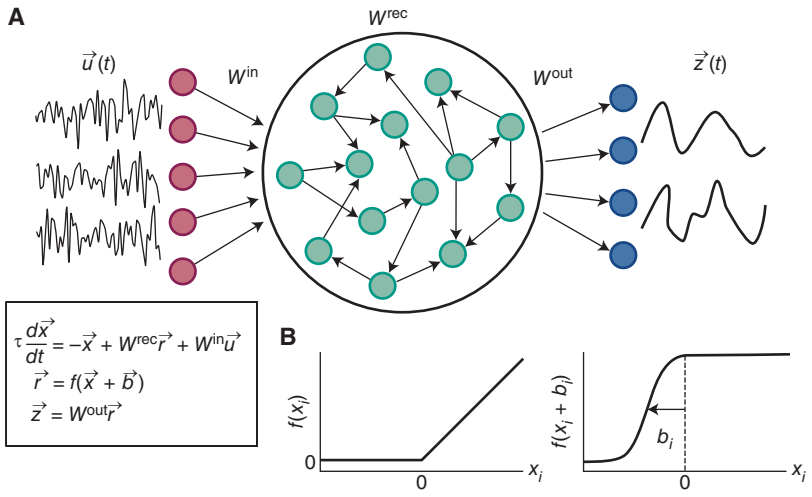

Figure 1. Basic principles of recurrent neural networks (RNNs). (*A*) An RNN transforms time-varying inputs $\vec{u}(t)$ into time-varying outputs $\vec{z}(t)$ using firing rate–type dynamics. The neurons' synaptic current $\vec{x}(t)$ and firing rate $\vec{r}(t)$ represent the RNNs' internal state that captures long-term relationships between inputs and outputs and τ sets the timescale of individual neurons' activity. Trainable parameters include the input W^{in}, recurrent W^{rec}, and output W^{out} weight matrices and the neurons' biases $\vec{b}$. (*B*) Currents are transformed into firing rates by an activation function (or *F–I* curve). Popular choices include nonsaturating functions like a rectified linear unit (ReLu, *left*) or saturating functions like tanh (·) (*right*). The bias b_i determines the amount of current required to activate the neuron.

ential equations of a typical firing rate model (Dayan and Abbott 2005):

$$\begin{aligned} \tau\frac{d\vec{x}}{dt} &= -\vec{x} + W^{rec}\vec{r} + W^{in}\vec{u}, \\ \vec{r} &= f(\vec{x} + \vec{b}), \\ \vec{y} &= W^{out}\vec{r}, \end{aligned} \tag{1}$$

where each entry i of the vector $\vec{x}$ is the net, low-pass filtered synaptic current of neuron i, $\vec{r}$ are corresponding instantaneous firing rates, and τ describes the timescale of activity of individual neurons. The feedforward firing rate input and readout of the system are $\vec{u}$ and $\vec{y}$, respectively. The activation function $f(\cdot)$ can be interpreted as an *F–I* curve (Izhikevich 2006) that mediates the relationship between injected current (I) and neuronal firing rate (F). Trainable parameters include W^{rec}, W^{in}, and W^{out}, the weight matrices that determine the efficacies of recurrent, input, and readout connections, respectively, and $\vec{b}$, the neurons' biases that set their activation threshold (analogous to rheobase). The resulting model is typically used to process sequential data such that the state of each neuron at time t is driven by the input data at t and previous neural states. Thus, RNNs have "memories" in which sustained activity represents information from prior input data by connection weights. This feature is key to learning long-term dependencies in sequence data.

Following the McCulloch–Pitts model (McCulloch and Pitts 1943), $f(\cdot)$ is a threshold-like activation function. Common choices for $f(\cdot)$ can be roughly divided into two categories: sigmoid functions that saturate as current increases (e.g., tanh (·); Fig. 1B, right) and those that do not (e.g., rectified linear units—ReLu; Fig. 1B, left). *F–I* curves of both types can be found in vivo (McCormick et al. 1985; Guan et al. 2014) and can change during development (Oswald and Reyes 2008). The functional significance of the specific activation function used remains poorly understood, hence a common approach is to analyze many RNNs with different activation functions (e.g., Yang et al. 2019; Driscoll et al. 2022). Note that this *F–I* curve classification differs from the taxonomy of Hodgkin–Huxley models (Hodgkin and Huxley 1952) in which neural dynamics are classified by the bifurcation at the activation threshold (Rinzel and Ermentrout 1989; Izhikevich 2006).

Although synaptic connection weights are primarily determined by training (discussed later), biological features can be imposed by constraining their values. For instance, an RNN can be divided into distinct excitatory and inhibitory subpopulations (Song et al. 2016). While a neuron can be made excitatory or inhibitory by simply giving all its outgoing connections the appropriate sign a priori, when weights are determined only by a training algorithm, neurons have outgoing connections of both signs. Recent work by Song et al. (2016) bypasses this problem by decomposing weight matrices such that $W = W^{+}M$, where the connection strength is determined by training the nonnegative matrix W^{+} and excitatory/inhibitory identity is fixed by a prior in the diagonal matrix M. This allows standard training algorithms to optimize the network weights while respecting the subpopulation identities.

Additional constraints can also be added to RNN circuit models (Molano-Mazón et al. 2023). Advances in large-scale neural recordings (Urai et al. 2022) have revealed that neurons distributed across many brain regions act in concert to produce behaviors and process sensory information (e.g., Musall et al. 2019; Steinmetz et al. 2019; Stringer et al. 2019). Recent RNN models applied to this spatial scale have incorporated brain region organization by imposing a block structure (Holland et al. 1983) on recurrent connections in which intra-area connections are dense and inter-area connections are sparse (Pinto et al. 2019; Kleinman et al. 2020; Perich and Rajan 2020). Furthermore, the input and output connection weights (W^{in} and W^{out}) can be structured to mirror brain region organization (Song et al. 2016). This allows distinct "sensory" and "motor" areas where external "sensory" input is routed to a subpopulation of neurons and the outgoing "motor" response is read exclusively from a different subpopulation.

Once the architecture of an RNN circuit model is determined, it can be trained to perform a specific task. This is often formulated as a supervised learning problem in which the data set consists of ordered pairs $D = \{(\vec{u}_i, \vec{y}_i^{label})\}_{i=1}^{N}$ and the RNN learns to produce a set of desired out-

puts $\vec{y}_i^{\text{label}}$ when driven by a corresponding input $\vec{u}_i$. Similar to feedforward networks, RNNs learn by adjusting their weights and biases. This can be formulated as an optimization problem:

$$\theta^* = \arg\min_{\theta} L(D; \theta), \tag{2}$$

where θ and θ^* represent all trainable parameters before and after training, respectively, typically W^{rec}, W^{in}, W^{out}, and $\vec{b}$, and L is a loss function that measures the differences between the RNN's actual and desired outputs. Loss is calculated for each item in the data set and parameters are adjusted to minimize the total loss. Despite their topological differences, the algorithms used to train both feedforward neural networks and RNNs are similar. Most notably, backpropagation can be implemented in RNNs as backpropagation-through-time by transforming the RNN into a deep feedforward network where each layer represents a timestep (Werbos 1990). Many additional training approaches are available including reinforcement learning (Song et al. 2017) and recursive least squares (Sussillo and Abbott 2009), and carry different assumptions that are amenable to specific types of tasks and can heavily influence the resulting neural dynamics (Mikhaeil et al. 2022).

Computations Mediated by RNNs

RNNs maintain an internal state that is ideal for performing computations involving sequences, and so can be trained to perform tasks with a temporal component (although see Wang et al. 2021). In a neuroscience context, RNNs have been used to study a wide variety of tasks, but cognition and motor control have been a particular focus (Vyas et al. 2020; Yang and Wang 2020). Within the cognitive task framework, RNNs have been trained to perform tasks that have been investigated experimentally (e.g., using a random dot motion approach) such as perceptual (Song et al. 2016) or context-dependent decision making (Mante et al. 2013) and working memory (Masse et al. 2019; Orhan and Ma 2019; Ghazizadeh and Ching 2021). The flexibility demonstrated by the cognitive task paradigm demonstrates the primary advantage of RNNs as a neural circuit model: A single training protocol can typically be reused to train an RNN on many tasks that would require a human to start over afresh when manually constructing circuit models.

RNNs' applicability to motor control tasks stems from a second, related advantage: RNNs are capable of approximating any dynamical system (Schäfer and Zimmermann 2006). For instance, hypotheses about the relationship between neural activity and motor control (Shenoy et al. 2013) have been explored by training RNNs to produce output signals that match experimental recordings of muscle actions and then comparing with corresponding neural recordings from the motor cortex (Hennequin et al. 2014; Sussillo et al. 2015; Saxena et al. 2022). Related brain–computer interface studies have demonstrated that RNNs can effectively and rapidly decode motor commands from neural activity (Sussillo et al. 2012; Saxena et al. 2022).

MODELING NEURAL CIRCUIT DEVELOPMENT WITH RNNS

Structurally Dynamic RNNs

In addition to the synaptic efficacy fine-tuning typical of adult learning, neurodevelopmental processes also include much broader-scale structural changes driven by neurogenesis, synaptogenesis, and pruning. RNN circuit models typically do not make this structure/efficacy distinction and implicitly assume that constituent neurons have all-to-all structural connections whose efficacy is tuned to perform specific computations. However, RNNs with structurally dynamic learning rules have granted insight into the functional significance of specific neurodevelopmental mechanisms, as we describe below (Fig. 2A–C).

Neurogenesis

The recurrent cascade correlation network (Fahlman 1990) was the first RNN to learn via neurogenesis. The network learns mappings between input and output sequences by iterating

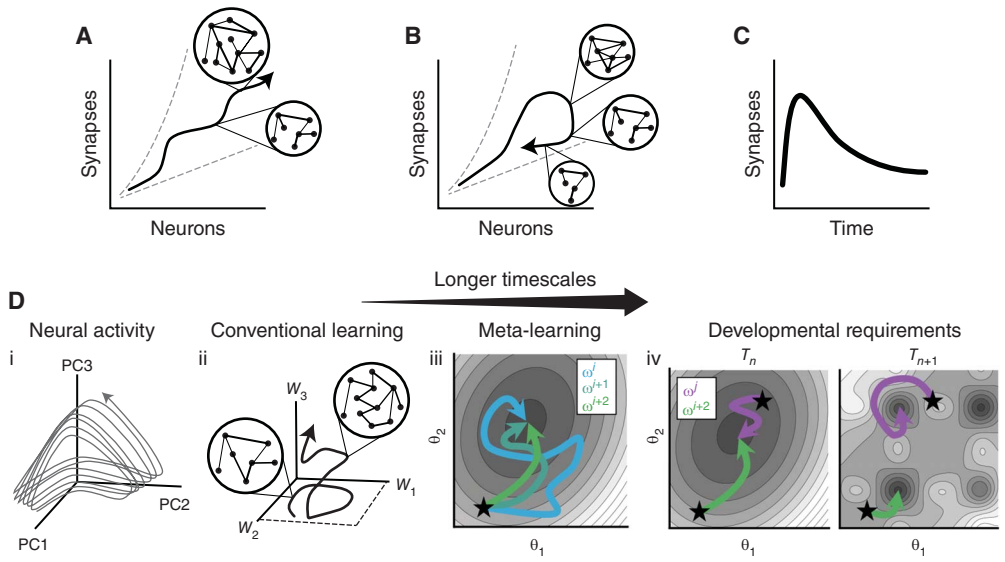

Figure 2. Recurrent neural network (RNN) models of neural development. (*A*) Some RNN models have incorporated structural dynamics into learning. Neurogenesis and synaptogenesis allow learning but restrict exploration of the possible search space (dashed lines indicate bounds on the number of synapses; Fahlman 1990). (*B*) In contrast grow-when-required learning allows more flexible model configurations by enabling synaptic pruning and apoptosis (Parisi et al. 2018). (*C*) Hebbian associative learning mediated by synaptogenesis and pruning can exhibit a phase of rapid growth followed by a phase of net synaptic pruning, which is characteristic of development (Millán et al. 2018). (*D*) Multiple timescales are relevant to neural computations and learning. (*i*) Computations are executed by neural activity at fast timescales. (*ii*) Conventional learning operates at an intermediate timescale. Learning can result both from traversing existing parameter space (e.g., the plane formed by synaptic weights w_1 and w_2) by adjusting synaptic efficacies and expanding parameter space through neuro and synaptogenesis (adding a new synapse with weight w_3). (*iii*) Metalearning occurs at a longer but still intermediate timescale where learning strategies, or metaknowledge ω^*, can be learned iteratively (blue to green arrows finding the minima of a loss function) resulting in more efficient conventional learning of model parameters θ_1, θ_2 (e.g., Goudar et al. 2023). (*iv*) At the longest timescales, the set of tasks needed to be performed is dynamic (T_n to T_{n+1}). Different structural priors (stars), possibly resulting from different learning strategies (ω^{i+2}, ω^i) developed over shorter timescales, may result in different solutions (Molano-Mazón et al. 2023).

over training data, each time reducing a loss function by adding an additional neuron with a recurrent self-connection and feedforward connections from existing neurons. This approach yields an RNN with the minimum number of neurons required to perform a computation when constructed using the specified rule. However, the resulting architecture is restricted to a small region of the possible search space and so has a limited capacity to learn.

A subsequent approach termed GNARL (generalized acquisition of recurrent links) used both neurogenesis and pruning to construct RNNs nonmonotonically with minimal restrictions to both RNN size and architecture (Angeline et al. 1994). This allowed the RNN to explore a much larger region of the parameter space by updating both its topology and its edge weights during learning. However, GNARL is an evolutionary algorithm in which populations of RNNs are initialized, ranked according to their ability to perform a computation (fitness), and then used to construct the next RNN generation according to this ranking, limiting its relevance to neural development.

A more recent approach used neurogenesis and pruning to implement lifelong learning (i.e., learning from a continuous stream of information) (Parisi et al. 2018). In neural networks, the primary obstacle to lifelong learning is cata-

strophic forgetting. The authors created a two-RNN circuit model based on complementary learning systems theory (McClelland et al. 1995) in which the hippocampus mediates episodic memories that are consolidated into semantic memories in the neocortex. Both RNNs learn via a grow-when-required algorithm (Marsland et al. 2002; Parisi et al. 2017) such that new neurons are created by a Hebbian learning rule only when necessary, and removed if they are no longer used. This allows the network to avoid catastrophic forgetting while adapting to nonstationary inputs. Together these RNN approaches illuminate the computational potential of apoptosis and synaptic pruning and their synergy with neural growth.

Synaptogenesis and Pruning

Neural development in many organisms including mammals can be characterized by initial rapid synaptic growth followed by a period of synaptic pruning that eventual plateaus at a more stable value (Huttenlocher 1979; Markus and Petit 1987; Bourgeois and Rakic 1993; White et al. 1997). Both activity-independent processes reliant on genetic, mechanical, and molecular mechanisms, and activity-dependent processes reliant on learning, sensory experiences, and spontaneous activity, are thought to play a role (Faust et al. 2021). Graph–theoretic analyses of neural development have indicated that overproduction-then-pruning algorithms can simultaneously enhance circuit robustness and efficiency, key measures of global network structure (Navlakha et al. 2018). However, the precise contributions of activity-dependent and independent processes to synaptic architecture development and their computational significance remain unclear.

Recent studies using RNNs have uncovered potential relationships between activity-dependent synaptogenesis and pruning and memory formation (Johnson et al. 2010; Millán et al. 2018, 2019, 2021). This approach modifies Hopfield networks, a classic RNN model capable of associative learning (Hopfield 1982), to learn via preferential attachment and detachment. More specifically, the RNN implements a stochastic, activity-dependent learning rule in which neurons gain and lose edges according to empirically derived probability distributions defined by the current each neuron receives and the number of synaptic connections in the RNN. The resulting RNNs exhibit synaptic overproduction-then-pruning that matches human and mouse data. Analysis revealed that strong coupling between activity and structure is necessary for memory formation, synaptic pruning can optimize neural circuits using local plasticity rules, and that an initial period of dense synaptic connectivity can enhance memory stability. Together, this approach synthesizes methods from machine learning, network science, and statistical physics to reveal how feedback between the dynamics of circuit structure and activity may interact throughout development to enhance computational performance.

Development of Circuit Function

Computational and behavioral requirements change over development. As a result, understanding the development of computational function is critical to understanding neural development. Complementary to the mechanism-first approaches described above, neural development can be investigated in a computation-first manner, where RNNs are trained to match the development of circuit function and then analyzed to reveal possible neurodevelopmental mechanisms. This approach leverages advances in machine learning to explore the functional principles of neural circuit development and produce hypotheses about the underlying mechanisms.

Development of Cognitive Function

Cognitive abilities improve during neural development. For example, monkeys exhibit an improvement in working memory tasks from adolescence to adulthood, that is accompanied by an increase in activity in prefrontal cortex neurons during the task's delay period (Zhou et al. 2016a). However, specific, causal relationships between neurobiological changes and cognitive improvement remain difficult to uncover. RNNs can be used to investigate this relationship by first rep-

licating specific computations across a developmental trajectory and then dissecting the underlying neural dynamics.

Recent work has taken this approach to link experimentally observed changes in neural activity to improvement in specific cognitive functions throughout development (Liu et al. 2021). The authors trained RNNs to match the performance of adolescent and adult monkeys (35%–65% and >65% correct trials, respectively) on working memory and response inhibition task variants (Zhou et al. 2013, 2016b). At each stage of development artificial and prefrontal cortex neural activity were compared to explore the extent to which computational optimization, rather than specific (possibly unrelated) biological processes, can explain changes in neural dynamics resulting from development. Intriguingly, RNN activity dynamics in both working memory and response inhibition tasks mirrored specific changes observed in prefrontal cortex neural activity and hypothesized to drive task improvement. Although not a full, causal explanation, this approach implicates the computational function of specific neurodevelopmental changes and demonstrates the utility of top-down RNN approaches to modeling neural circuit development. RNNs' flexibility allows this approach to serve as a promising template for exploring the computational relevance of other neurodevelopmental changes across model systems.

Modeling Considerations

Three key factors require special consideration when using RNN circuit models to investigate neural development. First, changes observed in circuit structure and activity during learning are not necessarily related to the task (Hennig et al. 2021). Even in a nondevelopmental context, neural representations in adult animals performing familiar tasks are well known to "drift" over time (Driscoll et al. 2017) due to many separate biological processes. The variety and scale of biological and cognitive changes present in neural development may obscure the relationship between RNN and neural circuit learning. Second, catastrophic forgetting (Ratcliff 1990; Kudithipudi et al. 2022) is a particular problem for modeling neural development since this involves learning a large amount of information over much longer time periods relative to learning problems considered in most RNN applications.

Third, neural circuits develop to concurrently perform a wide variety of tasks. Recent advances in multitask machine learning (Zhang and Yang 2018) enable individual RNN circuit models to learn many tasks (Yang et al. 2019). Understanding the distinctions between how tasks are represented within RNNs in these settings is an active area of research (Sucholutsky et al. 2023). Studies exploring the resulting neural circuits (in a nondevelopmental context) have observed compositionality in task variable representations: RNNs tend to develop functionally specialized subcircuits to execute computations that are reused when performing many tasks (Yang et al. 2019; Driscoll et al. 2022). However, not all task sets require this functional specialization, and may instead produce computations that are executed by the collective dynamics of all neurons (Dubreuil et al. 2022). Together, these studies indicate that "more is different" (Anderson 1972) and interactions between tasks within a task set may have a substantial impact on the computational mechanisms learned by RNNs. As a result, improving a task set's approximation of behavioral and computational requirements during development could be critical to constructing rich and informative theories about the underlying mechanisms of circuit development.

Learning at Multiple Timescales

Metalearning

Neural development involves learning at disparate timescales. For example, a specific task T (e.g., a single stimulus-reward association) can be learned on a shorter timescale, but recognizing commonalities in a sequence of tasks $\{T_1, T_2, T_3, \ldots\}$ may require learning over a much longer timescale. However, learning at different timescales is not an independent process. Rather, learning at longer timescales often benefits from ongoing increases in efficiency, indicating the presence of learning-to-learn (Harlow 1949).

Metalearning (Hospedales et al. 2021), offers a flexible framework to couple learning across different timescales by learning-to-learn (Fig. 2D). Conventional learning algorithms make specific assumptions about how to learn, including hyperparameter values of the learning algorithm or initial values of the model's parameters. The metalearning paradigm seeks to improve these assumptions. Formally, metalearning can be viewed as a nested optimization problem:

$$\theta^* = \arg\min_{\theta} L(D; \theta, \omega), \tag{3}$$

$$\omega^* = \arg\min_{\omega} L^{\text{meta}}(D; \theta^*(\omega), \omega). \tag{4}$$

The inner optimization problem (Equation 3) constitutes conventional supervised learning and ω represents an assumption of the learning algorithm (e.g., a specific hyperparameter). The outer, metalearning problem (Equation 4) seeks to optimize ω given θ^*, the parameters resulting from Equation 3, using its own learning algorithm and loss function L^{meta} (Hospedales et al. 2021; Wang 2021). Within this paradigm, learning at the fast (inner) timescale shapes learning at the slow (outer), developmentally relevant timescale, and vice versa.

Metalearning in RNNs

RNN metalearning studies can be categorized by their metarepresentations. Mirroring advances in machine learning (Finn et al. 2017), metalearning the RNN weight initialization has yielded insight into the computational process underlying learning over the disparate timescales relevant to neural development (Goudar et al. 2023; Molano-Mazón et al. 2023). Within this framework, RNNs are pretrained on a set of tasks $\{T_1, T_2, \ldots, T_n\}$ and then the RNN's ability to generalize to novel tasks $\{T_{n+1}, T_{n+2}, T_{n+3}, \ldots\}$ is assessed. The specific parameter configuration induced by pretraining corresponds to "structural priors," embodying preexisting knowledge learned over the course of development or possibly longer, evolutionary timescales (Zador 2019; Koulakov et al. 2022; Barabási et al. 2023). Recent analysis has demonstrated that RNNs with structural priors induced by pretraining on naturalistic tasks match suboptimal behavior exhibited by rats in a two alternative forced choice task (Molano-Mazón et al. 2023). In contrast, RNNs trained only on the two alternative force choice tasks exceeded the rats' performance. This indicates the importance of incorporating learning across different timescales in producing biologically relevant RNN models, but does not differentiate between development and evolution. From a neural manifold perspective, metalearning weight initializations have been shown to enable rapid generalization to new problems through the construction schema (Goudar et al. 2023), neural representations that abstract commonalities across previous experience and play a key role in developmental psychology (Piaget 2005). This metalearned solution minimizes the weight changes necessary to learn additional tasks, thus linking learning acceleration to possible wiring constraints resulting from the biophysical mechanisms of neural development.

From both a neurobiological (Doya 2002) and machine learning perspective (Li et al. 2017), the learning algorithm's hyperparameters (e.g., the learning rate) are a natural metarepresentation choice. Neuromodulators have been shown to selectively regulate learning across a wide variety of circumstances (Marder 2011). Work on the neural basis of reward-driven learning has linked neuromodulators to specific parameters and hyperparameters in reinforcement and metareinforcement learning algorithms (Schultz et al. 1997; Doya 2002). More recently, neuromodulation has been explicitly incorporated into RNN circuit models as an additional mechanism of experience-dependent plasticity that modulates neural activity throughout tasks (Wang et al. 2018; Jiang and Litwin-Kumar 2021). In this approach, RNN weights are trained via gradient descent to produce computations that track a changing environment using neuromodulation mechanisms. Consequently, the standard roles of synaptic weights as parameters and neuromodulators as hyperparameters are effectively transposed. RNN weights act as hyperparameters that are metalearned on a long, neural development-relevant timescale, while neuro-

modulation mechanisms implement parameters that learn by integrating reward information on a short, behaviorally relevant timescale.

RNN MODELS OF NEURODEVELOPMENTAL DISORDERS

A great variety of mental health conditions are neurodevelopmental in origin. Examples include schizophrenia, autism spectrum disorder (ASD), and attention-deficit/hyperactivity disorder. These cause deficits spanning a wide range of domains, including social, emotional, communicative, intellectual, and sensory functions (Morris-Rosendahl and Crocq 2020). Numerous studies investigating underlying mechanisms have revealed changes at the genetic, neuronal, and circuit levels (Sahin and Sur 2015). However, precise, causal relationships between observed pathologies and disruptions to circuit function remain poorly understood (Mizusaki and O'Donnell 2021; Hitchcock et al. 2022). By more accurately accounting for development in RNN paradigms, we gain the ability to make inferences about diseases in which specific mechanisms are known (Fig. 3).

Computational models of circuit dysfunction have spanned a range of levels, including networks of Hodgkin–Huxley neurons (O'Donnell et al. 2017; Onasch and Gjorgjieva 2020), leaky integrate-and-fire neurons (Cano-Colino and Compte 2012; Cavanagh et al. 2020; Lam et al. 2022), and population firing rate models (Murray et al. 2017). These generally follow a three-step process (Fig. 3A). First, the circuit model's parameters are tuned to fit a healthy state. Second, specific parameters are altered to match abnormalities found in a neurodevelopmental disorder. Third, the resulting computational deficits are compared to the symptoms of the disorder. In principle, this approach could be applied to any neurodevelopmental disorder (or psychiatric disorders more generally) but studies so far have focused primarily on ASD (Sahin and Sur 2015) and schizophrenia (Owen et al. 2011). Parameter alterations used to induce a disease state can be grouped into three broad categories (Lanillos et al. 2020): disconnection, where the network has an atypical decrease or increase in connections (Stevens 1992; Friston and Frith 1995), E/I imbalance, where the relative amounts of excitation and inhibition are abnormal (Rubenstein and Merzenich 2003), and hypopriors, in which circuits exhibit an abnormal reliance on sensory information relative to top-down predictions based on prior experiences (Pellicano and Burr 2012).

The advantages of RNNs as a model of neural circuit function described earlier also make RNNs suitable to investigate circuit dysfunction. The heterogeneous symptoms of neurodevelopmental disorders require a flexible modeling framework capable of diverse computations, and the ability of RNNs to learn many computations using the same training process, often within the same RNN (Yang et al. 2019), is well aligned to this goal. Similarly, the loss function of RNNs allows the exact quantification of computational errors across tasks. As a result, RNNs are equipped with a natural method to measure computational deficits induced by a parameter perturbation. Additionally, RNNs are universal approximators (Schäfer and Zimmermann 2006), often producing rich neural dynamics that can be analyzed to provide mechanistic explanations of computational deficits.

Schizophrenia

The earliest use of RNNs to investigate neurodevelopmental disorders applied Hopfield networks to explain how symptoms of schizophrenia might arise during memory formation (Hoffman 1987). The memory capacity of Hopfield networks is ~15% of the number of neurons (Hopfield 1982). Rather than applying a parameter alteration corresponding to a proposed circuit mechanism of schizophrenia, Hoffman (1987) induced a disease state in the network by overloading the number of stored memories. Surprisingly, the resulting errors were qualitatively different from those observed below the overloading threshold and were interpretable in the context of schizophrenia. In the subthreshold regime, memories were either recalled correctly or resulted in a "generalization" error in which the final state stabilized near a cluster of similar memories. In contrast, the overloaded

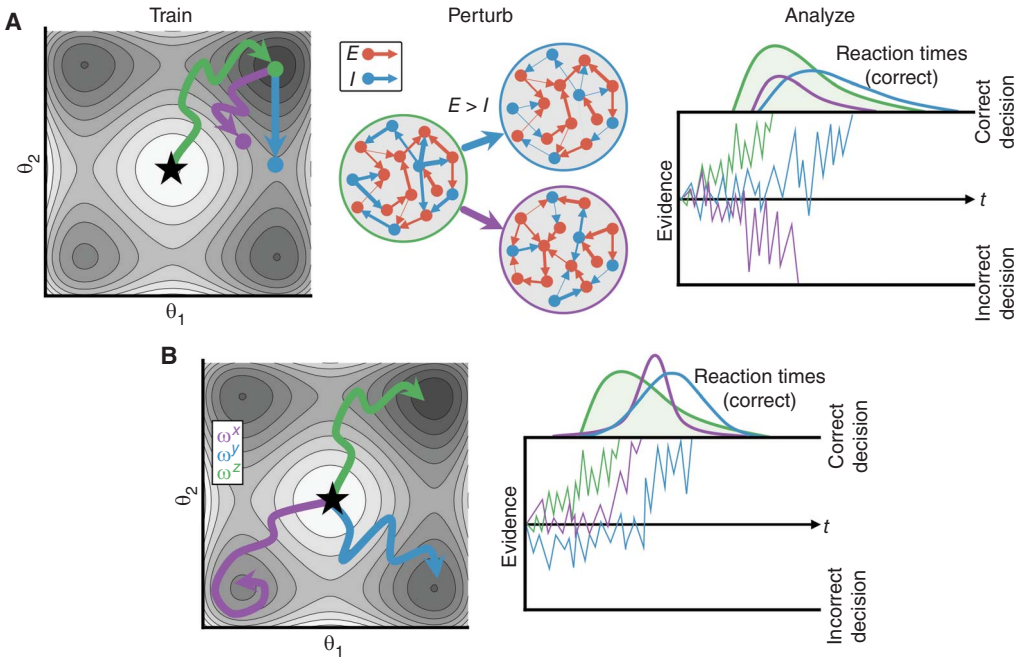

Figure 3. Recurrent neural network (RNN) modeling of neurodevelopmental disorders. (*A*) The train, perturb, analyze approach to RNN models of neurodevelopmental disorders. First, the RNN's parameters (θ_1 and θ_2) are trained (green line) such that a loss function is minimized and the RNN is in a "healthy" state (green dot). Then the parameters are perturbed (purple and blue lines) away from the optimal configuration to "disease" states. These perturbations may be systematic and easily interpretable, such as E/I imbalances in models of autism (e.g., weaking inhibitory connections, blue; Echeveste et al. 2022) or more abstract such as memory overloading in models of schizophrenia (purple; Hoffman 1987). The resulting computational changes can be analyzed to link circuit structure to computational deficits. For example, in a two alternative forced choice task, perturbations will result in lower performance (*lower* areas under the purple and blue distributions) but may also induce specific changes to how computations are performed (e.g., increasing the mean reaction time, i.e., translation of reaction time distributions to the *right*). (*B*) The metalearning perturbation approach to RNN models of neurodevelopmental disorders. Changes (ω^x, purple, and ω^y, blue) to the metalearned healthy state (ω^z, green) may result in distinct solutions to the learning problem (Philippsen and Nagai 2018). These solutions may have similar task performance (areas under the green, blue, and purple distributions are comparable) but perform computations in radically different ways (blue and gold distributions have different means and higher-order moments).

state produced two types of pathological errors: misperceptions, in which the network exactly recalled an incorrect memory, and more severe dysfunction—similar to delusions or hallucinations—in which many inputs recalled the same, nonmemory state. The insights into schizophrenia provided by this study were limited by the direction of its inference: observing a dysfunctional state in an RNN and then mapping it onto a psychiatric disorder without an underlying biological mechanism. Despite this limitation, its findings demonstrated the applicability of RNNs to psychiatric disorders (Lanillos et al. 2020).

Subsequent studies examined the ability of excessive synaptic pruning to account for symptoms of schizophrenia. Hoffman and Dobscha (1989) trained Hopfield networks to recall a set of memories and then removed connections from weakest to strongest while tracking recall performance using the same error classes described above. Surprisingly, they found that synaptic pruning induced "delusions" and "hallucinations" as described above, although not until ~80% of syn-

apses were removed. This demonstrated the feasibility of a biologically motivated alteration to an RNN as a model for a psychiatric disorder. Furthermore, it is aligned with significant, experimental observed pruning during healthy development (Huttenlocher 1979). However, the model predicts that delusions and hallucinations in schizophrenia are accompanied by memory loss, which is at odds with clinically observed symptoms.

A second series of RNN schizophrenia models addressed this mechanism/symptom inconsistency by analyzing an alternative theory due to Stevens (1992) in which recurrent connectivity strength in the hippocampus overcompensates for degraded external inputs (Horn and Ruppin 1995; Ruppin et al. 1996). The authors used a more biologically motivated Hopfield model variant (Tsodyks 1988; Tsodyks and Feigel'man 1988) with directed connections and a more realistic level of neural activity. Mean field theoretic analysis and simulations revealed the functional consequences of both proposed biological mechanisms. Weakening the model's external input impaired memory function and strengthening recurrent connections restored memory function at the cost of inducing hallucination-like memory retrieval errors. Combined, these models of schizophrenia highlight the ability of RNNs to parse redundant circuit mechanisms in psychiatric disorders (Mizusaki and O'Donnell 2021).

Autism Spectrum Disorder

The proportion of excitation and inhibition neurons receive is critical to sensory processing (Anderson et al. 2000; Wehr and Zador 2003) and is a fundamental description of the activity in neuronal networks (van Vreeswijk and Sompolinsky 1996; Ahmadian and Miller 2021). Similarly, an imbalance in E/I signaling is a prominent hypothesis for the mechanism underlying ASDs (Rubenstein and Merzenich 2003). However, the extent to which E/I imbalances can explain the symptoms of autism is not well understood (O'Donnell et al. 2017; Antoine et al. 2019).

A recent ASD study (Echeveste et al. 2022) used RNNs to provide a mechanistic link between the E/I imbalance theory of circuit dysfunction and the hypopriors theory of ASD, a variation of the Bayesian framework for perception (Pellicano and Burr 2012). The hypopriors theory of ASD proposes that higher uncertainty in prior distributions results in an overreliance on sensory information. To explore the neural basis of hypopriors, Echeveste et al. (2022) built on previous work that trained an excitatory–inhibitory RNN circuit model to implement Bayesian perceptual inference such that optimal estimates of the environmental state were encoded in the activity of the excitatory neurons (Echeveste et al. 2020). E/I imbalances were then induced in the trained model by weakening inhibitory connections, and the effect on perception was measured from the resulting posterior distributions. Critically, the E/I imbalances induced an overreliance on sensory information, as predicted by the hypopriors theory of ASD. This result represents one of the clearest links between circuit dysfunction and computational deficits and highlights the potential of RNNs to reveal and clarify the neural basis of developmental disorders.

Using the language of metalearning, the train-alter-compare approach induces a disease state by shifting the RNN out of an optimal, learned state (i.e., changing θ in Equation 3). Alternatively, differences in cognitive processes observed in neurodevelopmental disorders could be a distinct, locally optimal state resulting from changes throughout development. As a result, neurodevelopmental disorders could be more appropriately modeled by shifting the RNN out of a metalearned state (i.e., changing ω^* in Equation 4; Fig. 3B). Philippsen and Nagai (2018) took the latter approach, training an RNN to perform time series forecasting by estimating the mean and variance of an input signal. The authors separately altered two learning hyperparameters, external contribution and aberrant precision, and measured the resulting differences in task performance and internal task variable representation after a fixed number of training steps. The external contribution determined the amount of sensory information used when building its internal representation of the time series, and the aberrant precision determined

the RNNs' estimate of the signal variance. While most parameters allowed the RNN to learn the task, the degree of structure in the internal task variable representations varied independently of task performance. This highlights that differences in cognitive processes may not align with decreases in task performance. As a result, investigation of specific behaviors alone may not be sufficient for understanding the mechanisms and cognitive processes underlying autism and other neurodevelopmental disorders. Similarly, limits inherent to train–alter–compare studies may be addressed by metalearning models of neurodevelopmental disorders.

CONCLUSIONS

Theories of neural development require a framework capable of extensive changes to circuit structure and function spread across multiple timescales. RNNs can fit all these criteria and have already yielded crucial insight into the codevelopment of neural circuit structure, dynamics, and computation. Incorporation of metalearning, often paired with reinforcement learning, as an explicit model of multiple developmental timescales, represents a promising method for gaining insight into both normal (Wang et al. 2018; Jiang and Litwin-Kumar 2021; Goudar et al. 2023) and abnormal development (Philippsen and Nagai 2018). In addition, the implementation of structurally dynamic learning rules has yielded insight into the computational significance of neurodevelopmental mechanisms.

How can RNN models of neural development work synergistically with RNN circuit models more broadly? Major challenges in understanding neural development and neurodevelopmental disorders parallel challenges in RNN analysis of mature circuits. Although RNNs are often used to hypothesize the relationship between circuit structure and function (Sussillo 2014), alternative, mechanistically distinct RNN hypotheses often yield nearly identical solutions to the same problem (Vyas et al. 2020; O'Shea et al. 2022). Similarly, redundancy in neural circuit function (Fig. 3B) represents a major challenge in understanding the neural basis of neurodevelopmental disorders (Mizusaki and O'Donnell 2021). Because these are two sides of the same coin, techniques that identify the mechanisms of neurodevelopmental disorders could be used to guide analysis that differentiates the RNN hypothesis. Recent advances in multitask learning in RNNs (Yang et al. 2019) offer a potential solution to this problem. By training the same RNN to perform many computations, the functional consequences of circuit differences—regardless of origin—can be compared across tasks.

In summary, the ability of RNNs to flexibly perform complex computations using the basic components of neural circuits makes them an appealing model for understanding not just the brain in general, but also the emergence of network computations on developmental timescales.

REFERENCES

Ahmadian Y, Miller KD. 2021. What is the dynamical regime of cerebral cortex? *Neuron* **109:** 3373–3391. doi:10.1016/j.neuron.2021.07.031

Anderson PW. 1972. More is different. *Science* **177:** 393–396. doi:10.1126/science.177.4047.393

Anderson JS, Carandini M, Ferster D. 2000. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J Neurophysiol* **84:** 909–926. doi:10.1152/jn.2000.84.2.909

Angeline PJ, Saunders GM, Pollack JB. 1994. An evolutionary algorithm that constructs recurrent neural networks. *IEEE Trans Neural Netw* **5:** 54–65. doi:10.1109/72.265960

Antoine MW, Langberg T, Schnepel P, Feldman DE. 2019. Increased excitation-inhibition ratio stabilizes synapse and circuit excitability in four autism mouse models. *Neuron* **101:** 648–661.e4. doi:10.1016/j.neuron.2018.12.026

Barabási DL, Beynon T, Katona Á, Perez-Nieves N. 2023. Complex computation from developmental priors. *Nat Commun* **14:** 2226. doi:10.1038/s41467-023-37980-1

Bicknell BA, Dayan P, Goodhill GJ. 2015. The limits of chemosensation vary across dimensions. *Nat Commun* **6:** 7468. doi:10.1038/ncomms8468

Bourgeois J, Rakic P. 1993. Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. *J Neurosci* **13:** 2801–2820. doi:10.1523/JNEUROSCI.13-07-02801.1993

Cano-Colino M, Compte A. 2012. A computational model for spatial working memory deficits in schizophrenia. *Pharmacopsychiatry* **45**(Suppl 1)**:** S49–S56. doi:10.1055/s-0032-1306314

Cavanagh SE, Lam NH, Murray JD, Hunt LT, Kennerley SW. 2020. A circuit mechanism for decision-making biases and

NMDA receptor hypofunction. *eLife* **9:** e53664. doi:10.7554/eLife.53664

Dayan P, Abbott LF. 2005. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT Press, Cambridge, MA.

Doya K. 2002. Metalearning and neuromodulation. *Neural Netw* **15:** 495–506. doi:10.1016/S0893-6080(02)00044-8

Driscoll LN, Pettit NL, Minderer M, Chettih SN, Harvey CD. 2017. Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell* **170:** 986–999.e16. doi:10.1016/j.cell.2017.07.021

Driscoll L, Shenoy K, Sussillo D. 2022. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. bioRxiv doi:10.1101/20220815503870

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. doi:10.1038/s41593-022-01088-4

Echeveste R, Aitchison L, Hennequin G, Lengyel M. 2020. Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nat Neurosci* **23:** 1138–1149. doi:10.1038/s41593-020-0671-1

Echeveste R, Ferrante E, Milone DH, Samengo I. 2022. Bridging physiological and perceptual views of autism by means of sampling-based Bayesian inference. *Netw Neurosci* **6:** 196–212. doi:10.1162/netn_a_00219

Fahlman S. 1990. The recurrent cascade-correlation architecture. *Adv Neural Inf Process Syst* **3:** 190–196.

Faust TE, Gunner G, Schafer DP. 2021. Mechanisms governing activity-dependent synaptic pruning in the developing mammalian CNS. *Nat Rev Neurosci* **22:** 657–673. doi:10.1038/s41583-021-00507-y

Finn C, Abbeel P, Levine S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning, Vol. 70 of Proceedings of Machine Learning Research*, pp. 1126–1135. PMLR, Sydney, Australia.

Friston KJ, Frith CD. 1995. Schizophrenia: a disconnection syndrome. *Clin Neurosci* **3:** 89–97.

Ghazizadeh E, Ching S. 2021. Slow manifolds within network dynamics encode working memory efficiently and robustly. *PLoS Comput Biol* **17:** e1009366. doi:10.1371/journal.pcbi.1009366

Goodhill GJ. 1993. Topography and ocular dominance: a model exploring positive correlations. *Biol Cybern* **69:** 109–118. doi:10.1007/BF00226194

Goodhill GJ. 2018. Theoretical models of neural development. *iScience* **8:** 183–199. doi:10.1016/j.isci.2018.09.017

Goudar V, Peysakhovich B, Freedman DJ, Buffalo EA, Wang X-J. 2023. Schema formation in a neural population subspace underlies learning-to-learn in flexible sensorimotor problem-solving. *Nat Neurosci* **26:** 879–890. doi:10.1038/s41593-023-01293-9

Gregor T, Tank DW, Wieschaus EF, Bialek W. 2007. Probing the limits to positional information. *Cell* **130:** 153–164. doi:10.1016/j.cell.2007.05.025

Guan D, Armstrong WE, Foehring RC. 2014. Electrophysiological properties of genetically identified subtypes of layer 5 neocortical pyramidal neurons: Ca^{2+} dependence and differential modulation by norepinephrine. *J Neurophysiol* **113:** 2014–2032. doi:10.1152/jn.00524.2014

Harlow HF. 1949. The formation of learning sets. *Psychol Rev* **56:** 51–65. doi:10.1037/h0062474

Hebb DO. 1949. *The organization of behavior: a neuropsychological theory*. Psychology, New York.

Hennequin G, Vogels TP, Gerstner W. 2014. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82:** 1394–1406. doi:10.1016/j.neuron.2014.04.045

Hennig JA, OBY ER, Losey DM, Batista AP, Yu BM, Chase SM. 2021. How learning unfold in the brain: toward and optimization view. *Neuron* **109:** 3720–3735. doi:10.1016/j.neuron.2021.09.005

Hitchcock PF, Fried EI, Frank MJ. 2022. Computational psychiatry needs time and context. *Annu Rev Psychol* **73:** 243–270. doi:10.1146/annurev-psych-021621-124910

Hodgkin AL, Huxley AF. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* **117:** 500–544. doi:10.1113/jphysiol.1952.sp004764

Hoffman RE. 1987. Computer simulations of neural information processing and the schizophrenia-mania dichotomy. *Arch Gen Psychiatry* **44:** 178. doi:10.1001/archpsyc.1987.01800140090014

Hoffman RE, Dobscha SK. 1989. Cortical pruning and the development of schizophrenia: a computer model. *Schizophr Bull* **15:** 477–490. doi:10.1093/schbul/15.3.477

Holland PW, Laskey KB, Leinhardt S. 1983. Stochastic blockmodels: first steps. *Soc Netw* **5:** 109–137. doi:10.1016/0378-8733(83)90021-7

Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* **79:** 2554–2558. doi:10.1073/pnas.79.8.2554

Horn D, Ruppin E. 1995. Compensatory mechanisms in an attractor neural network model of schizophrenia. *Neural Comput* **7:** 182–205. doi:10.1162/neco.1995.7.1.182

Hospedales T, Antoniou A, Micaelli P, Storkey A. 2022. Meta-learning in neural networks: a survey. *IEEE Trans Pattern Anal Machine Intell* **44:** 5149–5169. doi:10.1109/TPAMI.2021.3079209

Huttenlocher PR. 1979. Synaptic density in human frontal cortex—developmental changes and effects of aging. *Brain Res* **163:** 195–205. doi:10.1016/0006-8993(79)90349-4

Izhikevich EM. 2006. *Dynamical systems in neuroscience*. MIT Press, Cambridge, MA.

Jiang L, Litwin-Kumar A. 2021. Models of heterogeneous dopamine signaling in an insect learning and memory center. *PLoS Comput Biol* **17:** e1009205. doi:10.1371/journal.pcbi.1009205

Johnson S, Marro J, Torres JJ. 2010. Evolving networks and the development of neural systems. *J Stat Mech* **2010:** P03003. doi:10.1088/1742-5468/2010/03/P03003

Kleinman M, Chandrasekaran C, Kao JC. 2020. Recurrent neural network models of multi-area computation underlying decision-making. bioRxiv doi:10.1101/798553

Koulakov A, Shuvaev S, Lachi D, Zador A. 2022. Encoding innate ability through a genomic bottleneck. bioRxiv doi:10.1101/20210316435261

Kudithipudi D, Aguilar-Simon M, Babb J, Bazhenov M, Blackiston D, Bongard J, Brna AP, Chakravarthi Raja S, Cheney N, Clune J, et al. 2022. Biological underpinnings

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

for lifelong learning machines. *Nat Mach Intell* **4:** 196–210. doi:10.1038/s42256-022-00452-0

Lam NH, Borduqui T, Hallak J, Roque A, Anticevic A, Krystal JH, Wang XJ, Murray JD. 2022. Effects of altered excitation-inhibition balance on decision making in a cortical circuit model. *J Neurosci* **42:** 1035–1053. doi:10.1523/JNEUROSCI.1371-20.2021

Lanillos P, Oliva D, Philippsen A, Yamashita Y, Nagai Y, Cheng G. 2020. A review on neural network models of schizophrenia and autism spectrum disorder. *Neural Netw* **122:** 338–363. doi:10.1016/j.neunet.2019.10.014

Li Z, Zhou F, Chen F, Li H. 2017. Meta-SGD: learning to learn quickly for few-shot learning. arXiv doi:10.48550/ARXIV170709835

Linsker R. 1986. From basic network principles to neural architecture: emergence of spatial-opponent cells. *Proc Natl Acad Sci* **83:** 7508–7512. doi:10.1073/pnas.83.19.7508

Liu YH, Zhu J, Constantinidis C, Zhou X. 2021. Emergence of prefrontal neuron maturation properties by training recurrent neural networks in cognitive tasks. *iScience* **24:** 103178. doi:10.1016/j.isci.2021.103178

Mante V, Sussillo D, Shenoy KV, Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503:** 78–84. doi:10.1038/nature12742

Marder E. 2011. Variability, compensation, and modulation in neurons and circuits. *Proc Natl Acad Sci* **108:** 15542–15548. doi:10.1073/pnas.1010674108

Markus EJ, Petit TL. 1987. Neocortical synaptogenesis, aging, and behavior: lifespan development in the motor-sensory system of the rat. *Exp Neurol* **96:** 262–278. doi:10.1016/0014-4886(87)90045-8

Marsland S, Shapiro J, Nehmzow U. 2002. A self-organising network that grows when required. *Neural Netw* **15:** 1041–1058. doi:10.1016/S0893-6080(02)00078-3

Masse NY, Yang GR, Song HF, Wang XJ, Freedman DJ. 2019. Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nat Neurosci* **22:** 1159–1167. doi:10.1038/s41593-019-0414-3

McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* **102:** 419–457. doi:10.1037/0033-295X.102.3.419

McCormick DA, Connors BW, Lighthall JW, Prince DA. 1985. Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J Neurophysiol* **54:** 782–806. doi:10.1152/jn.1985.54.4.782

McCulloch WS, Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* **5:** 115–133. doi:10.1007/BF02478259

Mikhaeil J, Monfared Z, Durstewitz D. 2022. On the difficulty of learning chaotic dynamics with RNNs. In *Advances in neural information processing systems* (ed. Koyejo S, et al.), Vol. 35, pp. 11297–11312. Curran Associates, Red Hook, NY.

Millán AP, Torres JJ, Johnson S, Marro J. 2018. Concurrence of form and function in developing networks and its role in synaptic pruning. *Nat Commun* **9:** 2236. doi:10.1038/s41467-018-04537-6

Millán AP, Torres JJ, Marro J. 2019. How memory conforms to brain development. *Front Comput Neurosci* **13:** 22. doi:10.3389/fncom.2019.00022

Millán AP, Torres JJ, Johnson S, Marro J. 2021. Growth strategy determines the memory and structural properties of brain networks. *Neural Netw* **142:** 44–56. doi:10.1016/j.neunet.2021.04.027

Miller KD, Keller JB, Stryker MP. 1989. Ocular dominance column development: analysis and simulation. *Science* **245:** 605–615. doi:10.1126/science.2762813

Mizusaki BEP, O'Donnell C. 2021. Neural circuit function redundancy in brain disorders. *Curr Opin Neurobiol* **70:** 74–80. doi:10.1016/j.conb.2021.07.008

Molano-Mazón M, Shao Y, Duque D, Yang GR, Ostojic S, de la Rocha J. 2023. Recurrent networks endowed with structural priors explain suboptimal animal behavior. *Curr Biol* **33:** 622–638.e7. doi:10.1016/j.cub.2022.12.044

Morris-Rosendahl DJ, Crocq MA. 2020. Neurodevelopmental disorders—the history and future of a diagnostic concept. *Dialogues Clin Neurosci* **22:** 65–72. doi:10.31887/DCNS.2020.22.1/macrocq

Murray JD, Jaramillo J, Wang XJ. 2017. Working memory and decision-making in a frontoparietal circuit model. *J Neurosci* **37:** 12167–12186. doi:10.1523/JNEUROSCI.0343-17.2017

Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK. 2019. Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci* **22:** 1677–1686. doi:10.1038/s41593-019-0502-4

Navlakha S, Bar-Joseph Z, Barth AL. 2018. Network design and the brain. *Trends Cogn Sci* **22:** 64–78. doi:10.1016/j.tics.2017.09.012

O'Donnell C, Gonç Alves T, Portera-Cailliau C, Sejnowski TJ. 2017. Beyond excitation/inhibition imbalance in multidimensional models of neural circuit changes in brain disorders. *eLife* **6:** e26724. doi:10.7554/eLife.26724

Onasch S, Gjorgjieva J. 2020. Circuit stability to perturbations reveals hidden variability in the balance of intrinsic and synaptic conductances. *J Neurosci* **40:** 3186–3202. doi:10.1523/JNEUROSCI.0985-19.2020

Orhan AE, Ma WJ. 2019. A diverse range of factors affect the nature of neural representations underlying short-term memory. *Nat Neurosci* **22:** 275–283. doi:10.1038/s41593-018-0314-y

O'Shea DJ, Duncker L, Goo W, Sun X, Vyas S, Trautmann EM, Diester I, Ramakrishnan C, Deisseroth K, Sahani M. 2022. Direct neural perturbations reveal a dynamical mechanism for robust computation. bioRxiv doi:10.1101/2022.12.16.520768

Oswald AMM, Reyes AD. 2008. Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex. *J Neurophysiol* **99:** 2998–3008. doi:10.1152/jn.01160.2007

Owen MJ, O'Donovan MC, Thapar A, Craddock N. 2011. Neurodevelopmental hypothesis of schizophrenia. *Br J Psychiatry* **198:** 173–175. doi:10.1192/bjp.bp.110.084384

Parisi GI, Tani J, Weber C, Wermter S. 2017. Lifelong learning of human actions with deep neural network self-organiza-

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

tion. *Neural Netw* **96:** 137–149. doi:10.1016/j.neunet.2017.09.001

Parisi GI, Tani J, Weber C, Wermter S. 2018. Lifelong learning of spatiotemporal representations with dual-memory recurrent self-organization. *Front Neurorobot* **12:** 78. doi:10.3389/fnbot.2018.00078

Pellicano E, Burr D. 2012. When the world becomes "too real": a Bayesian explanation of autistic perception. *Trends Cogn Sci* **16:** 504–510. doi:10.1016/j.tics.2012.08.009

Perich MG, Rajan K. 2020. Rethinking brain-wide interactions through multi-region "network of networks" models. *Curr Opin Neurobiol* **65:** 146–151. doi:10.1016/j.conb.2020.11.003

Philippsen A, Nagai Y. 2018. Understanding the cognitive mechanisms underlying autistic behavior: a recurrent neural network study. In *2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pp. 84–90. IEEE, Piscataway, NJ.

Piaget J. 1997. *Language and thought of the child*. Routledge, London. doi:10.4324/9780203992739

Pinto L, Rajan K, DePasquale B, Thiberge SY, Tank DW, Brody CD. 2019. Task-dependent changes in the large-scale dynamics and necessity of cortical regions. *Neuron* **104:** 810–824.e9. doi:10.1016/j.neuron.2019.08.025

Ponce-Alvarez A, Mochol G, Hermoso-Mendizabal A, de la Rocha J, Deco G. 2020. Cortical state transitions and stimulus response evolve along stiff and sloppy parameter dimensions, respectively. *eLife* **9:** e53268. doi:10.7554/eLife.53268

Ratcliff R. 1990. Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychol Rev* **97:** 285–308. doi:10.1037/0033-295X.97.2.285

Rinzel J, Ermentrout GB. 1989. Analysis of neural excitability and oscillations. In *Methods in neuronal modeling: from synapses to networks*, pp. 135–169. MIT Press, Cambridge, MA.

Rubenstein JLR, Merzenich MM. 2003. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav* **2:** 255–267. doi:10.1034/j.1601-183X.2003.00037.x

Rumelhart DE, Hinton GE, Williams RJ. 1986. Learning representations by back-propagating errors. *Nature* **323:** 533–536. doi:10.1038/323533a0

Ruppin E, Reggia JA, Horn D. 1996. Pathogenesis of schizophrenic delusions and hallucinations: a neural model. *Schizophr Bull* **22:** 105–121. doi:10.1093/schbul/22.1.105

Sahin M, Sur M. 2015. Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. *Science* **350:** aab3897. doi:10.1126/science.aab3897

Saxena S, Russo AA, Cunningham J, Churchland MM. 2022. Motor cortex activity across movement speeds is predicted by network-level strategies for generating muscle activity. *eLife* **11:** e67620. doi:10.7554/eLife.67620

Schäfer AM, Zimmermann HG. 2006. Recurrent neural networks are universal approximators. In *Artificial Neural Networks—ICANN 2006*. Lecture Notes in Computer Science, Vol. 4131. Springer, Berlin.

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science* **275:** 1593–1599. doi:10.1126/science.275.5306.1593

Shenoy KV, Sahani M, Churchland MM. 2013. Cortical control of arm movements: a dynamical systems perspective. *Annu Rev Neurosci* **36:** 337–359. doi:10.1146/annurev-neuro-062111-150509

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. doi:10.1371/journal.pcbi.1004792

Song HF, Yang GR, Wang XJ. 2017. Reward-based training of recurrent neural networks for cognitive and value-based tasks. *eLife* **6:** e21492. doi:10.7554/eLife.21492

Steinmetz NA, Zatka-Haas P, Carandini M, Harris KD. 2019. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576:** 266–273. doi:10.1038/s41586-019-1787-x

Stevens JR. 1992. Abnormal reinnervation as a basis for schizophrenia: a hypothesis. *Arch Gen Psychiatry* **49:** 238. doi:10.1001/archpsyc.1992.01820030070009

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. 2019. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364:** 225. doi:10.1126/science.aav7893

Sucholutsky I, Muttenthaler L, Weller A, Peng A, Bobu A, Kim B, Love BC, Grant E, Groen I, Achterberg J, et al. 2023. Getting aligned on representational alignment. arXiv doi:10.48550/arXiv.2310.13018

Sussillo D. 2014. Neural circuits as computational dynamical systems. *Curr Opin Neurobiol* **25:** 156–163. doi:10.1016/j.conb.2014.01.008

Sussillo D, Abbott LF. 2009. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63:** 544–557. doi:10.1016/j.neuron.2009.07.018

Sussillo D, Nuyujukian P, Fan JM, Kao JC, Stavisky SD, Ryu S, Shenoy K. 2012. A recurrent neural network for closed-loop intracortical brain–machine interface decoders. *J Neural Eng* **9:** 026027. doi:10.1088/1741-2560/9/2/026027

Sussillo D, Churchland MM, Kaufman MT, Shenoy KV. 2015. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat Neurosci* **18:** 1025–1033. doi:10.1038/nn.4042

Tkačik G, Dubuis JO, Petkova MD, Gregor T. 2015. Positional information, positional error, and readout precision in morphogenesis: a mathematical framework. *Genetics* **199:** 39–59. doi:10.1534/genetics.114.171850

Tsodyks MV. 1988. Associative memory in asymmetric diluted network with low level of activity. *Europhys Lett* **7:** 203–208. doi:10.1209/0295-5075/7/3/003

Tsodyks MV, Feigel'man MV. 1988. The enhanced storage capacity in neural networks with low activity level. *Europhys Lett* **6:** 101–105. doi:10.1209/0295-5075/6/2/002

Urai AE, Doiron B, Leifer AM, Churchland AK. 2022. Large-scale neural recordings call for new insights to link brain and behavior. *Nat Neurosci* **25:** 11–19. doi:10.1038/s41593-021-00980-9

van Ooyen A. 2011. Using theoretical models to analyse neural development. *Nat Rev Neurosci* **12:** 311–326. doi:10.1038/nrn3031

van Vreeswijk C, Sompolinsky H. 1996. Chaos in neuronal networks with balanced excitatory and inhibitory activity.

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Science **274:** 1724–1726. doi:10.1126/science.274.5293.1724

von der Malsburg C. 1973. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* **14:** 85–100. doi:10.1007/BF00288907

Vyas S, Golub MD, Sussillo D, Shenoy KV. 2020. Computation through neural population dynamics. *Annu Rev Neurosci* **43:** 249–275. doi:10.1146/annurev-neuro-092619-094115

Wang JX. 2021. Meta-learning in natural and artificial intelligence. *Curr Opin Behav Sci* **38:** 90–95. doi:10.1016/j.cobeha.2021.01.002

Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M. 2018. Prefrontal cortex as a meta-reinforcement learning system. *Nat Neurosci* **21:** 860–868. doi:10.1038/s41593-018-0147-8

Wang PY, Sun Y, Axel R, Abbott LF, Yang GR. 2021. Evolving the olfactory system with machine learning. *Neuron* **109:** 3879–3892.e5. doi:10.1016/j.neuron.2021.09.010

Wehr M, Zador AM. 2003. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426:** 442–446. doi:10.1038/nature02116

Werbos PJ. 1990. Backpropagation through time: what it does and how to do it. *Proc IEEE* **78:** 1550–1560. doi:10.1109/5.58337

White E, Weinfeld L, Lev D. 1997. A survey of morphogenesis during the early postnatal period in PMBSF barrels of mouse SmI cortex with emphasis on barrel D4. *Somatosens Mot Res* **14:** 34–55. doi:10.1080/08990229771204

Yamins DLK, 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* **111:** 8619–8624. doi:10.1073/pnas.1403112111

Yang GR, Wang X-J. 2020. Artificial neural networks for neuroscientists: a primer. *Neuron* **107:** 1048–1070. doi:10.1016/j.neuron.2020.09.005

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. doi:10.1038/s41593-018-0310-2

Zador AM. 2019. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat Commun* **10:** 3770. doi:10.1038/s41467-019-11786-6

Zhang Y, Yang Q. 2018. An overview of multi-task learning. *Natl Sci Rev* **5:** 30–43. doi:10.1093/nsr/nwx105

Zhou X, Zhu D, Qi XL, Lees CJ, Bennett AJ, Salinas E, Stanford TR, Constantinidis C. 2013. Working memory performance and neural activity in prefrontal cortex of peripubertal monkeys. *J Neurophysiol* **110:** 2648–2660. doi:10.1152/jn.00370.2013

Zhou X, Qi XL, Constantinidis C. 2016a. Distinct roles of the prefrontal and posterior parietal cortices in response inhibition. *Cell Rep* **14:** 2765–2773. doi:10.1016/j.celrep.2016.02.072

Zhou X, Zhu D, King SG, Lees CJ, Bennett AJ, Salinas E, Stanford TR, Constantinidis C. 2016b. Behavioral response inhibition and maturation of goal representation in prefrontal cortex after puberty. *Proc Natl Acad Sci* **113:** 3353–3358. doi:10.1073/pnas.1518147113

Zipser D, Andersen RA. 1988. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* **331:** 679–684. doi:10.1038/331679a0

Cold Spring Harbor Perspectives in Biology

Modeling Normal and Abnormal Circuit Development with Recurrent Neural Networks

Daniel Zavitz, ShiNung Ching and Geoffrey Goodhill

Cold Spring Harb Perspect Biol published online July 15, 2024