nature neuroscience

Article

https://doi.org/10.1038/s41593-025-02080-4

Prediction of neural activity in connectome-constrained recurrent networks

Received: 13 May 2024

Accepted: 5 September 2025

Published online: 27 October 2025

Check for updates

Manuel Beiran [1,2] & Ashok Litwin-Kumar [1,2,3]

Recent technological advances have enabled measurement of the synaptic wiring diagram, or 'connectome', of large neural circuits or entire brains. However, the extent to which such data constrain models of neural dynamics and function is debated. In this study, we developed a theory of connectome-constrained neural networks in which a 'student' network is trained to reproduce the activity of a ground truth 'teacher', representing a neural system for which a connectome is available. Unlike standard paradigms with unconstrained connectivity, the two networks have the same synaptic weights but different biophysical parameters, reflecting uncertainty in neuronal and synaptic properties. We found that a connectome often does not substantially constrain the dynamics of recurrent networks, illustrating the difficulty of inferring function from connectivity alone. However, recordings from a small subset of neurons can remove this degeneracy, producing dynamics in the student that agree with the teacher. Our theory demonstrates that the solution spaces of connectome-constrained and unconstrained models are qualitatively different and determines when activity in such networks can be well predicted. It can also prioritize which neurons to record to most effectively inform such predictions.

Establishing links between the connectivity of large neural networks and their emergent dynamics is a major goal of theoretical neuroscience. Many studies have attempted to develop methods to infer synaptic connectivity from functional correlations derived from recorded neural activity. However, this 'inverse problem' has proven to be challenging and often ill-posed[1–5], due to the degeneracy of the space of network connectivities that produce similar dynamics. Such inference is particularly difficult when neural dynamics are low dimensional or otherwise structured[1].

The recent availability of comprehensive synaptic connectome datasets has led to approaches that focus instead on the 'forward problem' of predicting neural dynamics from synaptic connectivity. The scale of such datasets has increased rapidly, from the 302 neurons of the nematode *Caenorhabditis elegans* identified decades ago[6] to recently acquired volumes containing entire nervous systems of *Drosophila* larvae[7] and adults[8–10] and larval zebrafish[11]. Several studies have compared connectomes with functional connectivity based on activity correlations between neurons in the resting state or in response to optogenetic perturbations[12]. This has highlighted notable differences for certain systems[13]. A complementary line of research has used connectome information to initialize or build explicit priors on the distribution of the parameters of neural network models[14,15]. In some cases, these models are then optimized to perform computations, and it has been found empirically that such biological constraints sometimes yield

[1]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA. [2]Kavli Institute for Brain Science, Columbia University Irving Medical Center, New York, NY, USA. [3]Department of Neuroscience, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY, USA. e-mail: mb4878@columbia.edu; a.litwin-kumar@columbia.edu

models with improved abilities to predict neural data[16–19]. However, the ill-posedness of the inverse problem and lack of one-to-one correspondence between structure and function call into question the reliability of such predictions.

A major challenge for connectome-constrained models is uncertainty in biophysical parameters that affect neural dynamics. Connectomes generated from electron microscopy imaging provide information on structural connections, neurotransmitter identities of chemical synapses[10,20] and connection strengths estimated by synapse count[21] or volume[22]. However, other biological processes are undetermined, such as the neuromodulatory environment, existence of electrical synapses and functional properties of individual neurons and synapses[23]. Changes in such parameters were previously shown to produce dramatic alterations in network activity[24–27].

In the present study, we develop a theory of the solution spaces of networks with specified synaptic weights but unmeasured and heterogeneous single-neuron biophysical parameters[28,29]. We use a 'teacher–student' paradigm in which the activity of a 'student' network is trained to reproduce the activity of a 'teacher' network. The teacher is a synthetic model that represents ground truth, analogous to biological circuits for which a connectome is available and from which we can record activity. Unlike previous theories in which the student and teacher neurons have the same input–output function and synaptic weights are trained[1,30], here the two networks have the same weights, but their biophysical properties differ a priori.

We found that training a connectome-constrained student network to generate the task-related readout of the teacher does not always produce consistent dynamics in the teacher and student. Multiple combinations of single-neuron parameters, each producing different activity patterns, can equivalently solve the same task. However, when connectivity constraints are combined with recordings of the activity of a subset of neurons, this degeneracy is broken. The minimum number of recordings depends on the dimensionality of the network dynamics, not the total number of neurons. This contrasts with student networks whose connectivity is unconstrained, which always display degenerate solutions. Interestingly, even when neural activity is well reconstructed, single-neuron parameters are often not recovered accurately, suggesting that some combinations of parameters are 'stiff', with strong effects on neural dynamics, whereas others are 'sloppy', with weak effects. Our qualitative predictions hold across a variety of simulated networks and networks constrained by true connectomes from invertebrates and vertebrates. Our theory can also rank neurons that should be recorded with higher priority to maximally reduce uncertainty in network activity, suggesting approaches that iteratively refine network models using neural recordings.

Results

Teacher–student recurrent networks

To explore how a connectome constrains the solutions of neural network models, we studied a teacher–student paradigm[31,32]: a recurrent neural network (RNN) that we call the teacher is constructed, and the parameters of a student RNN are adjusted to mimic this teacher. The teacher is used as a proxy for a neural system whose connectome has been mapped and whose output or neural activity can be recorded. To develop our theory, we will begin by examining synthetic teacher networks whose activity and function we specify. Later, we will consider teacher networks derived from empirical connectome data.

Both teacher and student are composed of N firing rate neurons, in which the activity of neuron i is described by a continuous variable $r_i(t)$ (see Methods for details). The activity is a nonlinear function, which we call the activation function, of the input current $x_i(t)$ received by the neuron and depends on a set of single-neuron parameters. For instance, if we describe this function using parameters g_i and b_i for neuron i's gain and bias, its activity is given by

$$r_i(t) = g_i\phi\left(x_i(t) + b_i\right), \tag{1}$$

where ϕ is a nonlinear function. The network dynamics follow:

$$\tau\frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij}r_j + I_i(t), \tag{2}$$

where J_{ij} is the synaptic weight from neuron j to neuron i, and $I_i(t)$ is the time-varying external input received by neuron i. For connectome-constrained networks, we begin by assuming that both the presence or absence of a connection between neurons as well as the strengths of these connections are known, and, thus, J_{ij} is the same for both teacher and student. Additionally, we assume that the external inputs and initial state $x(t=0)$ are the same for teacher and student (Discussion).

Note that the number of unconstrained parameters in the student network scales differently depending on whether single-neuron parameters or connectivity parameters are fixed. There are N^2 free synaptic weight parameters if the connectivity is unspecified, as in previous studies of teacher–student paradigms[31,32]. On the other hand, for connectome-constrained networks, the number of unconstrained parameters is proportional to N. For example, when we parameterize the activation functions of neurons with gains and biases, as in equation (1), there are $2N$ unknowns.

Student network constrained by task output

We first asked whether teacher and student networks that share the same synaptic weight matrix exhibit consistent solutions when the student is trained to reproduce a task performed by the teacher (Fig. 1). Because we are interested in whether connectivity constraints yield mechanistic models of the teacher, we measure the consistency of solutions using the similarity of the activity of neurons in the teacher and those same neurons in the student. Such a direct comparison is possible because the connectome uniquely identifies each individual neuron. We also measure the similarity of teacher and student single-neuron parameters. We refer to the dissimilarity between teacher and student activities or parameters as the 'error' associated with each respective quantity. We note that our notion of similarity between teacher and student is more precise than requiring similarity of collective dynamics as measured through dimensionality reduction methods, such as principal component analysis. Indeed, matching such dynamics can be accomplished by recording a small number of neurons without access to a connectome[33,34].

We built a teacher network that performs a flexible sensorimotor task. Specifically, the network implements a variant of the cycling task[35], which requires the production of oscillatory responses of different durations in response to transient sensory cues (Fig. 1a and Methods). In the network, firing rates are a non-negative smooth function of the input currents, and the unknown single-neuron parameters are the gains and biases (Fig. 1b, left). The synaptic weight matrix is sparse, and neurons are either excitatory or inhibitory (Fig. 1b, right).

We trained multiple students to generate the same readout as the teacher. Each student is initialized with different gains and biases before being trained via gradient descent. Trained networks successfully reproduce the teacher's readout (Fig. 1c,f). However, the error in the neural activity of the student, compared to the teacher, increases over training epochs (Fig. 1d). As a baseline, we computed the error of a student whose neurons match the activities of all neurons in the teacher but with shuffled identities (gray line in Fig. 1d). In this baseline, the manifold of neural activity is the same in teacher and student but not the activity of single neurons. In all networks, the error in activity after training remains above this baseline, indicating that training does not

https://doi.org/10.1038/s41593-025-02080-4

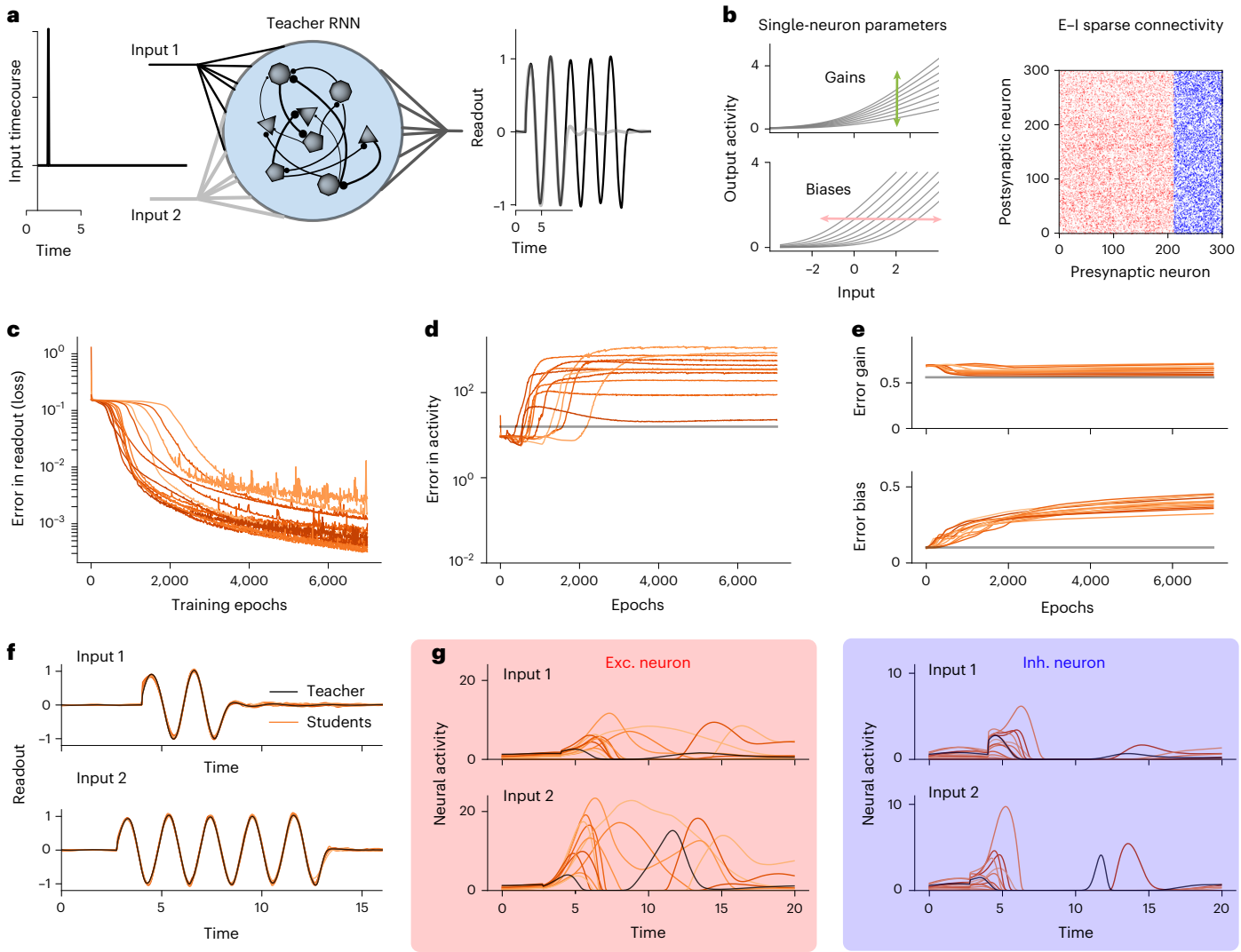


Fig. 1 | Task-trained networks with the same connectivity. a, A teacher RNN is trained to generate two different readout sequences in response to input pulses that produce two different patterns of activation (gray and black). **b**, Properties of the teacher RNN. The teacher RNN has heterogeneous single-neuron parameters (gains and biases of activation functions, left) and sparse connectivity with connection probability $p = 0.5$ (right), and neurons connect through either excitatory (E, red) or inhibitory (I, blue) synapses. **c**, Student networks with the same synaptic weights as the teacher are trained to produce the teacher's output. Error in the readout (training loss, mean squared error) as a function of training epoch. Each colored line corresponds to a different student network. **d**, Error (mismatch in neural activity) between teacher and student RNNs. Gray line, for reference, corresponds to the average error in activity when the student reproduces the teacher's activity but with shuffled neuron identities. **e**, Error in gains and biases versus training epoch. **f**, Readout of teacher and student networks after training, for the two trial types (top and bottom). Teacher and student networks both solve the task. **g**, Neural activity of an example excitatory (left) and inhibitory (right) neuron. Teacher and student neurons exhibit different single-neuron dynamics. Exc., excitatory; Inh., inhibitory.

produce a correspondence between the function of individual teacher and student neurons. Examining the activities of individual neurons shows that neuronal dynamics across different student networks are highly variable, and all students differ from the teacher (Fig. 1g).

Finally, we examined the error in single-neuron parameters between teacher and student (Fig. 1e). The error in gains varies little over training and is similar to a randomly shuffled baseline. The error in biases grows slightly but remains within the same order of magnitude as the baseline.

We conclude that knowledge of synaptic weights and task output is not always enough to predict the activity of single neurons in recurrent networks. For the task we considered, there is a degenerate space of solutions, with different combinations of single-neuron gains and biases, that solve the same task. There may be scenarios for which this degeneracy is reduced, such as small networks optimized for highly specific functions or networks trained on complex or high-dimensional task spaces (Discussion). Nonetheless, our results show that, even with N^2 connectivity constraints, task-optimized neural dynamics are, in general, highly heterogeneous.

Student network constrained by activity recordings

We next asked whether these conclusions change if, instead of recording only task-related readout activity, we record the activity of a subset of neurons in the teacher network. We use $M \leq N$ to denote the number of recorded neurons. Students are trained to reproduce this recorded activity, which provides additional constraints on the solution space (Fig. 2a,b). The recording of subsampled activity in the teacher is analogous to neural recordings in imaging or electrophysiology studies, where only a subset of neurons is registered. We trained two types of student networks: students that have access to the teacher connectome and students that are not constrained in connectivity. For connectome-constrained students (Fig. 2c,e), single-neuron

https://doi.org/10.1038/s41593-025-02080-4

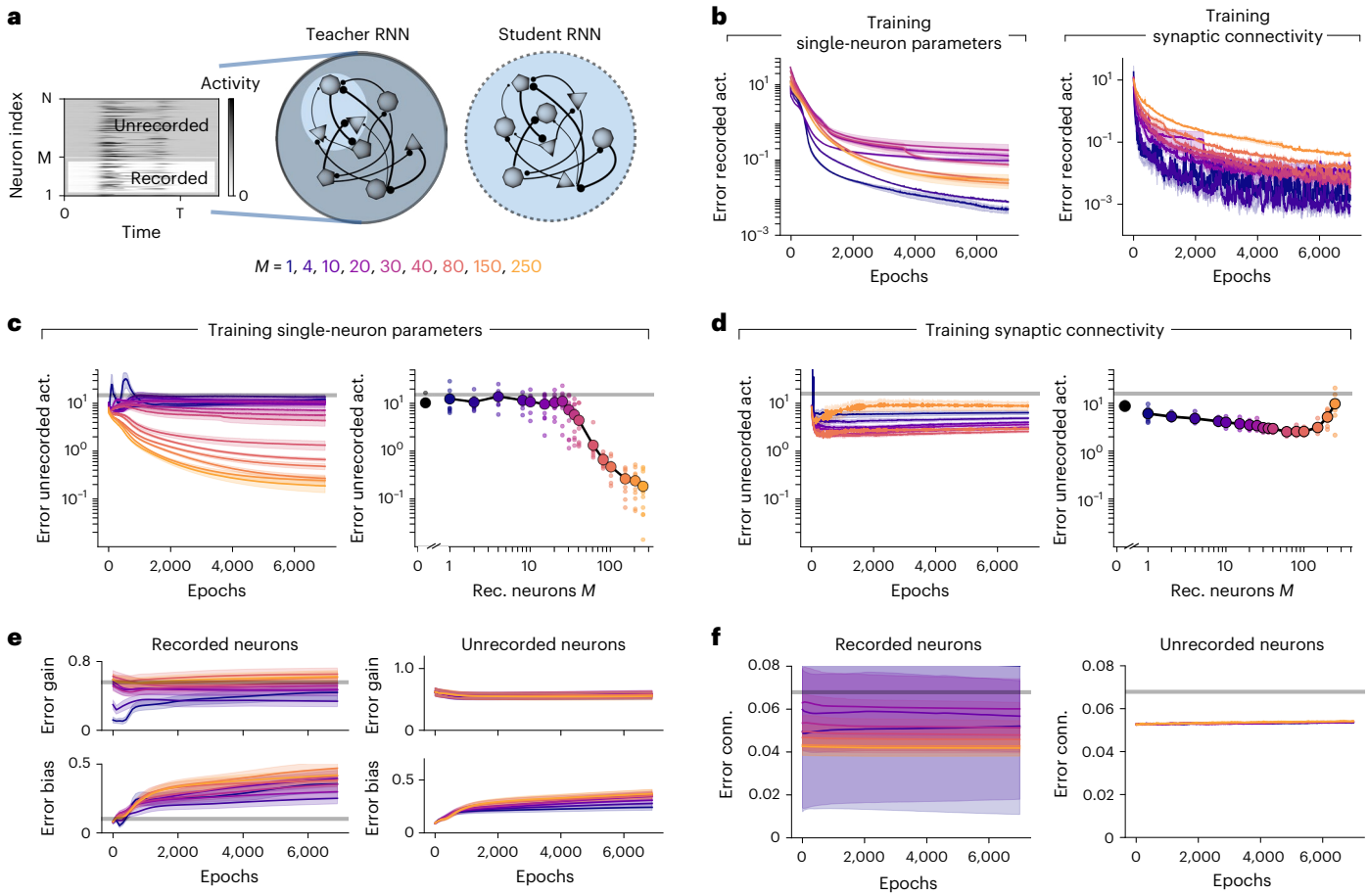


Fig. 2 | Predicting activity of unrecorded neurons when the activity of a subset of the network is observed. a, The student RNN is trained to mimic the activity of M recorded neurons in a teacher RNN. **b**, Error in recorded activity (loss) versus training epoch for students with trained single-neuron parameters (left) and students with trained synaptic weights (right). Lines correspond to different numbers of recorded neurons M and show mean over 10 random seeds. Error bands in all panels indicate ±s.e.m. All students successfully reproduce the recorded activity of the teacher after training. **c**, Left, error in activity of the $N - M$ unrecorded neurons versus training epoch. Right, error in unrecorded neuronal activity after training, as a function of number of recorded neurons M. Smaller dots correspond to each of the 10 trained students. Error is substantially reduced when recording from $M > 30$ neurons. The error corresponding to zero recorded neurons (black dots) is the error of the student network prior to training, with random single-neuron parameters. Gray line denotes shuffled baseline as in Fig. 1d. **d**, Analogous to **c** but training synaptic weights instead. The error in the activity of unrecorded neurons remains high across values of M. The M dependence is a consequence of the procedure of matching neurons across teacher and student (Methods). **e**, Error in gains and biases versus training epochs. Left, parameters of recorded neurons. Right, parameters of unrecorded neurons. **f**, Analogous to **e** for synaptic weights of recorded neurons (left) and unrecorded neurons (right). act., activity; Rec., recorded; conn., connectivity.

parameters of both recorded and unrecorded neurons are unknown and, therefore, trained. For students with unconstrained connectivity, synaptic weights are trained instead. In this case, the single-neuron parameters of the student are set equal to those of the teacher so that the networks differ only in synaptic weights (Fig. 2d,f). Additionally, because there is no direct map between unrecorded neurons in the teacher and the student when the connectome is not known, after training we searched for the mapping between student and teacher neurons that minimizes the mismatch in unrecorded activity at each training epoch (Methods).

We found that both connectome-constrained and unconstrained students are able to mimic the activity of the M recorded teacher neurons with small errors (Fig. 2b; the teacher has $N = 300$ neurons). We then asked whether this holds for the unrecorded neurons. When the connectivity is provided (Fig. 2c), the error for unrecorded neurons is reduced to values similar to the error for recorded neurons when more than $M^* = 30$ neurons are recorded (example task outputs are shown in Extended Data Fig. 1). In comparison, when training the synaptic weights (Fig. 2d), unrecorded neuron activities are not recovered substantially better than baseline even when most neurons are recorded. Thus, connectome-constrained, but not unconstrained, networks produce consistent solutions when M is large enough.

We then assessed whether the students' parameters converge to those of the teacher. For connectome-unconstrained students, the error in synaptic weights remains high, for connections between both recorded and unrecorded neurons (Fig. 2f). We may expect this to occur given that the activity of unknown neurons in these networks is not well predicted (Fig. 2d). More surprisingly, errors in the single-neuron parameters of connectome-constrained networks also remain high, even when the activity of unrecorded neurons is well predicted (Fig. 2e). We did not find qualitative differences in the behavior of single-neuron parameters for recorded and unrecorded neurons.

Thus far, we focused on a teacher whose neural activity is primarily generated through recurrent interactions, triggered by brief external pulses. We further explored whether similar results hold in networks driven by a time-varying external input (Extended Data Fig. 1). Additionally, we systematically varied the distributions of gains and connection sparsity (Extended Data Fig. 1). In all these networks, the qualitative dependence of the error on M was unchanged. Nevertheless, the error in unrecorded neural activity prior to training is different in

https://doi.org/10.1038/s41593-025-02080-4

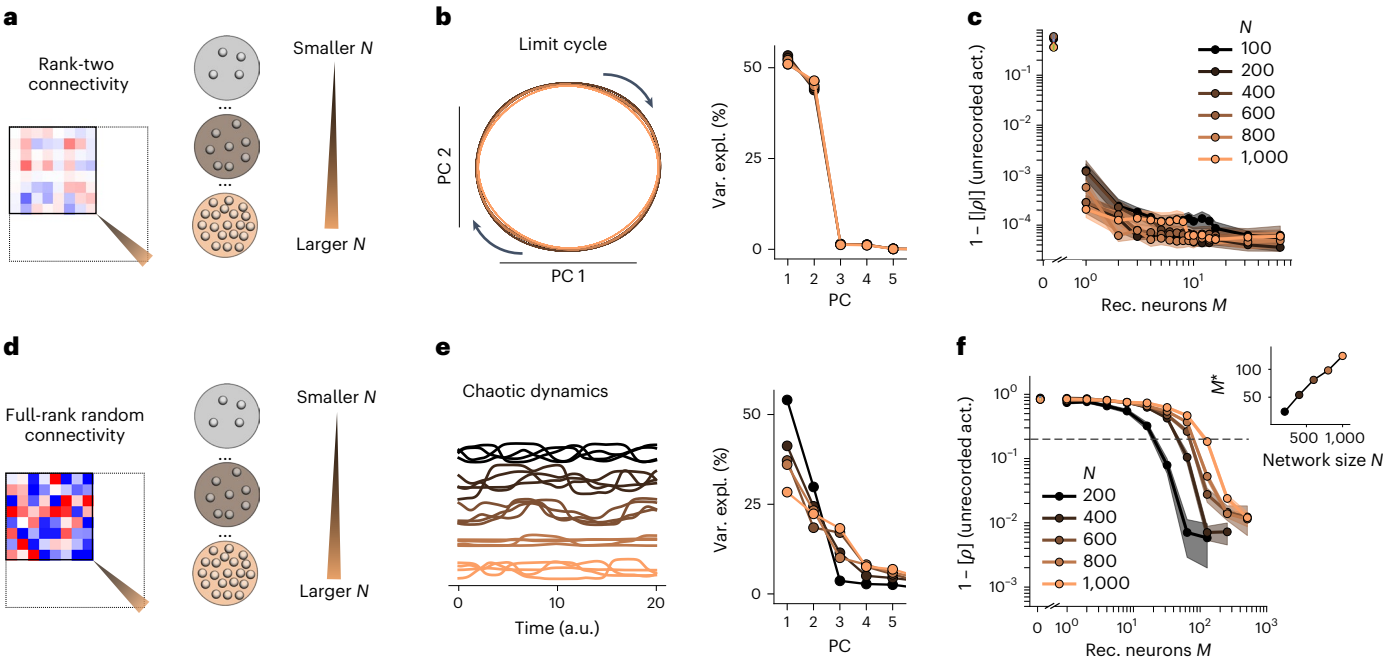


Fig. 3 | Prediction of unrecorded neuron activity depends on dimensionality, not network size. a, Set of teacher RNNs with variable network size N but fixed rank of the synaptic weight matrix (Methods). **b**, Teachers of different size produce the same low-dimensional dynamics. Left, dynamics projected on the top two principal components (PCs). All RNNs generate a limit cycle largely constrained to a two-dimensional linear subspace. Right, variance screeplot. **c**, Error in the activity of unrecorded neurons, after training. We measured the correlation distance between activity in the teacher and student. Plot shows empirical average and s.e.m. for each network size (10 networks per condition). **d**, Set of teacher RNNs with variable network size N and random full-rank synaptic weights. **e**, Left, networks generate high-dimensional chaotic dynamics. Sample activity of four units for networks of different sizes and a time window of 20τ. Right, variance scree plot of the recordings. Larger networks generate higher-dimensional dynamics. **f**, Error in the activity of unrecorded neurons after training. Larger networks require recording from a larger number of neurons M to predict unrecorded activity. Data are presented as mean ± s.e.m. over 10 networks, as in **c**. Inset: number of neurons M^* needed to predict unrecorded activity above a certain threshold (set to 0.2; dotted line), as a function of network size. act., activity; Rec., recorded; Var. expl., variance explained.

networks with strong inputs or weak recurrent connections. Unlike in Fig. 2, where the error prior to training is similar to a baseline with randomly shuffled neuron identities, the error for strongly input-driven networks lies below this baseline even before training. Thus, although certain features of neural activity may be predictable even with random parameters when the input is known, improving upon this initial baseline through training requires sufficiently many recorded neurons.

In summary, connectome-constrained networks are able to predict the activity of unrecorded neurons when further constrained by the activity of enough recorded neurons. By contrast, networks without a connectome constraint do not predict unrecorded activity. Nevertheless, in all cases, the unknown parameters are not precisely recovered, suggesting that multiple sets of biophysical parameters lead to the same neural activity.

Required number of recorded neurons is independent of network size

What features of a connectome-constrained RNN determine how many recorded neurons are required to predict unrecorded activity? We considered two alternatives: the required number is a fixed fraction of the total number of neurons in the network or the number is determined by properties of the network dynamics. The former alternative would pose a challenge for large connectome datasets.

To disambiguate these two possibilities, we examined a class of teacher networks whose population dynamics are largely independent of their size N. We generated networks with specific rank-two connectivity that autonomously generate a stable limit cycle[36] (Fig. 3a and Methods). In these networks, the currents received by each neuron oscillate within a two-dimensional linear subspace, independent of N (Fig. 3b).

We found no difference in a plot of error in unrecorded activity against number of recorded neurons M, for networks of different sizes (Fig. 3c), suggesting that accurate predictions can be made when recording from few neurons, even in large networks. Examining more closely the dependence of the error on M, we observed that when $M = 1$, the student produces oscillatory activity with the same frequency as the teacher, but the activity of unrecorded neurons exhibits consistent errors at particular phases of the oscillation (Extended Data Fig. 2). By contrast, when $M = 7$, errors in recorded and unrecorded neurons are similarly small.

This led us to hypothesize that the number of recorded neurons required to accurately predict neural activity scales with the dimensionality of the neural dynamics, not the network size. This would explain why networks with widely varying sizes but similar two-dimensional dynamics exhibit similar performance (Fig. 3c). To further test this hypothesis, we studied a setting in which we trained students to mimic another class of teacher networks: strongly coupled random networks[37] (Fig. 3d). In such networks, activity is chaotic, and, unlike low-rank networks (Fig. 3b), the linear dimensionality of the dynamics grows in proportion to N (ref. 38), a dependence that we verified for the time windows we considered (Fig. 3e). In this case, the required number of recorded neurons also grows proportionally with N (Fig. 3f). Together, these results suggest that recording from a subset of neurons, on the order of the dimensionality of network activity, is sufficient to predict unrecorded neural activity. Later, we will show that this numerical result is consistent with the predictions of an analytical theory.

Robustness to model mismatch

Thus far, we have considered teacher and student networks that belong to the same model class of firing rate networks with parameterized

https://doi.org/10.1038/s41593-025-02080-4

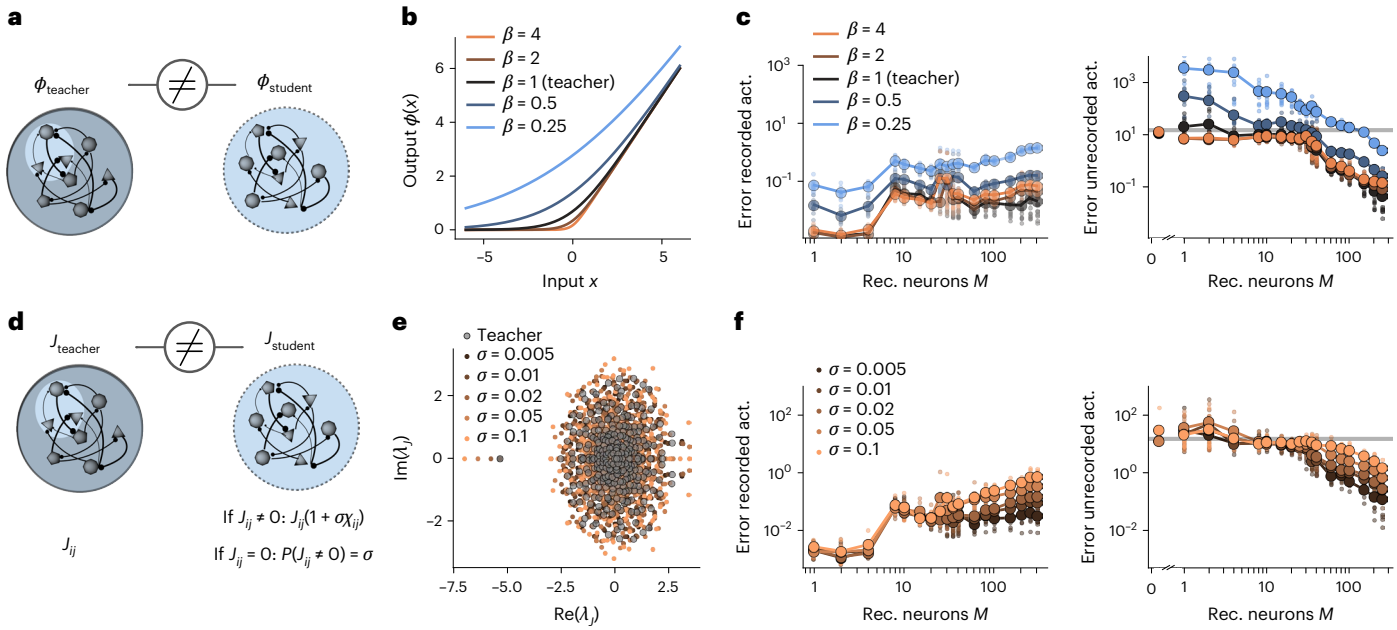


Fig. 4 | Model mismatch between teacher and student. a, Mismatch in activation functions of teacher and student neurons. **b**, The activation function is a smooth rectification but with different degrees of smoothness, parameterized by β. Teacher RNN from Fig. 2. **c**, Errors in the activity of recorded (left) and unrecorded (right) neurons for different values of model mismatch between teacher and student. Across a large range of mismatch, we observe a decrease of the error in unrecorded activity when $M > 30$. **d**, Mismatch in synaptic weights between teacher and student, mimicking errors in connectome reconstruction. **e**, Eigenvalues of the teacher and student synaptic weight matrices, for different levels of mismatch. **f**, Errors in the activity of recorded (left) and unrecorded (right) neurons for different levels of mismatch in the synaptic weights. act., activity; Rec., recorded.

activation functions and connectivity. However, models based on experimental data will possess some degree of 'model mismatch' due to unaccounted or incorrectly parameterized biophysical processes. Moreover, errors in synaptic reconstruction and inter-individual variability in connectomes imply that synaptic weight estimates may also be imprecise[39]. In this section, we examine whether our qualitative results hold when teacher and student exhibit model mismatch.

We used the same teacher as in Figs. 1 and 2. To study the case of mismatch in activation function (Fig. 4a), we parameterized the activation function with β, which controls the smoothness of the rectification, and used different values of β for student and teacher (Fig. 4b). Larger mismatch increases the error in both recorded and unrecorded activity (Fig. 4c). The effect is strongest in an extreme case of very small student β, for which very little rectification occurs. This makes it difficult for the student to match even the recorded activity of the teacher (Fig. 4c and Extended Data Fig. 3). Nevertheless, up to a considerable mismatch, there is a steep decrease in the error in unrecorded activity as more neurons are recorded.

Can model mismatch arising from single-neuron properties be compensated by allowing the synaptic weights to be trained, which introduces additional free parameters? We examined a student with activation function mismatch and a synaptic weight matrix that was initialized equal to that of the teacher but then trained (Extended Data Fig. 3). This performed worse than training single-neuron parameters, arguing against the feasibility of this approach. An alternative approach is to increase the number of single-neuron parameters. For instance, when β is trained together with gains and biases, the error in unrecorded activity is similar to the case without mismatch (Extended Data Fig. 3). We conclude that parameterizing uncertainty in activation function is important for dealing with this form of model mismatch.

We next considered mismatch between teacher and student connectomes. To simulate such errors, we added Gaussian noise to the strengths of existing connections and added spurious connections with probability σ (Fig. 4d and Methods). The resulting corrupted synaptic weight matrix was used by the student. Noise in the synaptic weight matrix shifts its eigenvalues (Fig. 4e) and modifies the corresponding eigenvectors. Trained students exhibit smooth increases of the error in recorded and unrecorded activity as this noise is increased (Fig. 4f). However, we again found a steep decrease of the error in unrecorded neural activity with M, suggesting that this qualitative behavior is not overly sensitive to connectome reconstruction errors.

Teacher networks constrained by empirical connectomes

Thus far, we have examined synthetic teachers, whose connectivity statistics and functional properties may differ from those of biological networks. We next study teachers whose synaptic weights are directly determined by empirical connectome datasets. We modeled three neural circuits for which a ground truth connectome is available and whose function has been characterized: the premotor–motor system in the ventral nerve cord of larval *Drosophila*[16], the heading direction system in the central complex of adult *Drosophila*[9,40] and the oculomotor neural integrator in the hindbrain of larval zebrafish[41].

When larval *Drosophilae* are engaged in forward or backward locomotion, recurrently connected premotor neurons in the ventral nerve cord drive motor neurons to produce appropriately timed muscle activity (Fig. 5a, left). Motor neurons in each body segment are segregated into functional groups whose sequences of activation differ across the two behaviors (Fig. 5a, right). A previous study showed that a connectome-constrained RNN recapitulates features of motor and premotor neuron activity when trained to produce such sequences in the A1 and A2 body segments[16]. We used such a model as a connectome-constrained teacher, whose 178 premotor neurons produce appropriately timed activity in 52 motor neurons (Methods). Student networks comprising the premotor circuitry were then trained to approximate recorded teacher activity. We found that the error in unrecorded activity is reduced when approximately 10 neurons are recorded (Fig. 5b). When few neurons are recorded, the error in activity is similar to a network with randomly chosen single-neuron parameters (two recorded neurons; Fig. 5c, left). Recording from more neurons

https://doi.org/10.1038/s41593-025-02080-4

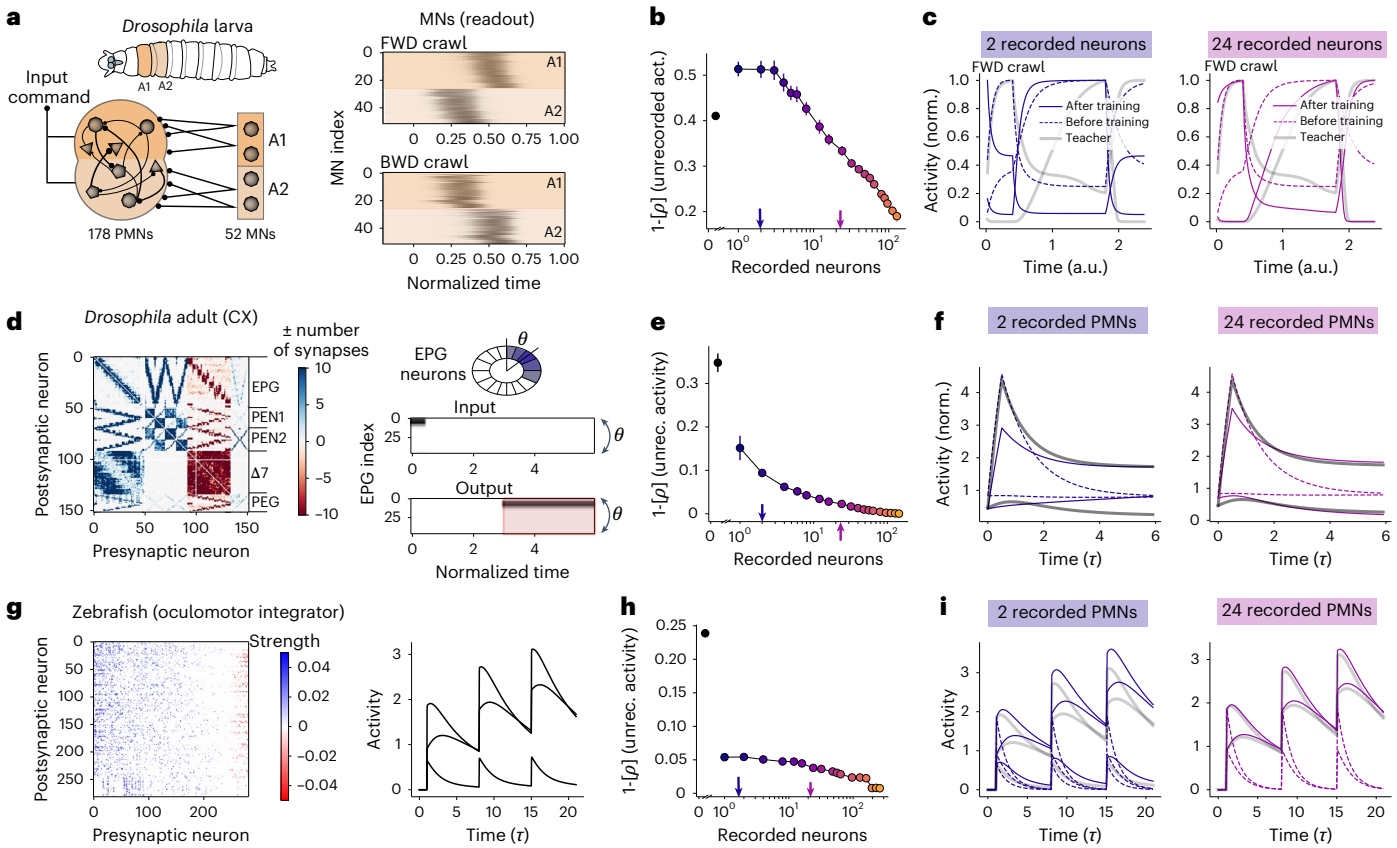


Fig. 5 | Prediction of neural activity in networks with empirical connectome constraints. a, Left, diagram of premotor neurons (PMNs) and motor neurons (MNs) in segments A1 and A2 of the *Drosophila* larva. The model synaptic weights are determined by a connectome, as in Zarin et al.[16]. Right, activity of MN readouts during forward (FWD, top) and backward (BWD, bottom) crawling. **b**, Error (based on the average single-neuron correlation between teacher and student) in unrecorded activity of PMNs as a function of the number of recorded neurons. Black dot indicates error before training—that is, the error before recording any neurons. Arrows indicate number of recorded neurons for which example traces are shown in the adjacent panel. **c**, Example activity traces of two unrecorded PMNs in the teacher network and in the student network before and after training, when two neurons are recorded (left) and when 24 neurons are recorded (right). Traces are normalized by their maximum value. **d**, Left, synaptic weight matrix for neurons of the central complex (CX) of adult *Drosophila*, based on the hemibrain connectome[9]. Right, input and target output of EPG neurons, which are arranged along the ellipsoid body according to their tuning to heading direction angle θ (top). On each trial, the input is a transient bump of activity presented at a random orientation. The target output requires this orientation be held in a persistent activity state. **e**, Similar to **b** for the CX model. **f**, Similar to **d**, example traces of one unrecorded neuron when two different presented stimuli (one preferred and one non-preferred). Traces are normalized by the mean firing rate of each neuron across stimuli. **g**, Left, diagram of synaptic wiring in the zebrafish brainstem, based on Vishwanathan et al.[41]. Right, traces of activity of three example neurons. There is a slow mode of activity that integrates eye velocity and is required for the oculomotor reflex. **h**, Similar to **b** and **e** for the oculomotor integrator model. **i**, Similar to **c** and **f** for the neurons shown in **g**, right. norm., normalized; unrec., unrecorded.

dramatically improves the prediction (Fig. 5c, right), which is qualitatively similar to the results of the synthetic teacher network (Fig. 2c).

Next, we studied the heading direction system in the central complex of adult *Drosophila*. This system has been the subject of numerous recent theoretical analyses, most of which examined models with idealized connectivity rather than directly incorporating connectome data[40,42,43]. We modeled a circuit reconstructed in the hemibrain dataset[9] comprising 153 neurons grouped into four cell types: the putatively excitatory EPG, PEN and PEG neurons and the putatively inhibitory Δ7 neurons (Fig. 5d). The 46 EPG neurons encode heading orientation and are arranged along a ring in the ellipsoid body based on their angular tuning. Recurrent connections among EPG neurons and other cell types form a stable 'bump' of neural activity representing heading angle, consistent with 'ring attractor' dynamical models[44]. We, therefore, constructed a teacher network in which EPG neurons maintained a bump representing a heading encoded by a brief stimulus (Fig. 5d and Methods). Student networks without access to recordings generated neural activity different from the teacher (Fig. 5e, black dots). In particular, these students did not behave as ring attractors, demonstrating that the central complex connectivity alone does not guarantee stable attractor dynamics (Fig. 5f). However, recording from a handful of neurons was enough to place the system in the correct dynamical regime and accurately predict the activity of unrecorded neurons (Fig. 5f).

Finally, we studied the oculomotor integrator in the hindbrain of larval zebrafish. This system persistently tracks eye position by integrating eye motor commands. The integration is supported by strong recurrent connections that produce a 'line attractor' in neural activity space. Such dynamics were previously modeled with a connectome-constrained linear RNN[41] (Fig. 5g and Methods). We used this network as the teacher and then trained the gains of student networks with the same synaptic weights. Although a random initialization of gain parameters did not produce the slow timescale necessary for accurate integration, recording from a few neurons substantially reduced the error in activity (Fig. 5h). This is consistent with the results of Vishwanathan et al.[41], who adjusted a global gain parameter to produce a slow timescale.

The weight matrices of empirical connectomes and synthetic teacher–student networks (Figs. 2 and 3) may exhibit statistical differences due to the level of sparsity, heterogeneity in the number and strength of synaptic connections and other higher-order structure.

However, in each of these examples, the qualitative phenomena present in synthetic teacher–student networks are recapitulated. Recording from a number of neurons determined by the dimensionality of the teacher activity (Extended Data Fig. 4)— a handful for the one-dimensional line attractor or two-dimensional ring attractor dynamics and approximately 10 for more complex sequential activity—produces consistent dynamics between teacher and student.

Linear network model

We developed an analytic theory of our connectome-constrained teacher–student paradigm. The theory aims to explain, first, how the teacher and student produce the same activity despite different single-neuron parameters and, second, the conditions under which the student's activity converges to that of the teacher.

We begin with a simplified linear model and later relax our assumptions: the teacher and student RNNs have linear single-neuron activation functions; the only unknown single-neuron parameters are the biases b_i; and the synaptic weight matrix J has rank D (Fig. 6a). This rank constraint implies that recurrent neural activity is confined to a D-dimensional subspace of the N-dimensional neural activity space. We focus on the network's steady-state activity at equilibrium, which depends linearly on the biases:

$$r_i = \sum_{j=1}^{N} A_{ij} b_j, \tag{3}$$

where we have defined $A \equiv (I - J)^{\dagger} J$.

Although we focus here on equilibrium activity, time-dependent trajectories also yield a linear relation between activity and single-neuron parameters (see Methods for the time-dependent derivation). For the same reason, we also assume no external input to each neuron ($I_i(t) = 0$). This linear relation between single-neuron parameters and activity, which underpins the mathematical tractability of the simplified model, is a consequence of the linear network dynamics and the additive influence of the bias parameters. Choosing multiplicative gains as the unknown single-neuron parameters, for instance, would produce a nonlinear relation.

The student is trained using gradient descent updates to the single-neuron parameters. In the limit of small learning rate η, the learning trajectory in parameter space can be expressed in continuous time t' (with t' proportional to training epoch) as:

$$\frac{db_i}{dt'} = -\eta \sum_{k=1}^{M} \sum_{j=1}^{N} A_{ik}^T A_{kj} \left(b_j - b_j^* \right), \tag{4}$$

where M is the number of recorded neurons. Using these learning dynamics, we can analytically calculate the expected error in recorded and unrecorded activity and in single-neuron parameters (Fig. 6c and Methods). This reveals a transition to zero error in the activity of unrecorded neurons when $M = D$, the rank of the synaptic weight matrix (Fig. 6c, gray line). There are, however, large errors in single-neuron parameters (Fig. 6c, red line) even when the activity of the full network is accurately recovered.

To understand these results, we analyzed the properties of the loss function, which describes how the difference in activity between teacher and student depends on single-neuron parameters. We differentiate the loss function for the full network, which is determined by errors in both recorded and unrecorded neural activity, from the loss function for the recorded neurons, which is the function optimized during training. These loss functions are convex, as illustrated in Fig. 6d. The minima are surrounded by a valley-shaped region of low loss (Fig. 6d, right). We refer to directions for which the loss changes quickly or slowly as 'stiff' or 'sloppy' parameter modes, respectively[45]. Stiff modes both have the greatest effect on the loss and are learned most quickly. Each mode's degree of stiffness is determined by the corresponding singular value of the matrix A (Methods). A mode is infinitely sloppy when its associated singular value is zero, implying that parameter differences between teacher and student along that mode produce no differences in neural dynamics.

The parameter modes that affect the recorded activities and, thus, the loss function for the M recorded neurons are determined by $A_{1:M,:}$ (the submatrix of A containing the rows corresponding to these neurons), whose stiff and sloppy modes are generally different from those of the fully sampled matrix A (Fig. 6e versus Fig. 6f). Recording from a subset of neurons introduces additional modes with zero singular value when $M < D$, because $A_{1:M,:}$ has, at most, M non-zero singular values. The stiff modes of the loss function of the recorded activity will also, typically, not be fully aligned with those of the fully sampled system (Fig. 6f, inset), leading to errors in prediction.

To illustrate these results, we plotted the error in single-neuron activity and biases for M below and above the critical number D (Fig. 6g,h). When recording from few neurons, the error in these parameters for unrecorded neurons remains high (Fig. 6g, left). The error in biases quickly converges to a small value along the stiffest mode, whereas it barely changes for sloppy modes (Fig. 6g, right). The stiffest mode of the subsampled network is not completely aligned with the stiffest mode of the fully sampled network, explaining why it converges to a small but non-zero value. Only when more neurons are recorded does the error in unrecorded activity, and along the stiffest parameter mode, converge to zero (Fig. 6h, left).

The simplified model demonstrates that specific patterns of single-neuron parameters determine the error between teacher and student. Stiff parameter modes are learned, whereas sloppy modes are not. The number of stiff parameter modes is bounded by the rank of J and, thus, the dimensionality of neural activity. Recording from increasingly many neurons provides increasingly many constraints on this activity. When enough neurons are sampled, the stiff modes for the M recorded neurons align with the stiff modes for the full network, leading to correct prediction of unrecorded activity. These conclusions do not rely on the choice of gradient descent as a learning algorithm, as an analysis using linear system identification methods yields the same conclusions (Supplementary Note). We also note that we have assumed here that the loss function is determined by the difference in recorded neural activity. However, similar conclusions would be reached if it were determined by other linear projections of activity, such as projections onto task-related dimensions (Discussion).

Loss landscape

We next generalized our theory to nonlinear networks. To facilitate analysis, we studied a class of low-rank RNNs whose activity can be understood analytically[36,46,47]. We focused on a teacher network with $N = 1{,}000$ neurons and a nonlinear, bounded activation function. Each neuron is parameterized only by the gain parameter. We designed the network's synaptic weight matrix to be rank-two, with two different subpopulations. For this network, there are only two stiff parameter modes: the average single-neuron gain for each subpopulation (Fig. 7a). We set the weights of the teacher to generate two different pairs of non-trivial fixed points, and we recorded activity as the neural dynamics approached one of these fixed points (Fig. 7b).

Because the parameter space is two dimensional, we can visualize the loss function for the full network across a grid of parameters (Fig. 7c). The function has a single minimum, similar to the linear model. However, due to the nonlinearity, the function is non-convex (contour lines are not convex in Fig. 7c), and the curvature for parameter values away from the global minimum is different than at the minimum. Despite this non-convexity, gradient descent on this fully sampled loss function will still approach the single minimum.

We next visualized the loss function for the activity of one recorded neuron (Fig. 7d). We repeated this for two different choices of the single recorded neuron, each of which exhibited distinct dynamics

https://doi.org/10.1038/s41593-025-02080-4

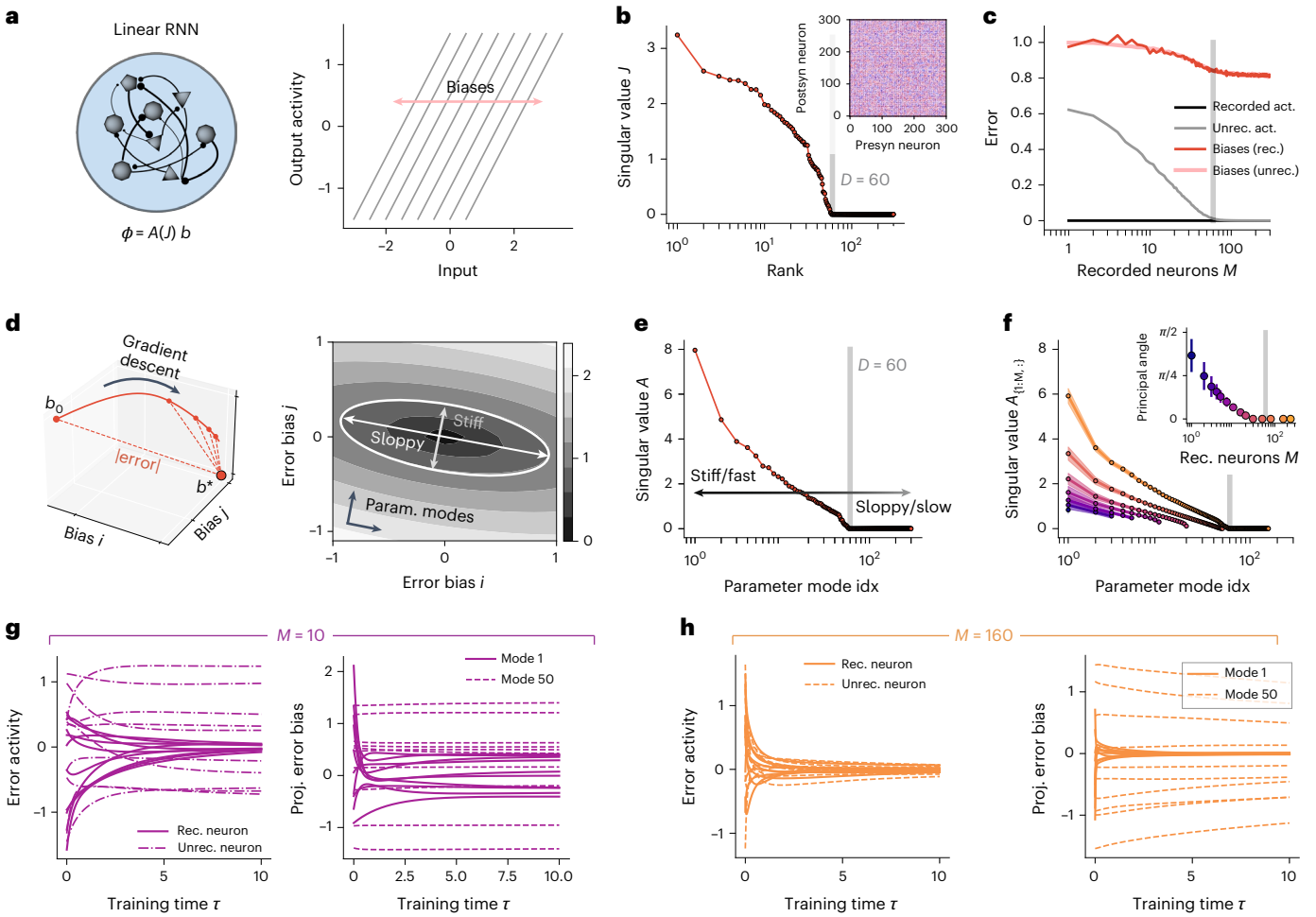


Fig. 6 | Linear teacher–student model. a, Left, the activity of a neuron is a linear mapping $A(J)$, which depends on the synaptic weight matrix J of the single-neuron parameters b. Right, neuronal activation functions are linear with heterogeneous biases. **b**, Singular values of the synaptic weight matrix, which is random and has rank $D = 60$ (gray line). **c**, Errors in activity and biases as a function of the number of recorded neurons. **d**, Single-neuron biases evolve over training through gradient descent. Parameter modes are described as stiff or sloppy based on the effect of changes along each mode near the optimal solution. **e**, Singular value decomposition of the mapping A determines stiff and sloppy parameter modes. Stiffer modes are learned more quickly. Gray line at $D = 60$ corresponds to the rank of the synaptic weight matrix. **f**, Effective singular value decomposition when recording from a subset of M neurons. Inset shows the maximum angle between the M stiffest modes and the M subsampled parameter modes. Different colors correspond to different numbers of recorded neurons M (as shown in the inset). **g**,**h**, Evolution of errors in activity and biases for $D > M = 10$ (**g**) and $D < M = 160$ (**h**), for 10 different initializations of parameters. Error in biases is projected along one stiff (1st) and one sloppy (50th) parameter mode. act., activity; Param., parameter; Postsyn, postsynaptic; Presyn, presynaptic; idx., index; rec., recorded; unrec., unrecorded.

(black lines in Fig. 7b). For these loss functions, there is an additional sloppy mode that is not present in the fully sampled loss (black valleys in Fig. 7d). These results are similar to those of the linear case, although due to the nonlinearity, the sloppy modes correspond to curved regions in parameter space.

The sloppy mode is different for each of the two recorded neurons. When running gradient descent on these subsampled loss functions, randomly initialized parameter values will evolve toward the dark regions of Fig. 7d—for example, toward the blue dot when neuron 1 is sampled or toward the red dot when neuron 2 is sampled. However, both of these two solutions produce high error in unrecorded activity (Fig. 7e,f). This mismatch in unrecorded activity occurs because recording from a single neuron constrains activity along only one dimension of the two-dimensional activity space defined by the rank-two synaptic weight matrix (Fig. 6e).

To test whether the same insights also apply to nonlinear networks with high-dimensional parameter spaces, we computed the stiff and sloppy modes of the fully sampled loss function in the network of Figs. 1 and 2. We approximated the loss function in parameter space to second order at the optimum. We then projected the average error in parameter space, before and after training, along the estimated stiff and sloppy modes (Fig. 7g,h). When few neurons are recorded, the average changes in parameter space before and after training are not aligned with the stiff modes of the loss function. However, when recording from many neurons, there is a large decrease in error along the estimated stiff modes while the error along sloppy modes barely changes, as predicted by our theory. Thus, a second-order approximation of the non-convex loss function qualitatively describes the behavior of gradient descent. Some other effects of non-convexity, however, cannot be explained by the linear theory—for instance, the growth in errors in the bias parameters over the course of training (Figs. 1e and 2e).

We conclude that the qualitative behavior of the linear model holds for the nonlinear networks studied in previous sections. Specifically, when the loss function is determined by recordings of a small number of neurons, the parameter modes become sloppier on average, and new sloppy parameter modes are added that do not align with those of the fully sampled loss function.

https://doi.org/10.1038/s41593-025-02080-4

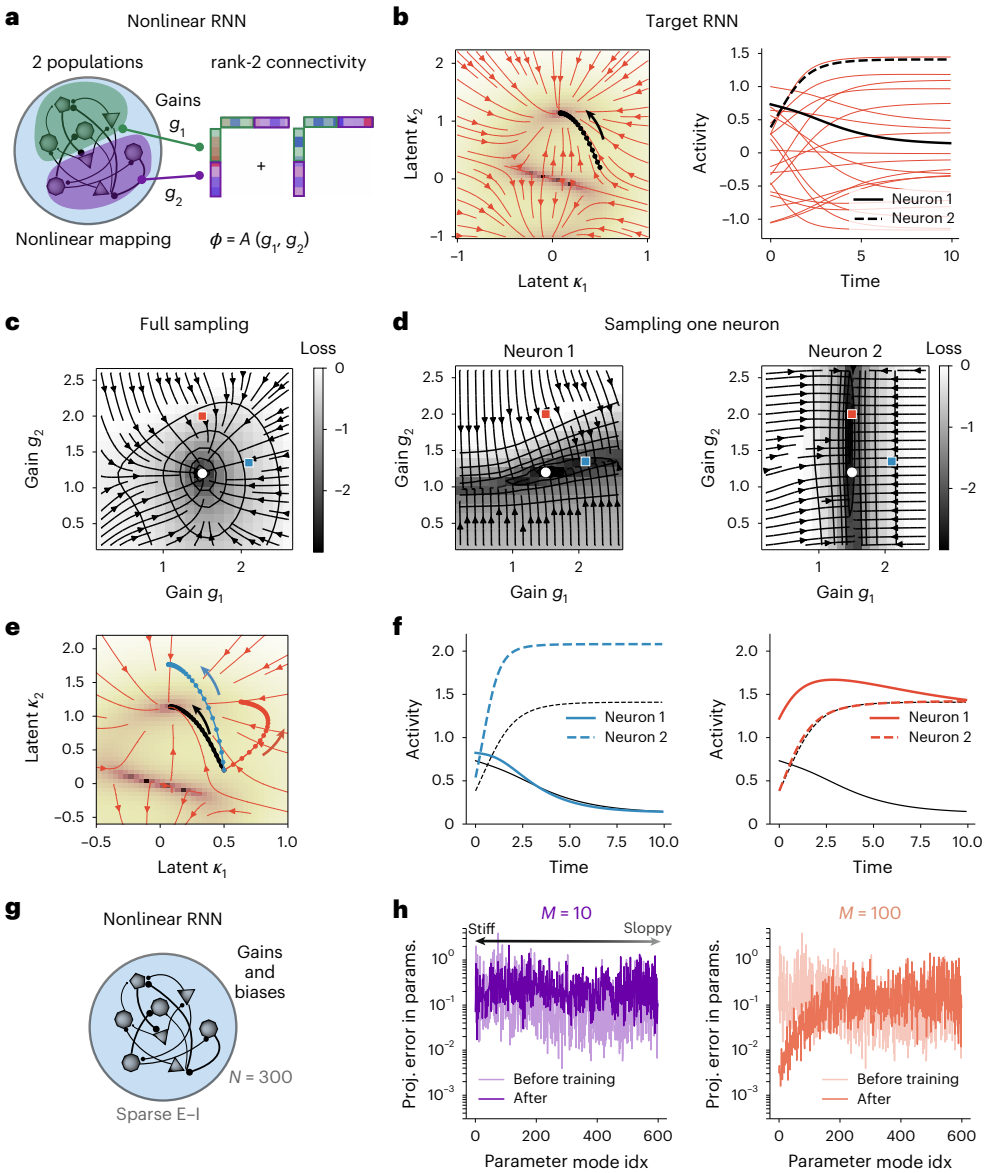


Fig. 7 | Loss landscape in nonlinear networks. a, We study a rank-two RNN with two populations. Neurons in each population share the same gains and connectivity statistics. **b**, Dynamics of the target network. Left, phase space of the two-dimensional latent variables. Right, activity as a function of time for 20 sampled neurons. For illustrative purposes, neurons 1 and 2 are selected based on their alignment with the two latent variables. **c**, The loss function of the full network depends only on the gains of each population, g_1 and g_2. White dot indicates the parameters of the teacher RNN. **d**, Loss function when recording the activity of neuron 1 (left) or neuron 2 (right). Blue and red squares correspond, respectively, to solutions where the training loss is close to zero. **e**, Target trajectory (black) and dynamics of the teacher RNN. Blue and red trajectories correspond to the solutions found in **d**. **f**, Predicted activity for neurons 1 and 2 for the solutions found in **d**. Left, error in the activity of the recorded neuron (neuron 1) is small, whereas error for the unrecorded neuron (neuron 2) is large. Right, similar to left but when neuron 2 is recorded and neuron 1 is unrecorded. **g**, Trained nonlinear RNN from Fig. 1. **h**, Average squared error in parameters projected on the different stiff and sloppy parameter modes. The stiff and sloppy dimensions are determined by approximating the full-sampled loss function around the teacher's values (Methods). Average over 10 realizations. E, excitatory; I, inhibitory; params., parameters; Proj., projected; idx., index.

Optimal selection of single neurons

Thus far, recorded neurons have been selected randomly from the teacher network. As we have seen, different sets of recorded neurons define different loss functions and gradient descent dynamics, suggesting the possibility of selecting recorded neurons to minimize the expected error in unrecorded activity (Fig. 8a). Specifically, we aim to select recorded neurons to maximize the alignment of stiff modes of the subsampled loss and those of the fully sampled loss function.

In the simplified linear model, subsampling neurons corresponds to selecting rows of the matrix A that relates single-neuron parameters to activity (Fig. 8b). In this case, it is possible to exactly determine which neurons are most informative to record. The most informative neuron i is the one whose corresponding row $A_{i,:}$ overlaps most with the weighted left singular vectors of A (Methods). The second most informative neuron is the one whose row overlaps most with the weighted left singular vectors of A projected onto the space orthogonal to the previously selected neuron's row and so on. It is also possible to define the least informative sequence of recorded neurons by minimizing rather than maximizing these overlaps. We compared the error in unrecorded activity for the most and least informative sequence of selected neurons as well as random selection, finding that the optimal strategy indeed improves the efficiency of training (Fig. 8d).

For nonlinear networks, the mapping between parameters and network activity is also nonlinear and depends on the unknown parameters

https://doi.org/10.1038/s41593-025-02080-4

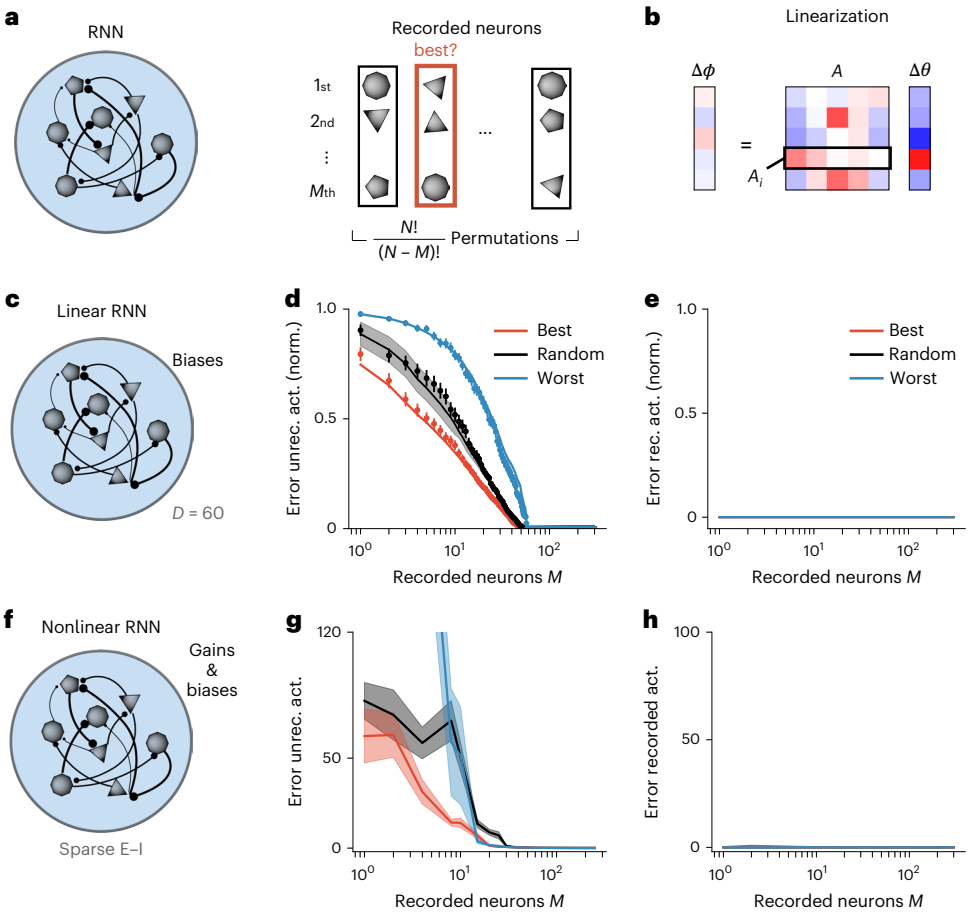


Fig. 8 | Optimal selection of recorded neurons. a, Recording from specific subsets of neurons (right) in the teacher RNN leads to different performance. **b**, We linearized the mapping from changes in single-neuron parameters to changes in neural activity. **c**, Teacher RNN with linear single-neuron activation functions, unknown biases and synaptic weight matrix with rank $D = 60$ (as in Fig. 6). **d**, Error in activity of unrecorded neurons as a function of number of recorded neurons M. Lines correspond to theoretical prediction, dots to numerical simulation (mean ± s.e.m.). We selected neurons following the estimated best ranking (red), five different random rankings (black) and the worst ranking (blue). **e**, Error in recorded neurons for the same networks. **f**–**h**, Analogous to **c**–**e** but for a nonlinear network; data show mean ± s.e.m. The teacher is the RNN from Fig. 2. Single-neuron parameters are both gains and biases. The linearization in the mapping from parameters to activity assumes homogeneous single-neuron parameters (Methods). Note the linear scale for the error in **g**, which highlights the errors when few neurons are recorded. act., activity; E, excitatory; I, inhibitory; norm., normalized; rec., recorded; unrec., unrecorded.

of the teacher (Methods). As a result, the globally optimal sequence cannot be determined a priori. Nevertheless, the mapping between parameters and activity can be linearized based on an initial guess of the single-neuron parameters and then iteratively refined. In practice, we found that linearization works well for nonlinear networks, with the optimal selection strategy dramatically reducing the error compared to random selection, especially when there are few recorded neurons. For the network studied in Fig. 2, the error using the best 10 predicted neurons is 60% smaller than random selection (Fig. 8g).

The singular vectors used to determine which neurons are most informative depend on the global connectivity structure and cannot be exactly reduced to any single-neuron property. Such properties, including in-degree, out-degree, average synaptic strength or neuron firing rate, may be correlated with the singular value decomposition score developed here but are not guaranteed to be good proxies for informativeness. This argues for the use of models like those studied here to guide the selection of recorded neurons.

Discussion

Building connectivity-constrained neural network models has become increasingly viable as the scale of connectome datasets has grown. Our theory cautions against overinterpreting such models when they are insufficiently constrained (Fig. 1) but also shows that correctly parameterized models paired with sufficiently many neural recordings can provide consistent predictions (Figs. 2 and 5). This consistency is a consequence of the qualitatively different solution spaces associated with connectome-constrained and unconstrained models (Figs. 2c,d and 7). The theory also suggests that models can be used to inform targets for physiological recordings (Fig. 8).

Challenges for connectome-constrained neural networks

We have studied the properties of connectome-constrained neural networks using simulations of neural activity based on synthetic network models and three different connectomics datasets[9,10,41]. Our results suggest that the 'forward problem' of predicting neural activity using a connectome is not as ill-posed as the corresponding 'inverse problem' studied previously[1]. However, although we demonstrated that this result is robust to model mismatch and inaccuracy in synaptic reconstruction (Fig. 4), it is likely that, for some neural systems, the degree of model mismatch is too severe. Such systems likely include those largely driven by unmodeled processes such as the effects of neuropeptides or gap junction couplings[13,23]. Moreover, systems for which the firing rate models described here are a poor match, such as systems that operate based on spike synchrony rather than rate codes[48], highly compartmentalized interactions[49] or dynamics of specific ion channels[50], may be out of reach of the present approach. Nonetheless,

our results establish that the dynamics in RNNs with order N, rather than N^2, unknown parameters can be accurately predicted.

We also assumed in all our analyses that time-varying external inputs, together with the initial state, are known. Consistent with this, recent work using connectome information to infer function has focused on regions close to the sensory periphery[19,51], where input statistics are better characterized. We do not expect connectomes to provide substantial constraints on strongly input-driven neural activity when inputs are not controlled.

Relatedly, for our studies of empirical connectomes (Fig. 5), we modeled systems for which detailed descriptions of the function of individual cell types exist[16,41,42]. This was necessary to generate realistic teacher activity, as recordings of neural activity aligned to whole-brain connectomes are not yet available for the systems that we studied. We did not apply our theory to *C. elegans* recordings as it has been argued that chemical synapses are not predictive of functional interactions[13].

We note that the parameters in our models may reflect state-dependent modulation. Neuromodulators, for instance, are known to modify effective neuronal excitabilities[27]. In our networks, gains and biases do not necessarily account for a single biophysical process but, rather, the coordinated effects of multiple processes. As long as the timescale of these processes is slower than the dynamics being predicted, we expect an approach similar to the one described here to be appropriate. However, this state dependence may also imply that the inferred parameters do not generalize to new behavioral states.

Assessment of connectome-constrained solutions

It is known that recordings from $M \geq D$ different neurons are required to estimate neural dynamics lying in a manifold of linear dimensionality D, independent of network size[33]. Connectome-constrained models go beyond such population-level descriptions of neural dynamics, as they are also concerned with how each specific neuron contributes to global activity patterns. This requires knowledge of unrecorded neurons' loadings onto the low-dimensional manifold. This benchmark is appropriate when such models are used to predict the function of specific neurons or neuron types or to guide experiments that manipulate specific neurons[14,16,19].

The match between student and teacher activities in our models depends on multiple properties. These include whether random choices of single-neuron parameters produce similar dynamics in the two networks (Fig. 3c and Extended Data Fig. 1), the extent of model mismatch (Fig. 4) and the degree to which training the student using activity recordings may compensate for the mismatch (Extended Data Fig. 3). These properties depend on specific features of the teacher network. Recent studies have demonstrated above-chance prediction of function using uniform or random parameters in models of the *Drosophila* nervous system[14,19]. In one case, accurate predictions of motor neuron responses to optogenetic stimulation was achieved without any adjustment of single-neuron parameters, which may reflect the presence of strong and direct excitatory pathways between stimulated neurons and output neurons[14]. In another case, further training of single-neuron parameters based on a task objective of detecting visual motion led to an improvement in predictions[19]. On the other hand, failure of a related approach in *C. elegans* was argued to be a consequence of model mismatch from unmodeled peptidergic interactions[13].

We found that training student networks to match the teacher only led to improvements when, in addition to connectivity constraints, sufficiently many neuron activities were constrained. This result was independent of the initial performance of the system with random parameters (Extended Data Fig. 1). We focused on the improvement that can be achieved through knowledge provided by neural recordings, which was motivated by the observation that training the student only on the task-related readout of a teacher did not predict unrecorded neural activity (Fig. 1). However, it is possible that high-dimensional task readouts may improve predictions. Indeed, for our linear model, constraining activity along one task-related dimension is analogous to recording one additional neuron, as both correspond to a linear projection of the network's vector of neural activities. In networks that perform multiple tasks or process diverse inputs, recording activity under multiple task or input conditions may improve the prediction of unrecorded activity, as has been argued for the *Drosophila* visual system[19]. This would likely require an assumption that single-neuron parameters are not strongly modulated across these conditions.

Properties of connectome-constrained and unconstrained network solutions

The loss functions of task-trained feedforward neural networks have been shown to exhibit multiple minima, with often counterintuitive geometrical properties[52,53]. The multiplicity of minima arises from symmetries such as weight permutations in the network parameterization[54]. It remains unclear whether such ideas extend to recurrent neural networks. Our results for connectome-constrained networks are consistent with the existence of a single minumum or connected set of minima with stiff and sloppy parameter modes around the optimal solution (Fig. 7). The alignment between sloppy parameter modes in subsampled versus fully sampled loss functions explains the success in generalizing to unrecorded neurons.

Robustness to a large range of structural parameters and perturbations is a hallmark of biological systems, with a few stiff parameter combinations determining function[45,55–59]. We have shown that this is also true of connectome-constrained networks. One consequence of this observation is that, in underconstrained, data-driven models for neuroscience and machine learning, the distribution of parameters, such as synaptic weights or single-neuron excitabilities found after successful training, may not be predictive of task performance. Our work argues in favor of identifying stiff parameter combinations in such networks and using these to assess the similarity of network solutions[60].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-025-02080-4.

References

1. Das, A. & Fiete, I. R. Systematic errors in connectivity inferred from activity in strongly recurrent networks. *Nat. Neurosci.* **23**, 1286–1296 (2020).
2. Haber, A. & Schneidman, E. Learning the architectural features that predict functional similarity of neural networks. *Phys. Rev. X* **12**, 021051 (2022).
3. Levina, A., Priesemann, V. & Zierenberg, J. Tackling the subsampling problem to infer collective properties from limited data. *Nat. Rev. Phys.* **4**, 770–784 (2022).
4. Liang, T. & Brinkman, B. A. Statistically inferred neuronal connections in subsampled neural networks strongly correlate with spike train covariances. *Phys. Rev. E* **109**, 044404 (2024).
5. Dinc, F., Shai, A., Schnitzer, M. & Tanaka, H. CORNN: convex optimization of recurrent neural networks for rapid inference of neural dynamics. *Adv. Neural Inf. Process. Syst.* **36**, 51273–51301 (2023).
6. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).
7. Ohyama, T. et al. A multilevel multimodal circuit enhances action selection in *Drosophila*. *Nature* **520**, 633–639 (2015).
8. Zheng, Z. et al. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell* **174**, 730–743 (2018).

9. Scheffer, L. K. et al. A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9**, e57443 (2020).
10. Dorkenwald, S. et al. Neuronal wiring diagram of an adult brain. *Nature* **634**, 124–138 (2024).
11. Hildebrand, D. G. C. et al. Whole-brain serial-section electron microscopy in larval zebrafish. *Nature* **545**, 345–349 (2017).
12. Turner, M. H., Mann, K. & Clandinin, T. R. The connectome predicts resting-state functional connectivity across the *Drosophila* brain. *Curr. Biol.* **31**, 2386–2394 (2021).
13. Randi, F., Sharma, A. K., Dvali, S. & Leifer, A. M. Neural signal propagation atlas of *Caenorhabditis elegans*. *Nature* **623**, 406–414 (2023).
14. Shiu, P. K. et al. A *Drosophila* computational brain model reveals sensorimotor processing. *Nature* **634**, 210–219 (2024).
15. Pospisil, D. A. et al. The fly connectome reveals a path to the effectome. *Nature* **634**, 201–209 (2024).
16. Zarin, A. A., Mark, B., Cardona, A., Litwin-Kumar, A. & Doe, C. Q. A multilayer circuit architecture for the generation of distinct locomotor behaviors in *Drosophila*. *eLife* **8**, e51781 (2019).
17. Keller, A. J. et al. A disinhibitory circuit for contextual modulation in primary visual cortex. *Neuron* **108**, 1181–1193 (2020).
18. Kohn, J. R., Portes, J. P., Christenson, M. P., Abbott, L. F. & Behnia, R. Flexible filtering by neural inputs supports motion computation across states and stimuli. *Curr. Biol.* **31**, 5249–5260 (2021).
19. Lappalainen, J. K. et al. Connectome-constrained networks predict neural activity across the fly visual system. *Nature* **634**, 1132–1140 (2024).
20. Eckstein, N. et al. Neurotransmitter classification from electron microscopy images at synaptic sites in *Drosophila melanogaster*. *Cell* **187**, 2574–2594 (2024).
21. Barnes, C. L., Bonnéry, D. & Cardona, A. Synaptic counts approximate synaptic contact area in *Drosophila*. *PLoS ONE* **17**, e0266064 (2022).
22. Kasai, H., Fukuda, M., Watanabe, S., Hayashi-Takagi, A. & Noguchi, J. Structural dynamics of dendritic spines in memory and cognition. *Trends Neurosci.* **33**, 121–129 (2010).
23. Bargmann, C. I. & Marder, E. From the connectome to brain function. *Nat. Methods* **10**, 483–490 (2013).
24. Marder, E. Neuromodulation of neuronal circuits: back to the future. *Neuron* **76**, 1–11 (2012).
25. Gutierrez, G. J., O'Leary, T. & Marder, E. Multiple mechanisms switch an electrically coupled, synaptically inhibited neuron between competing rhythmic oscillators. *Neuron* **77**, 845–858 (2013).
26. Stroud, J. P., Porter, M. A., Hennequin, G. & Vogels, T. P. Motor primitives in space and time via targeted gain modulation in cortical networks. *Nat. Neurosci.* **21**, 1774–1783 (2018).
27. Ferguson, K. A. & Cardin, J. A. Mechanisms underlying gain modulation in the cortex. *Nat. Rev. Neurosci.* **21**, 80–92 (2020).
28. Connors, B. W. & Gutnick, M. J. Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci.* **13**, 99–104 (1990).
29. Kubota, Y., Hatada, S., Kondo, S., Karube, F. & Kawaguchi, Y. Neocortical inhibitory terminals innervate dendritic spines targeted by thalamocortical afferents. *J. Neurosci.* **27**, 1139–1150 (2007).
30. Perich, M. G. et al. Inferring brain-wide interactions using data-constrained recurrent neural network models. Preprint at *bioRxiv* https://doi.org/10.1101/2020.12.18.423348 (2021).
31. Seung, H. S., Sompolinsky, H. & Tishby, N. Statistical mechanics of learning from examples. *Phys. Rev. A* **45**, 6056 (1992).
32. Saad, D. & Solla, S. A. Exact solution for on-line learning in multilayer neural networks. *Phys. Rev. Lett.* **74**, 4337 (1995).
33. Gao, P. et al. A theory of multineuronal dimensionality, dynamics and measurement. Preprint at *bioRxiv* https://doi.org/10.1101/214262 (2017).
34. Kim, C. M., Finkelstein, A., Chow, C. C., Svoboda, K. & Darshan, R. Distributing task-related neural activity across a cortical network through task-independent connections. *Nat. Commun.* **14**, 2851 (2023).
35. Russo, A. A. et al. Motor cortex embeds muscle-like commands in an untangled population response. *Neuron* **97**, 953 (2018).
36. Beiran, M., Dubreuil, A., Valente, A., Mastrogiuseppe, F. & Ostojic, S. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Comput.* **33**, 1572–1615 (2021).
37. Sompolinsky, H., Crisanti, A. & Sommers, H. J. Chaos in random neural networks. *Phys. Rev. Lett.* **61**, 259 (1988).
38. Clark, D. G., Abbott, L. F. & Litwin-Kumar, A. Dimension of activity in random neural networks. *Phys. Rev. Lett.* **131**, 118401 (2023).
39. Hamood, A. W. & Marder, E. Animal-to-animal variability in neuromodulation and circuit function. *Cold Spring Harb. Symp. Quant. Biol.* **79**, 21–28 (2014).
40. Turner-Evans, D. B. et al. The neuroanatomical ultrastructure and function of a biological ring attractor. *Neuron* **109**, 1582 (2021).
41. Vishwanathan, A. et al. Predicting modular functions and neural coding of behavior from a synaptic wiring diagram. *Nat. Neurosci.* **27**, 2443–2454 (2024).
42. Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).
43. Noorman, M., Hulse, B. K., Jayaraman, V., Romani, S., Hermundstad, A. M. Maintaining and updating accurate internal representations of continuous variables with a handful of neurons. *Nat. Neurosci.* **27**, 2207–2217 (2024).
44. Ben-Yishai, R., Bar-Or, R. L. & Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848 (1995).
45. Prinz, A. A., Bucher, D. & Marder, E. Similar network activity from disparate circuit parameters. *Nat. Neurosci.* **7**, 1345–1352 (2004).
46. Mastrogiuseppe, F. & Ostojic, S. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* **99**, 609–623 (2018).
47. Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F. & Ostojic, S. The role of population structure in computations through neural dynamics. *Nat. Neurosci.* **25**, 783–794 (2022).
48. Singer, W. Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physiol.* **55**, 349–374 (1993).
49. Koch, C. & Segev, I. The role of single neurons in information processing. *Nat. Neurosci.* **3**, 1171–1177 (2000).
50. Goaillard, J.-M. & Marder, E. Ion channel degeneracy, variability, and covariation in neuron and circuit resilience. *Annu. Rev. Neurosci.* **44**, 335–357 (2021).
51. Seung, H. S. Predicting visual function by interpreting a neuronal wiring diagram. *Nature* **634**, 113–123 (2024).
52. Li, H., Xu, Z., Taylor, G., Studer, C. & Goldstein, T. Visualizing the loss landscape of neural nets. *Adv. Neural Inf. Process. Syst.* **31**, 6391–6401 (2018).
53. Fort, S. & Jastrzebski, S. Large scale structure of neural network loss landscapes. *Adv. Neural Inf. Process. Syst.* **32**, 6709–6717 (2019).
54. Simsek, B. et al. Geometry of the loss landscape in overparameterized neural networks: symmetries and invariances. In *Proceedings of the International Conference on Machine Learning* 9722–9732 (PMLR, 2021).
55. Gutenkunst, R. N. et al. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189 (2007).
56. Daniels, B. C., Chen, Y. J., Sethna, J. P., Gutenkunst, R. N. & Myers, C. R. Sloppiness, robustness, and evolvability in systems biology. *Curr. Opin. Biotechnol.* **19**, 389–395 (2008).

57. Fisher, D., Olasagasti, I., Tank, D. W., Aksay, E. R. & Goldman, M. S. A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit. *Neuron* **79**, 987–1000 (2013).
58. Naumann, E. A. et al. From whole-brain data to functional circuit models: the zebrafish optomotor response. *Cell* **167**, 947–960 (2016).
59. Otopalik, A. G. et al. Sloppy morphological tuning in identified neurons of the crustacean stomatogastric ganglion. *eLife* **6**, e22352 (2017).
60. O'Leary, T., Sutton, A. C. & Marder, E. Computational models in the age of large datasets. *Curr. Opin. Neurobiol.* **32**, 87–94 (2015).

Methods

Recurrent network models

We focused on recurrent neural networks where the activity of each neuron i is described by a continuous variable, a firing rate $r_i(t)$, for $i = 1...N$. The firing rate of each neuron is calculated by applying a parametric function to the input $x_i(t)$ that the neuron receives at each timepoint,

$$r_i(t) = F(x_i; \theta_i), \tag{5}$$

where θ_i denote the single-neuron parameters that modulate the function F. We denote this input-to-rate function F as the activation function. The activation function may depend on various single-neuron parameters θ_i, such as the gains g_i and biases b_i in equation (1).

The dynamics of the recurrent neural network follow

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij} r_j + I_i(t) + \eta_i(t). \tag{6}$$

The matrix J is the synaptic weight matrix, with each element J_{ij} indicating the signed synaptic strength of the connection from neuron j to neuron i. The constant τ indicates the timescale of single neurons. We assume the same τ for all neurons and set it equal to 1, unless specified otherwise, such that all other timescales are given in units of τ. We denote the external input by $I_i(t)$, which includes the task-related input. Additional private white noise may be provided to each neuron, denoted by $\eta_i(t)$.

The dynamical landscape that a network can implement is thus determined by the order-N single-neuron parameters and the N^2 synaptic weights. In the section 'Teacher RNNs and training parameters' below, we specify the choice of activation function, single-neuron parameters and synaptic weight matrices used in each figure.

Teacher–student: setup and training

We focus on a set of two RNNs. The first is the teacher RNN, which represents the network whose connectome is known and from which we can record neural activity. The other is the student RNN, which is trained (that is, its parameters are optimized) to match the recorded neural activity in the teacher. The dynamics of both networks are determined by equation (6). Asterisks are used to refer to teacher network parameters.

Unless otherwise specified, the teacher and the student network share the same synaptic weight matrix J. Additionally, the external input $I_i(t)$ and initial conditions $x_i(t = 0)$ are the same for teacher and student. The noise intensity is chosen to be zero in the teacher, because we focus on the case of no measurement noise. The noise intensity in the students is weak, to provide additional stability over training. The possible structural differences between teacher and student come from the set of single-neuron parameters $\{\theta_i\}$. These single-neuron parameters are optimized so that student activity matches the recorded activity of the teacher. In Fig. 2d and Extended Data Fig. 1, where we train students with unconstrained connectivity, we instead set the single-neuron parameters to be equal but the synaptic weight matrices to be different in teacher and student. In these cases, the student synaptic weights are trained.

The trained parameters are optimized to minimize the loss function

$$\mathcal{L} = \frac{1}{M} \sum_{m=1}^{M} \left[\left(r_m(t) - r_m^*(t) \right)^2 \right], \tag{7}$$

where the square brackets denote an average over the timepoints in the recorded window, and we sorted neurons such that the first M neurons are those that are recorded. In Fig. 1, instead of defining the loss based on the recorded activity traces $r_m(t)$, we used the task readout, $z = \sum_i w_i^{\text{out}} r_i(t)$.

We trained the parameters of the student using standard gradient descent methods applied to time-varying signals: we implemented backpropagation through time via the Adam optimizer using PyTorch[61–63]. Learning rates varied between 0.0001 and 0.01, and decay rates of the first and second moments varied between 0.9 and 0.999.

Quantifying performance

We used two different metrics to assess the deviations between teacher and student. The first is the root mean squared error. For Figs. 3 and 5, networks where the firing rates are very heterogeneous across neurons, or where the temporal profile of the responses is more relevant (which would be the case when comparing to, for instance, calcium fluorescence traces), we instead used a correlation-based score, which is not affected by the average amplitude of the responses. This score is defined as

$$\text{Error} = 1 - \langle \rho_i \rangle, \tag{8}$$

where the angular brackets indicate the mean over neurons i, and ρ_i is the Pearson correlation coefficient for the i-th neuron:

$$\rho_i = \frac{\left[\left(r_i(t) - \bar{r}_i \right) \left(r_i^*(t) - \bar{r}_i^* \right) \right]}{\sqrt{\left[\left(r_i(t) - \bar{r}_i \right)^2 \right] \left[\left(r_i^*(t) - \bar{r}_i^* \right)^2 \right]}}. \tag{9}$$

The square brackets indicate the average across timepoints, and $\bar{r}_i$ is the time-averaged activity of the i-th neuron. When there are multiple trials (Fig. 5a–f), we calculated one score ρ_i per neuron and trial and then averaged over trials. For Fig. 3c, we took the absolute value of ρ_i before averaging over neurons, because we were interested in whether single neurons of the teacher reproduce the oscillatory dynamics of the teacher, independently of the sign, although our qualitative conclusions do not depend on this choice.

In Fig. 2, to match unrecorded neurons, we paired unrecorded neurons in the teacher with unrecorded neurons in the student at each training epoch, by solving the linear sum assignment problem using the function 'linear_sum_assignment' in Scipy. We calculated the mean squared error between all unrecorded neurons i in the teacher and unrecorded neurons j in the student and stored them in the matrix element C_{ij}. The linear sum assignment problem finds the matrix X such that minimizes $\sum_{i,j} C_{i,j} X_{i,j}$, with the constraint that each row in X maps exactly one single neuron in the teacher to one neuron in the student.

Teacher RNNs and training parameters

In this section, we detail the choice of single-neuron parameters, the features of the teacher RNN and the training parameters used in each figure. Unless otherwise specified, we used $\Delta t = 0.1$ in units of the single-neuron time constant. We injected noise at each timestep of the dynamics with s.d. 0.002. See also the shared code (to be made available upon publication) for reproducing all the numerical experiments in the study.

Figures 1 and 2. The teacher RNN ($N = 300$) is trained with learning rate 0.001 during 1,400 epochs to solve the cycling task. The initial connectivity is chosen as follows. First, each synapse is drawn from a random Gaussian distribution with mean zero and variance $2.4/\sqrt{N}$. The sparse weights (fraction $p = 0.5$) are randomly chosen, set to zero and not trained. A fraction $p_E = 0.7$ of neurons selected randomly is set to be excitatory, such that their synaptic strengths are set to their absolute value, whereas the remaining fraction of neurons, $1 - p_E$, is set to be inhibitory. Synapses are rectified to their assigned sign after each

training epoch. Trials are 20 time units long. The single-neuron activation function is given by Softplus:

$$\text{Softplus}\ (x;\beta) = \beta^{-1}\log\left(1+\exp\left(\beta x\right)\right), \tag{10}$$

where we set the smoothness parameter to $\beta = 1$.

The student RNNs with unknown single-neuron parameters are trained for 7,000 epochs and learning rate 0.001. The student RNN with unconstrained connectivity shares the same single-neuron parameters as the teacher RNN, to facilitate the comparison. Figure 2 shows results for 14 different trained students. The learning rate was set to 0.005. The synaptic weights are initialized randomly following a Gaussian distribution, and the weight signs are correctly assigned (that is, the student knows whether a neuron is excitatory or inhibitory). To compare the weights after training, we picked a random unrecorded neuron from the teacher and matched it with the unrecorded neuron with the most similar activity profile. Then, the selected neurons in teacher and student are discarded, and we picked a new neuron to be matched in the teacher. This procedure is repeated until all neurons are paired.

Figure 3. The teacher networks in the top row are rank-two networks whose synaptic weight matrices are given by:

$$J_{ij} = \frac{1}{N}\left(m_i^{(1)}n_j^{(1)} + m_i^{(2)}n_j^{(2)}\right). \tag{11}$$

The activation function for each neuron is the tanh function, and we consider gains as the only single-neuron parameter. Networks have variable numbers of neurons N, but the distribution of connectivity loadings is given by fixed parameters, such that all the networks generate the same dynamics for large N. The connectivity loadings of the i-th neuron, $\{m_i^{(1)}, m_i^{(2)}, n_i^{(1)}, n_i^{(2)}\}$, are sampled from a four-variable Gaussian distribution with mean 0. The variance parameters are $\sigma_{m^{(r')}m^{(r)}} = \delta_{r'r}$, $\sigma_{n^{(r')}n^{(r)}} = 2\delta_{r'r}$, $\sigma_{n^{(r)}m^{(r)}} = 1.5$ for $r, r' = 1, 2$ and $\sigma_{n^{(1)}m^{(2)}} = -1.5$, $\sigma_{n^{(2)}m^{(1)}} = 1.5$. These parameters lead to a limit cycle in the dynamics[36,47].

The gains are chosen to be Gaussian, with mean 1 and s.d. 0.9, uncorrelated with all the other connectivity loadings. The precise shape of the gain distribution does not affect the dynamics, only its mean and variance, as long as the gains are uncorrelated with the connectivity loadings. We selected trajectories that start on the limit cycle and evolve during 20 time units. The student network is initialized in this case with homogeneous unitary gains and is trained for 7,000 epochs with learning rate 0.005.

The teacher networks in the bottom row have random synaptic weights as in ref. 37, where the J_{ij} are randomly drawn from a Gaussian distribution with mean 0 and s.d. $1.7/\sqrt{N}$. The single-neuron gains are drawn from a Gaussian distribution of unit mean and s.d. 0.5. One trial with fixed and known initial conditions is considered.

The student networks were initialized before training with homogeneous gains, $g_i = 1$.

Figure 4. The teacher network is the same biologically inspired network as in Fig. 2.

Figure 5. *Drosophila* larva (top row). The teacher network was trained, following Zarin et al.[16], such that the motor neurons produce activation sequences consistent with forward and backward crawling, with an additional L2-regularization loss on the activity of premotor units. The synaptic weights were fixed based on existing synapses, using neurotransmitter identity when available, and normalizing based on the percent input received by the postsynaptic target, as in Zarin et al.[16]. There is a different tonic input for each motion type. The trained parameters are the biases, gains and the tonic input patterns for the two motion directions. The single-neuron time constant in premotor and motor neurons is 0.2, and the activation function is Softplus. Gains were bounded during training to be between 0.5 and 5.

The student network focused on the 178 premotor neurons, during both forward and backward motion. Gains and biases were trained starting from a random permutation of the teacher's parameters. We added a small L2-regularization penalty to the loss function to reduce instabilities in the solutions.

Drosophila adult (central complex). The orientation selectivity of EPG neurons is assigned based on Turner-Evans et al.[40], where each EPG cell type is maximally selectively 22.5° from their neighboring EPGs, tiling the whole range of angular directions. The teacher was trained such that the EPG neurons are able to produce a bump of activity for 2.5 time units, 2 time units after a brief pulse (0.3 time units) is presented. The activities of PEN1s, PEN2s, PEGs and Δ7s were not constrained. The nonlinearity was Softplus, with a parameter of $\beta = 5$. The signs of the synaptic weights, based on the hemibrain connectome, were given by the predicted neurotransmitter, and the strength of each connection was set proportional to the number of synapses. The overall scaling of the synaptic weight matrix was set such that the largest eigenvalue has a real part of 0.8, and then gains and biases were trained.

To build the student network, the gains and biases of each cell type were shuffled within their own cell group. The gains were further scaled down by a factor 0.8. The gains were forced to be non-negative. We used 60 different trials, with randomly set orientations.

Larval zebrafish (hindbrain). The teacher network is a linear network based on Vishwanathan et al.[41]. The network was not trained; the overall strength of recurrent connections was fixed such that the real part of the largest eigenvalue is less but close to zero. The input pattern was randomly set but made sure to overlap with the subspace of the slowest activity mode. We then randomly assigned a set of gains from a log-normal distribution with mean 1 and s.d. 0.3 to each neuron. To keep the same dynamics, we rescaled each column of the synaptic weight matrix by the inverse of the corresponding gain.

Figure 6. For the connectivity of the teacher in Fig. 6b–h, we drew each synaptic strength J_{ij} from a Gaussian distribution with mean 0 and s.d. $1.4/\sqrt{N}$ and, based on the singular value decomposition, kept the first 60 rank-one components. The network size was $N = 300$.

In Fig. 6f (inset), to calculate the principal angle between the first M singular vectors of the subsampled matrix $A_{1:M,:}$ and the full matrix A, we calculated the M left singular vectors of the matrix $A_{1:M,:}$, $\mathbf{v'_m}$, and the first M left singular vectors of the matrix A, $\mathbf{v_m}$. The principal angle measures the maximum angle between two linear subspaces. We computed the principal angle as the maximum singular value of the matrix product $\mathbf{v'_m}^T\mathbf{v'_n}$, for $m, n = 1 \ldots M$.

Figure 7. We designed a teacher RNN with a rank-two synaptic weight matrix and two populations[36], tanh activation function and gains as single-neuron parameters; the biases are set to 0, $N = 1{,}000$ neurons. The parameters in the first population are $\sigma^{(1)}_{n_1m_1} = 1.89$, $\sigma^{(1)}_{n_1m_2} = 0.25$, $\sigma^{(1)}_{n_2m_1} = 0.10$, and $\sigma^{(2)}_{n_2m_2} = 0.11$ and in the second population $\sigma^{(1)}_{n_1m_1} = -0.11$, $\sigma^{(1)}_{n_1m_2} = 0.22$, $\sigma^{(1)}_{n_2m_1} = -0.02$, and $\sigma^{(2)}_{n_2m_2} = 2.26$. The gains are reduced to a two-dimensional parameter space, where all the gains of neurons in population 1 have the same value, g_1, and all the gains of neurons in population 2 have value, g_2. In the teacher network, $g_1^* = 1.2$ and $g_1^* = 1.5$. The parameters are chosen such that the first population has more control of the dynamics along the variable κ_1, and the second population controls κ_2.

Figure 8. The linear network corresponds to the same network as in Fig. 6. The nonlinear network is the same network as in Fig. 2.

Statistics and reproducibility

For each teacher network, we trained at least 10 student networks to obtain robust estimates of the prediction accuracy. All trained student

networks were included in the analysis. Student networks whose activity diverged during training were retrained with a different random seed until convergence. No other data were excluded from the analyses.

Prediction in linear recurrent networks

In the linear model, the RNNs are linear networks with dynamics

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij}\left(x_j + b_j\right). \tag{12}$$

We define the activity in this linear network as $r_i(t) = x_i(t)$. We assume that the real part of all eigenvalues of the connectivity matrix J are smaller than unity, so that the linear dynamics are stable. The single-neuron parameters b_i correspond to the bias. Throughout the results section, we focused on the fixed point activity, which is given in vector form by

$$\mathbf{r} = (I - J)^{\dagger} J\, \mathbf{b}, \tag{13}$$

where I is the identity matrix. There is a linear mapping between single-neuron parameters $\mathbf{b}$ and activity $\mathbf{r}$, given by a matrix A, in this case defined as $A = (I - J)^{\dagger} J$. The notation $A^{\dagger}$ indicates the pseudo-inverse.

In linear networks, such as the simplified model studied here, it is possible to formulate the teacher–student setup as a system identification problem and estimate the parameters using alternative approaches to gradient-based training, such as subspace methods[64]. For consistency with our approach to nonlinear networks, we use gradient descent to optimize the parameters of the student. However, the same conclusions can be reached from a system identification perspective (Supplementary Note).

Fully sampled teacher. The loss function when all neurons are recorded is given by the quadratic form

$$\mathcal{L} = \left(\boldsymbol{b} - \boldsymbol{b}^*\right)^T A^T A \left(\boldsymbol{b} - \boldsymbol{b}^*\right), \tag{14}$$

such that there is one global minimum when the student and teacher are identical to each other, $\boldsymbol{b} = \boldsymbol{b}^*$, and the Hessian of the loss is independent of the teacher's biases $\boldsymbol{b}^*$. Running gradient descent, in the limit of small learning rates η, leads to equation (4) for the estimated biases in the student over the timecourse t' of learning, which reads in vector form:

$$\frac{d\boldsymbol{b}}{dt'} = -\eta \nabla \mathcal{L} = -\eta A^T A \left(\boldsymbol{b} - \boldsymbol{b}^*\right). \tag{15}$$

Equation (15) shows that the evolution of single-neuron parameters $\boldsymbol{b}$ is given by a linear dynamical system. The eigenvalue decomposition of $A^T A$, or, equivalently, the singular vector decomposition of A, therefore determines how fast and along which modes the parameters $\boldsymbol{b}$ decay toward the teacher values, $\boldsymbol{b}^*$. Given the singular vector decomposition $A = \sum_{k=1}^{N} s_k \mathbf{u}_k \mathbf{v}_k^T$, we denote the left singular vector $\mathbf{u}_k$ an activity mode and the right singular vector $\mathbf{v}_k$ a parameter mode. The error in parameter mode $\mathbf{v}_k$ decreases over training with timescale $\eta^{-1} s_k^{-2}$, reducing the error in activity mode $\mathbf{u}_k$. An initial guess $\mathbf{b}_0$, which is a distance of 1 away from the teacher $\mathbf{b}^*$ along mode $\mathbf{v}_k$, generates an error in neural activity along mode $\mathbf{u}_k$ of magnitude s_k. Thus, parameter modes that have large effects on activity are learned quickly, whereas parameter modes that have small effects on activity are learned more slowly. We refer to parameter modes corresponding to large and small singular values as 'stiff' and 'sloppy' modes, respectively.

If the connectivity J is not full rank, some singular values of the mapping matrix A will be zero. In that case, the parameter values along the modes $\mathbf{v}_k$ corresponding to singular value $s_k = 0$ (the extreme case of sloppy parameter modes) cannot be inferred through gradient descent, although that mismatch does not cause any error in the activity of unrecorded neurons.

All the results can be directly extended to linear networks where transient trajectories are considered, given an initial state $\mathbf{x_0}$. For time-dependent responses, the dynamics follow

$$\mathbf{x}(t) = A(t)\mathbf{b} + \exp\left((-I + J)\, t\right)\mathbf{x_0}, \tag{16}$$

where there is an affine mapping from $\mathbf{x}(t)$ to the parameters $\mathbf{b}$, given by $A(t)$:

$$A(t) = \left(I - \exp\left((-I + J)\, t\right)\right)(I - J)^{\dagger} J. \tag{17}$$

The second term in equation (16) is the same for the teacher and the student because we assume that the initial state is known. Thus, the difference in activity between networks is:

$$\mathbf{x}(t) - \mathbf{x}^*(t) = A(t)\left(\mathbf{b} - \mathbf{b}^*\right). \tag{18}$$

The loss function is the time-averaged squared error of the activity:

$$\mathcal{L} = \frac{1}{T}\int dt (\mathbf{x}(t) - \mathbf{x}^*(t))^T (\mathbf{x}(t) - \mathbf{x}^*(t)) = \left(\boldsymbol{b} - \boldsymbol{b}^*\right)^T \left[A^T(t) A(t)\right] \left(\boldsymbol{b} - \boldsymbol{b}^*\right), \tag{19}$$

where the square brackets indicate a time average. The matrix that determines the stiff and sloppy modes is, therefore, the time-averaged matrix $\left[A(t)^T A(t)\right]$. The eigenvalues of this matrix determine the level of stiffness, and the eigenvectors determine the parameter modes.

Note that, in this model, we have assumed that all the recurrent dimensions are explored by the teacher, such that the rank of the connectivity determines the dimensionality of the activity. In practice, neural activity is recorded for a limited time window in response to a small set of inputs, so the dimensionality of the activity is much lower. The rank of the connectivity sets an upper bound on the network's activity (see Extended Data Fig. 4 for a comparison of the dimensionality of activity and rank of the connectivity in connectome-constrained recurrent networks).

Subsampled activity. Recording from a subset of M neurons is equivalent to selecting the rows of matrix A corresponding to the recorded neurons and removing the rest. We refer to this matrix as matrix $[A]_{1:M,:}$. Equations (14) and (15) still hold, when substituting $[A]_{1:M,:}$ for A.

The effect of subsampling limits the number of learnable or stiff parameter modes of the loss function used for training, which cannot exceed M. The fact that the initial parameter guess $\mathbf{b}_0$ can be corrected only along M modes makes the error in unrecorded activity non-zero when the rank of A is larger than M—that is, when not enough neurons are sampled. Furthermore, the parameter modes and activity modes without a non-zero eigenvalue of the training loss need not align with the stiffest modes of the fully sampled loss function.

One recorded neuron. In linear networks, we can calculate the average error when we record only from neuron i. We use the vector $\mathbf{a}_i$ to refer to the row of the subsampled matrix $A_{i,:}$. After training has converged for a student with initial parameters $\mathbf{b}_0$, which is equivalent to assuming zero training error for the recorded activity of neuron i given the absence of measurement noise, the vector of biases after training is

$$\mathbf{b}_f - \mathbf{b}^* = \left(I - \frac{\mathbf{a}_i \mathbf{a}_i^T}{\mathbf{a}_i^T \mathbf{a}_i}\right)\left(\mathbf{b}_0 - \mathbf{b}^*\right). \tag{20}$$

The squared error in single-neuron parameters (combining both recorded and unrecorded neurons), e_f, is calculated based on the norm of the vector $\mathbf{b}_f - \mathbf{b}_0$ given by equation (20), which reads:

$$e_f^2 = e_0^2 - \left(\mathbf{b}_0 - \mathbf{b}^*\right)^T \frac{\mathbf{a}_i \mathbf{a}_i^T}{\mathbf{a}_i^T \mathbf{a}_i} \left(\mathbf{b}_0 - \mathbf{b}^*\right)^T. \tag{21}$$

Assuming that the initial guesses $\mathbf{b}_0$ are unbiased with respect to the teacher parameters $\mathbf{b}^*$, on average over initial conditions, the improvement in the error in parameters is

$$\left\langle \frac{e_f^2}{e_0^2} \right\rangle = 1 - \frac{1}{N}. \tag{22}$$

Therefore, on average, the error in parameter space is equally reduced for any selected neuron.

Similarly, the error in the activity of neurons reads:

$$\mathbf{x}_f - \mathbf{x}^* = A\left(\mathbf{b}_f - \mathbf{b}^*\right) \tag{23}$$

such that the squared error, using the singular value decomposition of A, is

$$E_f^2 = \sum_{k=1}^{N} s_k^2 \left(\mathbf{v}_k^T \left(\mathbf{b}_0 - \mathbf{b}^*\right)\right)^2. \tag{24}$$

The squared error E_f^2 can be larger or smaller than the error before training, unlike the error in parameter space, which can only decrease. Nevertheless, on average over initial conditions, the expected error always decreases and is given by

$$1 - \left\langle \frac{E_f^2}{E_0^2} \right\rangle = \frac{1}{\sum_k s_k^2} \left(\sum_{k=1}^{N} s_k^2 \cos^2 \theta_k \right), \tag{25}$$

where $\cos\theta_k$ is the angle between $\mathbf{a}_i$ and $\mathbf{v}_k$. Equation (25) is used to calculate the theoretical predictions in Fig. 8e.

Error in unrecorded neural activity versus single-neuron parameters. From the perspective of a single neuron i, we can write the following identity using the equation for the linear network dynamics at the fixed point:

$$(1 - J_{ii})\left(x_i^f - x_i^*\right) = \sum_{j \neq i} J_{ij}\left(x_j^f - x_j^*\right) + \sum_j J_{ij}\left(b_j^f - b_j^*\right). \tag{26}$$

The first term corresponds to the error due to incorrect prediction of the activity of other neurons in the network, whereas the second term corresponds to the error due to parameter mismatch between teacher and student.

If neuron i is a recorded neuron, then, after training has converged, equation (26) equals 0, imposing the constraint

$$\sum_{j \neq i} J_{ij}\left(x_j^f - x_j^*\right) = -\sum_j J_{ij}\left(b_j^f - b_j^*\right). \tag{27}$$

In other words, the weighted sum of errors from incorrectly inferring parameters (r.h.s.) compensates for the weighted sum of errors from the incorrect prediction of activity (l.h.s.).

If neuron i is not a recorded neuron, then both terms in equation (26) in general contribute to the squared error. Which term has a stronger contribution depends on the strength of recurrent connectivity. For strong recurrence, the first term will dominate, whereas, for weakly connected networks, the second term will dominate. As more neurons are recorded, the amplitudes of both contributions decay similarly.

Optimal selection of neurons: linear RNN. To calculate the best and worst strategy for sampling neurons (Fig. 8), we used a greedy strategy, where we first selected the neuron with the highest and lowest expected reduction in activity error, based on equation (25). Then, we proceeded iteratively, projecting out the component $\mathbf{a}_i^{(l)}$ from the (l)-th selected neuron from the matrix $A^{(t)}$, calculating the matrix $A^{(l+1)}$:

$$A^{(l+1)} = \left(I - \frac{\mathbf{a}_i^{(l)} \left(\mathbf{a}_i^{(l)}\right)^T}{\left(\mathbf{a}_i^{(l)}\right)^T \mathbf{a}_i^{(l)}} \right) A^{(l)}. \tag{28}$$

We then selected again the row-vector $\mathbf{a}_i^{(l+1)}$ that maximizes (minimizes) the decrease error in equation (25), for the best (worst) greedy selection of neurons.

Optimal selection of neurons: nonlinear RNN. For any teacher RNN with unknown gains or nonlinear activation functions, the mapping between unknown single-neuron parameters and activity is not given by a linear transformation via a matrix A. Moreover, the linearization of the gradient dynamics (equation (15)) close to the teacher parameter depends on the specific parameters, unlike the linear case. Nevertheless, we can still compute the best and worst selection of neurons based on an initial guess of the target parameters.

We focus on networks with firing rates given by $\mathbf{r} = \hat{\mathbf{g}}\phi\left(\mathbf{x} + \mathbf{b}\right)$, where the notation $\hat{\mathbf{x}}$ indicates a diagonal matrix whose non-zero elements are given by vector $\mathbf{x}$, and we assume the function ϕ is invertible. We are interested in the linearization $\Delta\mathbf{r}/\Delta\mathbf{b}$ and $\Delta\mathbf{r}/\Delta\mathbf{g}$. We are focused on fixed point activity, and, thus, using equation (6), we can define the function:

$$F\left(\mathbf{r}, \mathbf{b}, \mathbf{g}\right) = -\phi^{-1}\left(\hat{\mathbf{g}}^{-1}\mathbf{r}\right) + \mathbf{b} + J\mathbf{r} = 0. \tag{29}$$

By applying the implicit function theorem to F, we can calculate the linearized mapping from parameters to activity:

$$\Delta\mathbf{r} = -\left(\frac{dF}{d\mathbf{r}}\right)^{-1}\left(\frac{dF}{d\mathbf{b}}\Delta\mathbf{b} + \frac{dF}{d\mathbf{g}}\Delta\mathbf{g}\right) \tag{30}$$

$$\Delta\mathbf{r} = \left(\left(\hat{\boldsymbol{\phi}}^{-1}\right)'\hat{\mathbf{g}}^{-1} - J\right)^{-1}\left(\left(\hat{\boldsymbol{\phi}}^{-1}\right)'\frac{\hat{\mathbf{r}}_0}{\hat{\mathbf{g}}^2}\Delta\boldsymbol{g} + \Delta\boldsymbol{b}\right). \tag{31}$$

This linear relationship is analogous to the parameter-to-activity mapping A defined previously for linear RNNs, allowing us to use the same procedure iteratively. This amounts to assuming that the curvature of the loss function close to the current parameter estimate is similar to the curvature close to the teacher parameters.

In Fig. 7g,h, the Jacobian of the mapping between time-varying activity and single-neuron parameters around the teacher's parameter values was estimated numerically, using PyTorch's automatic differentiation.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The connectomics data used in this study were published in Zarin et al.[16] for *Drosophila* larva, in Scheffer et al.[9] for the central complex of adult *Drosophila* and in Vishwanathan et al.[41] for the brainstem of the larval zebrafish. All generated data shown in the main results, together with the teacher and student recurrent networks, are publicly available at https://doi.org/10.5281/zenodo.16618353 (ref. 65).

Code availability

All simulations and analyses were performed using custom code written in Python (https://www.python.org). The code used to generate all the results and can be found in ref. 65 and https://github.com/emebeiran/connconstr.

References

61. Werbos, P. J. Backpropagation through time: what it does and how to do it. *Proc. IEEE* **78**, 1550–1560 (1990).

62. Kingma, D. P. & Ba, J. L. Adam: a method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations* (ICLR, 2015).
63. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**, 8026–8037 (2019).
64. Van Overschee, D. P. & De Moor, B. *Subspace Identification for Linear Systems: Theory–Implementation–Applications* (Springer Science and Business Media, 2012).
65. Beiran, M. & Litwin-Kumar, A. Dataset and code for generating the figures of publication Beiran, M., Litwin-Kumar A., Prediction of neural activity in connectome-constrained recurrent networks. *Zenodo* https://doi.org/10.5281/zenodo.16618353 (2025).

Acknowledgements

We are grateful to L. F. Abbott for helpful discussions and comments on the paper. M.B. and A.L.-K. were supported by the Kavli Foundation, the Gatsby Charitable Foundation (GAT3708), the Burroughs Wellcome Foundation and National Institutes of Health awards R01EB029858 and RF1DA060772. A.L.-K. was supported by the McKnight Endowment Fund. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

Author contributions

M.B. and A.L.-K. conceived the study. M.B. performed simulations and analyses, with contributions from A.L.-K. M.B. and A.L.-K. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

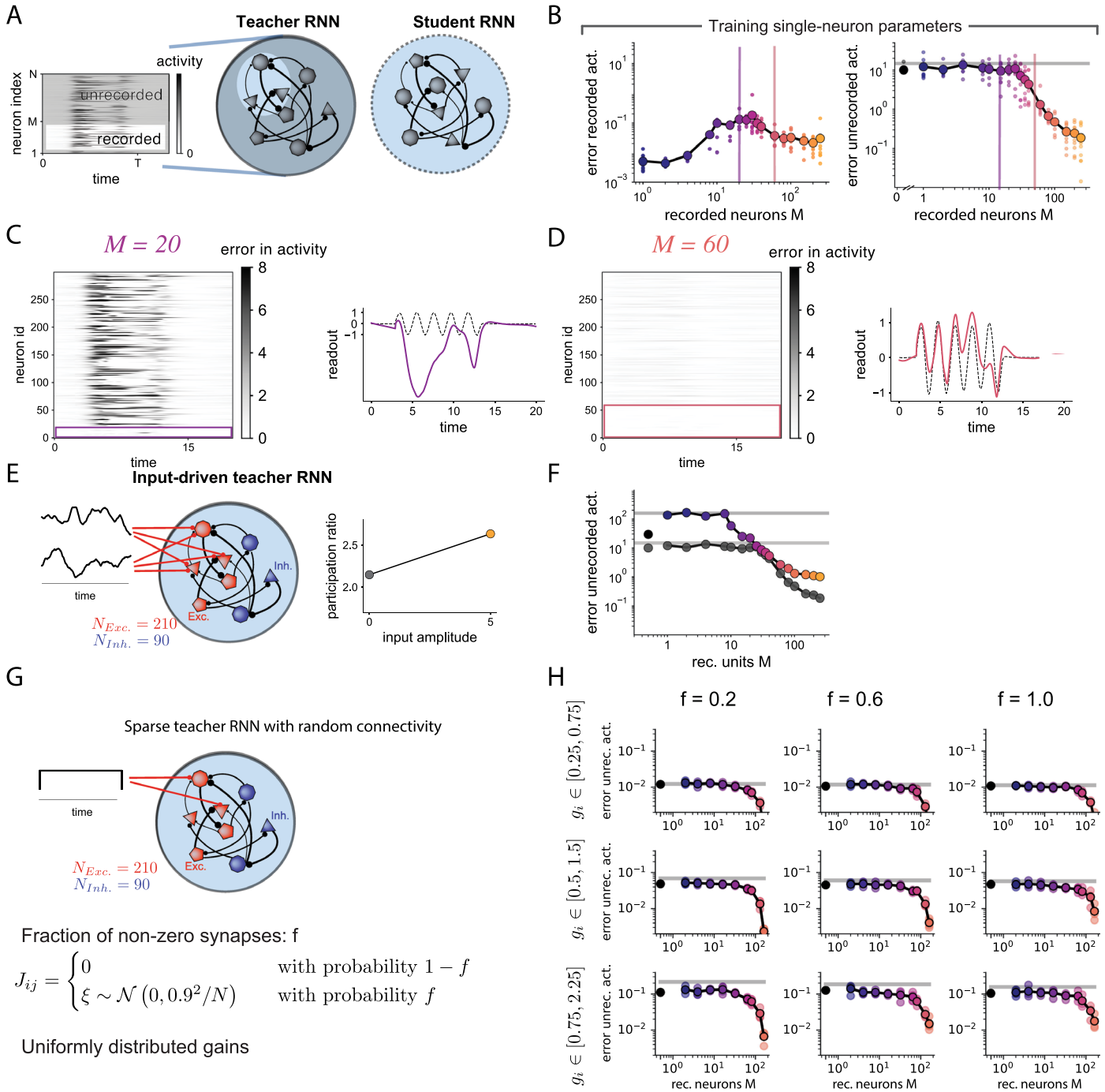
Extended data is available for this paper at https://doi.org/10.1038/s41593-025-02080-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-025-02080-4.

Correspondence and requests for materials should be addressed to Manuel Beiran or Ashok Litwin-Kumar.

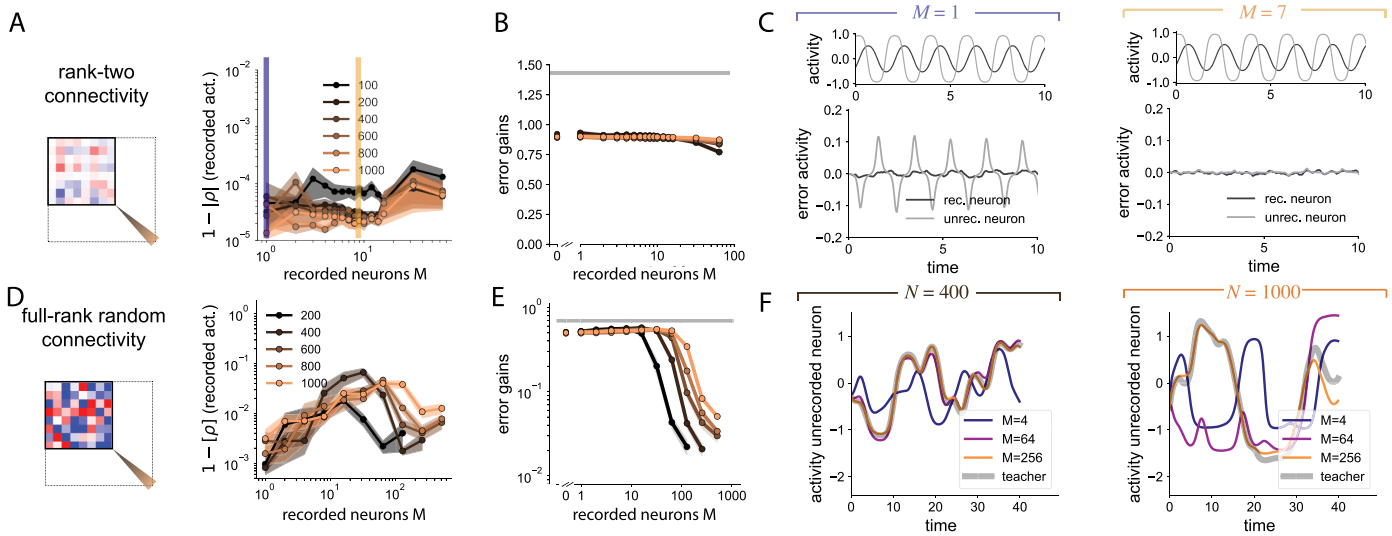
Peer review information *Nature Neuroscience* thanks Jakob Macke and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41593-025-02080-4



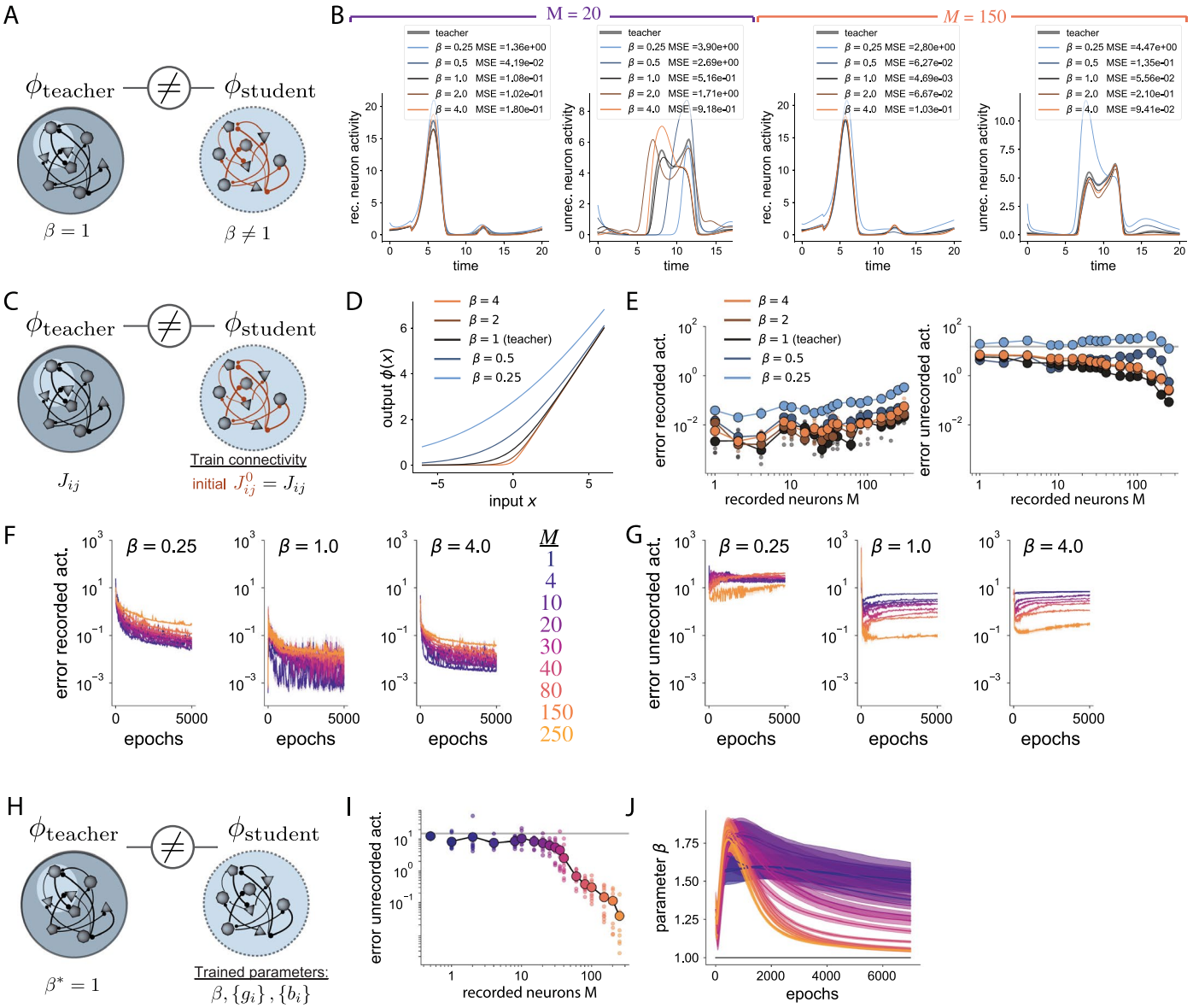
Extended Data Fig. 1 | Related to Fig. 2. A Teacher as in Fig. 2. The students are trained on a varying number of recorded neurons *M*. **B** Average error in the recorded and unrecorded activity between teacher and students. **C** Left: Error in the network activity for a given student network in a given trial, when $M = 20$ neurons are recorded. Right: Error in the task-related readout signal. While the recorded neurons have low error, the unrecorded neurons in the student display large deviations. **D** Analogous to **C**, when more neurons are recorded, $M = 60$. In this case, the activity of unrecorded neurons and the readout are well predicted. **E** Teacher network from panel **A** receives a strong external two-dimensional time-varying input, fed to a subset of 100 excitatory neurons. Middle: The dimensionality of the activity, measured by the participation ratio, increases with the input. **F** Error in unrecorded neuronal activity after training student networks to match the input-driven teacher (color dots), compared to the non-driven teacher (grey dots). Fewer recorded neurons are required to predict activity of unrecorded neurons in this example input-driven network. **G** Input-driven teacher network with different levels of connectivity sparsity and gain heterogeneity. Teachers have E-I random connectivity, and are initialized at the fixed point. A positive input of unit strength is delivered to 5 excitatory neurons. Recorded neurons correspond to excitatory neurons, while unrecorded neurons can be both excitatory or inhibitory. Teacher networks are generated with different fractions *f* of non-zero weights, and different ranges for the uniformly distributed gains. Both gains and biases are trained in the students. **H** Error in unrecorded activity after training vs number of recorded neurons, for different level of sparsity *f* and gain distributions. While the overall magnitude of the error changes for different gain strengths, the decay of the error as a function of *M* does not change.

https://doi.org/10.1038/s41593-025-02080-4



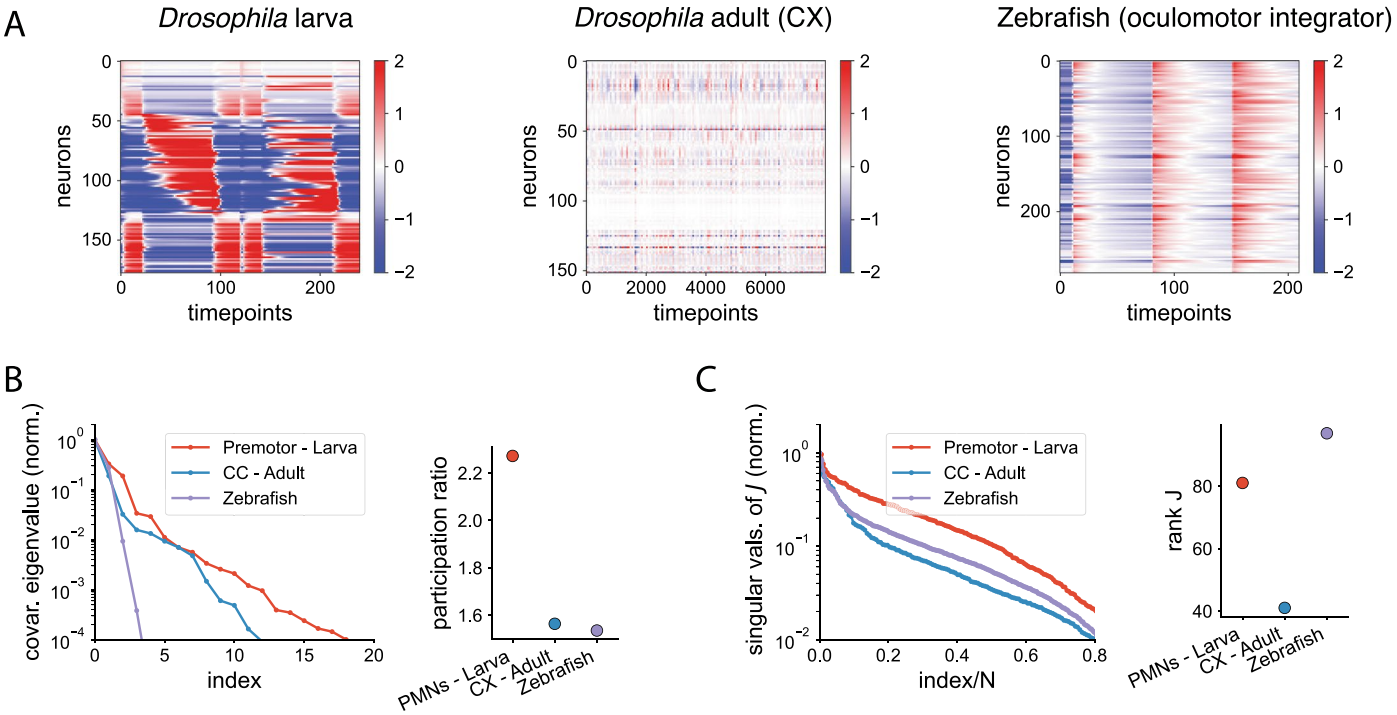
Extended Data Fig. 2 | Teacher networks with different dynamics, related to Fig. 3. A Teachers with variable network size and fixed rank-two connectivity, generating a limit cycle. Right: Error in the activity of recorded neurons after training. The students always learn the dynamics of the teacher. **B** Error in the single-neuron gains after training. **C** Example of error in the activity of a recorded neuron and an unrecorded neuron, when there is only one recorded neuron (left), compared to when 7 neurons are recorded (right). For one recorded neuron, the student learns the frequency of the limit cycle, but the temporal profile of the unrecorded neurons does not much the profile of the teacher network. Example for $N = 400$. **D** Teachers with variable network size and random connectivity, generating chaotic dynamics. Right: Error in the activity of recorded neurons after training. The students always learn the dynamics of the teacher. **E** Error in the single-neuron gains for the chaotic teachers. Note that the single-neuron parameters are much better inferred given enough recorded neurons when the teacher is chaotic than when it is low-rank, because there are many more stiff dimensions. **F** Traces of one example neuron in teacher and student networks with size $N = 400$ (left) and $N = 1000$ (right). For $N = 400$, $M = 64$ recorded neurons is sufficient to accurately match unrecorded neural activity from the teacher (gray line), while for $N = 1000$, $M = 64$ recorded neurons is insufficient but $M = 256$ is sufficient.

https://doi.org/10.1038/s41593-025-02080-4



Extended Data Fig. 3 | Training connectivity with model mismatch, related to Fig. 4. A Teacher with model mismatch in the activation function, from Fig. 4a-c. **B** Example traces of one recorded neuron and one unrecorded neuron in the teacher and after training the student with mismatch in the β parameter. The students networks were trained with 20 recorded neurons (left) and with 150 recorded neurons (right). **C** Teacher-student framework with mismatch. We train the connectivity of the student, given the teacher's connectivity as initial condition. The single-neuron parameters are the same in teacher and student, while there is a mismatch in the activation function. Same network as in Fig. 4. **D** The activation function is a smooth rectification but with different degrees of smoothness, parameterized by a parameter β. Teacher RNN from Fig. 2. **E** Errors in the activity of recorded (left) and unrecorded (right) neurons for different values of model mismatch between teacher and student. We observe a minor decrease in the error in unrecorded neurons when recording from a large number of neurons, $M \approx 150$. **F** Error in the recorded activity (loss function) for three different mismatch values as a function of training epochs ($\beta = 1$. means no mismatch). **G** Error in the unrecorded activity (loss function) for three different mismatch values as a function of training epochs. **H** Removing the mismatch in activation by training an additional parameter. We train a student network with the same connectivity as the teacher and different single-neuron parameters. However, the student also does not know the smoothness parameter β. The trained parameters are therefore the gains and biases of each neuron and the smoothness β. **I** Error in unrecorded activity after training on a subset of M recorded units, similar to C. Training the smoothness parameter of the nonlinearity provides the student with the same prediction power as students without mismatch (see Fig. 2). **J** Estimated parameter β during training (average and SEM over 10 different initializations). Networks do not retrieve the exact teacher value ($\beta^* = 1$) although converge to values not far from it on average. Students have a bias towards estimating sharper activation functions ($\beta > 1$). Both bias and variance are reduced as the number of recorded neurons is increased.

https://doi.org/10.1038/s41593-025-02080-4



Extended Data Fig. 4 | Dimensionality of the activity and rank of connectivity in the data-constrained RNNs, related to Fig. 5. A Neural activity traces (centered) used for training the student networks for the three different data constrained RNNs: the premotor network in the *Drosophila* larva, the central complex in the adult *Drosophila*, and the oculomotor integrator in larval zebrafish. Different trials/conditions have been concatenated. **B** Left: First eigenvalues of the covariance spectrum of the datasets. Right: Participation ratio of the activity covariance. The dimensionality of neural activity is higher in the premotor system, then the CX and then the premotor network, indicated by how fast the eigenvalues decay. **C** Left: Singular values of the connectivity matrix. Right: Estimated rank of the connectivity matrix *J*, calculated using the participation ratio of the distribution of singular values of *J*. Given the sparsity and heterogeneity in connectomes, the rank of the connectivity is high.

nature portfolio

Corresponding author(s): Manuel Beiran

Last updated by author(s): Jul 31, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No commercial software was used. We used Python3 and PyTorch 1.13 for the numerical experiments. Code is shared in a community repository: https://doi.org/10.5281/zenodo.16618353
Data analysis	No commercial software was used. We used Python3 and PyTorch 1.13 for the numerical experiments. Code is shared in a community repository: https://doi.org/10.5281/zenodo.16618353

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Connectomics data was used from larval Drosophila (Zarin et al. 2019), from adult Drosophila -the hemibrain dataset (Scheffer et al. 2020), and from the hindbrain

of larval zebrafish (Vishwanathan et al. 2024). The preprocessed datasets and the simulated data shown in the results are available in a public repository: https://doi.org/10.5281/zenodo.16618353

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We systematically studied multiple student networks (at least 10) linked to the same teacher network to ensure that all results are consistent across random initializations.
Data exclusions	No data were excluded from the analysis, except for the few student networks whose activity became unstable over training.
Replication	Code and data are available to replicate the findings of the study
Randomization	No random allocation of samples was relevant to this study on recurrent neural networks. We performed statistical controls involving random shuffling neuronal identities to estimate how much better than chance are neural predictions.
Blinding	Blinding group allocation is not relevant for this study on recurrent neural networks.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*

Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used — *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation — *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s) — *State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.*

Authentication — *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination — *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See ICLAC register) — *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance — *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*

Specimen deposition — *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods — *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight — *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals — *For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*

Wild animals — *Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Reporting on sex — *Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall*

	numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks — *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes — *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication — *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.
For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission — *Provide a list of all files available in the database submission.*

Genome browser session
(e.g. UCSC)
Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates — *Describe the experimental replicates, specifying number, type and replicate agreement.*

Sequencing depth — *Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*

Antibodies — *Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Peak calling parameters — *Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*

Data quality — *Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*

Software — *Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation — *Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*

Instrument — *Identify the instrument used for data collection, specifying make and model number.*

Software — *Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Indicate task or resting state; event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

(See Eklund et al. 2016)

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*
Multivariate modeling and predictive analysis	*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*