COMMENT

Measurement, manipulation and modeling of brain-wide neural population dynamics

Krishna V. Shenoy [1,2,3,4,5,6] & Jonathan C. Kao[7,8]

Neural recording technologies increasingly enable simultaneous measurement of neural activity from multiple brain areas. To gain insight into distributed neural computations, a commensurate advance in experimental and analytical methods is necessary. We discuss two opportunities towards this end: the manipulation and modeling of neural population dynamics.

Neural circuits comprise networks of individual neurons that perform sensory, cognitive, and motor functions. Neuronal biophysics, together with these circuits, give rise to neural population dynamics, which express how the activity of the neural population evolves through time in principled ways. Neural population dynamics provide a framework for understanding neural computation. Prior studies have modeled neural population dynamics to gain insight into computations involved in decision-making, timing, and motor control[1]. Here, we present emerging opportunities for new experiments and analyses that use a dynamical systems framework to better understand brain circuits, how they interact, and how they relate to behavior.

The simplest model of neural population dynamics is a linear dynamical system (LDS). An LDS (Fig. 1a) is described by a *dynamics* equation ($\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$) and an *observation* equation ($\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{d}$). Typically, $\mathbf{y}(t)$ reflects experimental measurements, such as a vector where each element is the number of action potentials fired by a neuron in a brief time bin (e.g., 10 ms). The vector $\mathbf{x}(t)$ is a "neural population state" that captures information in $\mathbf{y}(t)$. This neural population state can be thought of as a representation of the dominant activity patterns in the experimental neural recordings. Typically, $\mathbf{x}(t)$ is an abstract representation in a low-dimensional subspace (or manifold) found via dimensionality reduction[2] (Fig. 1b, neural state), reflecting that the neural activity is correlated and the dominant patterns can be described by a relatively small number of variables. The neural population state can also represent the activity of each neuron in the original dimensionality of the measured data (e.g., 100D if 100 neurons). The observation equation relates the observed action potentials ($\mathbf{y}(t)$) to the neural population state ($\mathbf{x}(t)$) through an observation matrix ($\mathbf{C}$). The vector, $\mathbf{d}$, is a constant offset (e.g., to model baseline firing). The neural population state moves through neural state space, constituting a neural population trajectory. The dynamics equation expresses how the neural population state ($\mathbf{x}(t)$) progresses through time as a function of a dynamics matrix ($\mathbf{A}$), an input matrix ($\mathbf{B}$) and

[1]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. [2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [3]Department of Neurobiology, School of Medicine, Stanford University, Stanford, CA 94305, USA. [4]Wu Tsai Neuroscience Institutes, Stanford University, Stanford, CA 94305, USA. [5]Bio-X Institute, Stanford University, Stanford, CA 94305, USA. [6]Howard Hughes Medical Institute (HHMI) at Stanford University, Stanford, CA 94305, USA. [7]Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA, USA. [8]Neurosciences Program, University of California, Los Angeles, Los Angeles, CA, USA. ✉email: shenoy@stanford.edu
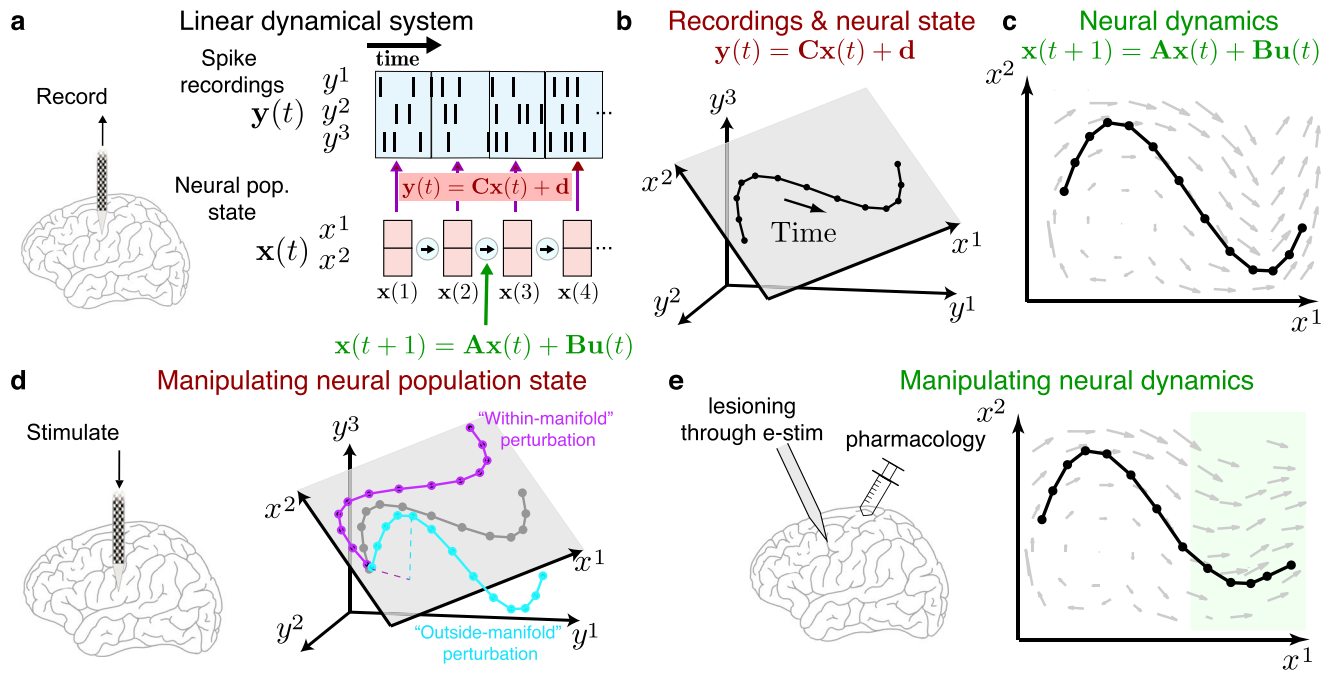
Fig. 1 Overview of neural dynamics and manipulations. a Linear dynamical system using neural recordings. Binned spiking activity, **y**(t), relates to a latent neural population state, **x**(t), that evolves according to linear dynamics, **A**, with inputs from other cortical areas, **Bu**(t). The neural population state is typically a low-dimensional trajectory, and the dynamics can be conceptualized as a flow field. **b** In many cases, the neural state can be represented as a low-dimensional trajectory in a subspace of the higher-dimensional recordings. **c** In this subspace, the neural state evolves according to neural dynamics, which define a flow field. In a LDS, the dynamics of **A** can be contractive, expansive, rotational, or a fixed point. Inputs may cause the flow field to exhibit more complex motifs, such as shown in this panel. Generally, dynamics and dimensionality reduction can also be modeled to be nonlinear. **d** By electrical or optogenetic stimulation, or applying perturbations to the sensory-behavior or force-behavior relationship, it is possible to make perturbations to the neural state, **x**(t). These perturbations can be "within-manifold", which perturbs the neural state along its natural modes, or they can be "outside-manifold," which perturbs the neural state along dimensions outside the plane spanned by x^1 and x^2. **e** The neural dynamics (flow field, **A**) can also be perturbed, for example by applying pharmacology or lesioning the circuit. The neural state changes as a result of the changing dynamics (highlighted in green). Panels **a** and **b** are modified by J. C. Kao and appeared in Pandarinath and colleagues 2018[38].

inputs (**u**(t)) from other brain areas and sensory pathways (Fig. 1c, neural dynamics). The neural population state and its dynamics are informative of behavior[3,4].

LDSs and nonlinear dynamical systems models[5,6], paired with measurements from populations of individual neurons from one or more brain areas[7–14], have produced new insights into the putative computational functions being performed[1]. Here we discuss emerging opportunities to expand dynamical systems insights into brain function. We focus on (1) manipulating neural dynamics and states, enabling future experiments to causally probe neural dynamics and their roles in computation, and (2) modeling multi-area dynamics spanning multiple brain areas, which leverages brain-wide measurements enabled by new large-scale electrophysiological neural recording methods[9,10].

Manipulating neural dynamics
To date, dynamical systems studies have primarily modeled neural population recordings during behaviors. By building on work to perturb the neural circuit, future experiments may further elucidate properties of the circuit and determine causal circuit roles. Two important opportunities are to (1) casually perturb the neural activity, which is present in the neural circuitry, and (2) causally alter the neural circuit dynamics, which reflect the neural circuitry.

First, causally perturbing **x**(t) and observing how the neural circuit dynamics counteract this perturbation helps us learn more about **A**[1]. We can also gain insight by perturbing inputs from other brain areas, **u**(t)[15]. There are several ways to casually perturb neural activity, including electrical microstimulation[16] and

optogenetic stimulation[17,18]. Neural activity can also be causally perturbed through task manipulations[4], including changing visual targets during a computation[19] and/or during behavior[20], and changing the sensory-behavior relationship[21,22]. An emerging challenge is to perturb a population of neurons with spatio-temporal patterns that activate the circuit 'within-manifold.' These within-manifold perturbations (Fig. 1d) alter neural activity in a manner consistent with the circuit's natural activations[17,18,23]. Within-manifold perturbations can therefore be viewed as displacements of the neural state in the activity's low-dimensional manifold (Fig. 1d). In contrast, an 'outside-manifold' perturbation[24] would result in neural activity that the circuit would not naturally exhibit. In general, perturbations such as optogenetic or electrical stimulation that do not explicitly consider the low-dimensional manifold of the circuit are outside-manifold. Outside-manifold perturbations may be informative, for example, by revealing interesting dynamics in unexplored dimensions previously unaccounted for. We highlight that precise within-manifold perturbation at millisecond precision will likely lead to significant insights on computation through dynamics, enabling experimenters to causally test the impact of the neural state on behavior[4]. However, these types of perturbations are challenging because they generally require the ability to deliver precise excitation and inhibition to individual neurons at milli-second precision to induce a desired change in neural state. Overall, causal perturbation of **x**(t) and **u**(t), and examination of their effects on behavior, may identify dimensions that are causally linked to behavior and learning, and how neural dynamics respond to both natural and unusual perturbations in the circuit.
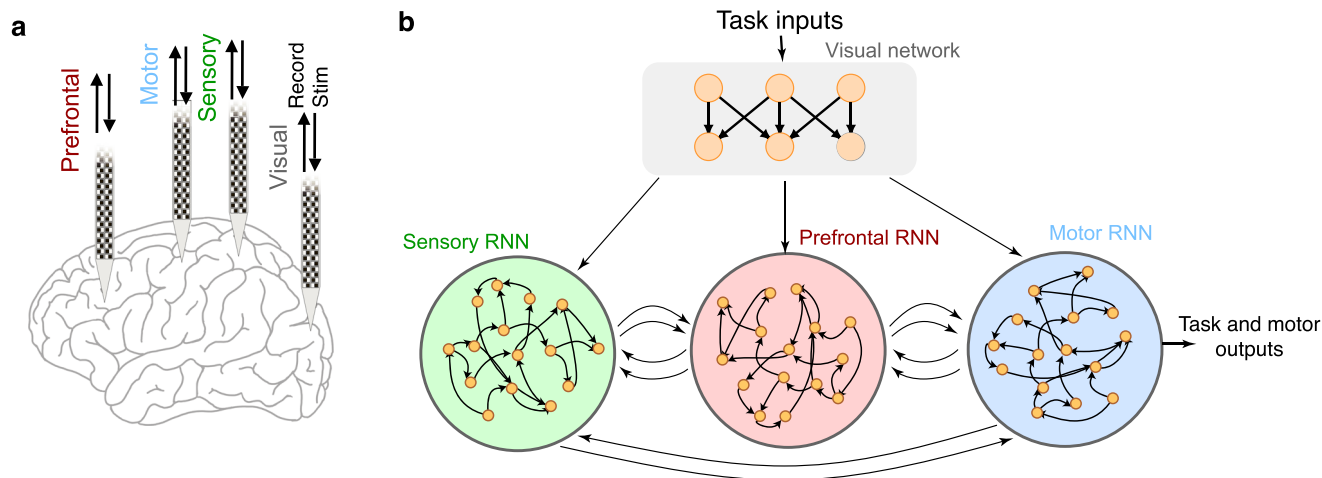
Fig. 2 Using neural networks to model multi-area computation. a Multi-area, brain-wide electrical recording and stimulation of neural activity is rapidly becoming possible, and these data require new analyses and modeling to provide new scientific insights and theories of neural computation. High-density NeuroPixel electrodes are shown inserted in four locations, and many additional insertions are possible in both cortical and subcortical regions[9]. Each insertion can access several different brain areas (M). Thus the size of the data is proportional to N × M. The proportionality constant depends on the type of neural recordings, with full broadband data including low-frequency local field potentials (LFPs) and action potentials (APs, or spikes) requiring the fastest sampling. For stimulation, it is possible to stimulate arbitrary waveforms on each electrode in each area and thus there are enormous combinatorial possibilities[39]. Single and two photon optical imaging of genetically encoded calcium indicators and voltage indicators, and optogenetic neural modulation, are also widely used but are not shown for simplicity. **b** Multi-area RNNs can be trained to model each of these insertions. Visual areas may be modeled via convolutional neural networks[40, 41] or related artificial networks that incorporate recurrence. Areas like the prefrontal and motor cortex are typically modeled by RNNs[6,30]. Regularizations may be employed so that activity within an RNN area resembles those recorded from each electrode array recordings.

A related second emerging opportunity is to alter the dynamics matrix **A** by experimentally changing the neural circuit. This can be achieved by local infusion (e.g., muscimol, chemogenetics) or systemic delivery (e.g., oral methylphenidate) of transiently acting pharmacological agents, altering local activity by delivering energy (e.g., continuous optogenetic stimulation, cooling[25], transcranial stimulation, focal ultrasound stimulation), or lesioning (which can be performed in a variety of ways). Modifying **A** may have diverse effects. For example, cooling appears to slow down trajectories within-manifold[4], while lesioning may change the manifold and its dynamics by permanently removing neurons from a circuit. Pharmacological agents may have brain-wide changes in dynamics across multiple areas, or more local changes if a local acting agent is used (e.g., muscimol). Understanding how modifications to the neural circuit's dynamics influence behavior will be important for future treatments of, and recovery from, neurological and psychiatric disorders.

Models of large scale, brain-wide neural population dynamics
Advanced neural technologies enable recording from many thousands of neurons across multiple interacting brain areas[10] (Fig. 2a). There are several modeling challenges and opportunities for dynamical systems analyses, including how to increase the modeling capacity of dynamical systems models, denoising neural data from multiple areas[16], incorporating physiological constraints into dynamical systems models, and interpreting dynamical models to generate new hypotheses for neural computation. We focus on one particular opportunity where we believe dynamical systems modeling is important: modeling distributed brain-wide computations that span multiple areas, each playing a distinct and critical role.

In cognitive and motor tasks, task-related activity arises in multiple brain areas. Evidence shows many brain areas are necessary to perform tasks at high performance, indicative that distributed computation between areas plays a critical role[26]. Multi-area recordings and analyses ought to enable new insights

into how distributed computation occurs across brain areas, addressing key questions including the following. What are the neural dynamics within each area, and how do these relate to the overall distributed computation and behavior? How are neural representations similar and different across brain areas, and what are the computational benefits of these representations? What types of information are conveyed between brain areas through inter-area connections? How are dynamics and inter-area connections coordinated? There are significant computational challenges to answering these questions. Even with multi-area recordings, information about axonal connectivity between areas may generally be unknown, requiring new computational approaches to model multi-area interactions.

Traditionally, inter-area communication between cortical areas is thought to rely on temporal coordination and communication through coherences[27]. With new multi-area datasets, recent studies viewing cortical computations as low-dimensional systems have resulted in new hypotheses for inter-area communication and multi-area dynamics. To help conceptualize multi-area dynamics, consider a didactic and oversimplified example of coupled LDSs for two areas. Here, one LDS models area 1 (subscript 1) and another LDS models area 2 (subscript 2). They are coupled through axonal projections: $\mathbf{x}_1(t+1) = \mathbf{A}_1\mathbf{x}_1(t) + \mathbf{B}_1\mathbf{u}_1(t) + \mathbf{B}_{2\text{-to-1}}\mathbf{x}_2(t)$ and $\mathbf{x}_2(t+1) = \mathbf{A}_2\mathbf{x}_2(t) + \mathbf{B}_2\mathbf{u}_2(t) + \mathbf{B}_{1\text{-to-2}}\mathbf{x}_1(t)$. $\mathbf{B}_{1\text{-to-2}}$ maps the neural state from area 1 as inputs to area 2, and vice versa for $\mathbf{B}_{2\text{-to-1}}$. Although axonal projections may not be recorded between areas, dynamical systems models using recordings from both areas 1 and 2 can provide insight into the information communicated between areas. In particular, $\mathbf{B}_{1\text{-to-2}}$ can be thought of as a communication subspace (CS) that selectively extracts features of $\mathbf{x}_1$ to propagate to $\mathbf{x}_2$, summarizing the role of inter-area axonal projections[11,12,28]. The CS may not be aligned with neural dimensions of highest variance (such as the principal components) but may instead communicate activity along low variance dimensions that are necessary for downstream computation. By conceptualizing inter-area communication as

this matrix multiplication, the CS builds on the principle of "output-null" spaces: information not necessary for downstream areas may be attenuated through alignment with the effective nullspace of the CS matrix. This phenomenon was initially observed for preparatory activity in PMd, which is attenuated in M1, likely due to its partial alignment with an output-null space[7].

There are challenges towards using dynamical system models to study multi-area computation. One challenge is to design models that couple within-area dynamics with inter-area connections. For example, what dynamical computations are performed along dimensions that are either orthogonal or read out by a CS? An important area of future research will be developing systems identification techniques to learn parameters of coupled dynamical systems from multi-area neural recordings. Second, it could be that key inputs to a brain area are not recorded. Future techniques should consider how to address these missing data if they are not recorded. One example approach is to couple dynamical systems and perturbation techniques; a recent motor learning study demonstrated that disrupting activity in one brain region can provide insight into the computations performed by another recurrently-connected brain region, despite not directly observing the activity of that second region[29]. Future dynamical models will need to not only recapitulate multi-area observations, but also respond to causal manipulations in a consistent manner. Third, it will be important to disentangle the role of feedforward and feedback connections. Finally, current approaches assume linear correlations in neural state between areas, an assumption that may need to be relaxed.

A related approach likely to be of importance for modeling multi-area distributed computations is to model nonlinear dynamical systems via neural networks. Recurrent neural networks (RNNs) have been successfully used to model the dynamics of local, single-area computations in cognitive[30,31], timing[32], navigation[33], and motor[6,34] tasks. Following optimization (e.g., with backpropagation-through-time), RNNs often have synthetic activity that resembles electrophysiological activity, although if not, regularizations or other training techniques can typically be applied to induce strong resemblance to neurophysiological recordings[6,35]. This enables the RNN to act as an in silico model of the cortical area, where the RNN can be probed to propose dynamical hypotheses for the neural computations appearing in an area[30]. It is worth noting there exists a realism gap between RNNs and neural circuits. RNNs typically model network rates instead of spikes, use deterministic weights in place of dynamic synaptic connectivity, and are occasionally unconstrained in architecture. An important area of future research is to bridge the realism gap by determining what features of neural circuit computation can and cannot be abstracted in RNNs, which involves comparisons to data and testing of RNN proposed hypotheses. Another concern may be that RNNs will converge to different solutions based on experimenter-chosen hyperparameters, like the size of the networks, the machine learning hyperparameters of training, or other features. Intriguingly, a recent study suggests that key dynamical features, including fixed point structure, are robust to hyperparameter variation[36].

One modeling opportunity to account for brain-wide computation is to expand RNN models to be multi-area, modeling within-area dynamics and inter-area communication (Fig. 2b). RNNs can be straightforwardly extended to incorporate multiple areas[31] at a resolution that enables incorporation of anatomical constraints, including E–I cell types, proportion of connectivity between areas, and Dale's law[37]. In these models, distinct RNNs can be treated as brain areas, with interactions defined by connections that implement CSs. While recent studies have similarly found that optimization can lead to RNN areas resembling brain areas[26,37], training multi-area RNNs to resemble brain areas may

be challenging and require additional training considerations, such as regularizing neural population trajectories of each RNN area to resemble cortical areas. Multi-area RNNs may propose hypotheses for how within-area dynamics perform computation and how inter-area connections selectively propagate upstream activity as inputs to downstream areas[37]. Extension of RNN modeling tools to multiple areas may therefore be an excellent candidate for generating new hypotheses for how behavior is shaped through distributed computation across multiple cortical areas.

Neural population dynamics offer a principled approach to the study of how neural circuits distributed across many brain areas orchestrate motor and cognitive function. We believe there are rich experimental and modeling opportunities to further our understanding of how multiple areas coordinate their dynamics to produce behavior.

References

1. Vyas, S., Golub, M. D., Sussillo, D. & Shenoy, K. V. Computation through neural population dynamics. *Annu. Rev. Neurosci.* **43**, 249–275 (2020).
2. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014).
3. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
4. Jazayeri, M. & Afraz, A. Navigating the neural space in search of the neural code. *Neuron* **93**, 1003–1014 (2017).
5. Hennequin, G., Vogels, T. P. & Gerstner, W. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82**, 1394–1406 (2014).
6. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* **18**, 1025–1033 (2015).
7. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* **17**, 440–448 (2014).
8. Ames, K. C. & Churchland, M. M. Motor cortex signals for each arm are mixed across hemispheres and neurons yet partitioned within the population response [Internet]. *eLife* https://doi.org/10.7554/elife.46159 (2019).
9. Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
10. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain [Internet]. *Nature* 266–273 (2019) https://doi.org/10.1038/s41586-019-1787-x
11. Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical areas interact through a communication subspace. *Neuron* **102**, 249–259.e4 (2019).
12. Semedo, J. D., Gokcen, E., Machens, C. K., Kohn, A. & Yu, B. M. Statistical methods for dissecting interactions between brain areas. *Curr. Opin. Neurobiol.* 59–69 (2020) https://doi.org/10.1016/j.conb.2020.09.009
13. Perich, M. G., Gallego, J. A. & Miller, L. E. A neural population mechanism for rapid learning. *Neuron* **100**, 964–976.e7 (2018).
14. Perich, M. G. et al. Motor cortical dynamics are shaped by multiple distinct subspaces during naturalistic behavior. Preprint at https://www.biorxiv.org/content/10.1101/2020.07.30.228767v2.abstract (2020).
15. Sauerbrei, B. A. et al. Cortical pattern generation during dexterous movement is input-driven. *Nature* **577**, 386–391 (2020).
16. Pandarinath, C. et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat Methods* **15**, 805–815 (2018).
17. Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A. & Yuste, R. Controlling visually guided behavior by holographic recalling of cortical ensembles. *Cell* **178**, 447–457.e5 (2019).
18. Marshel, J. H., et al Cortical layer-specific critical dynamics triggering perception. *Science* [Internet] **365**, (2019) https://doi.org/10.1126/science.aaw5202
19. Peixoto, D. et al. Decoding and perturbing decision states in real time. Preprint at https://doi.org/10.1101/681783 (2019); *Nature* (2021) (in press).
20. Ames, K. C., Ryu, S. I. & Shenoy, K. V. Neural dynamics of reaching following incorrect or absent motor preparation. *Neuron* **81**, 438–451 (2014).

21. Vyas, S. et al. Neural population dynamics underlying motor learning transfer. *Neuron.* **97**, 1177–1186 (2018).
22. Sun, X. et al. Skill-specific changes in cortical preparatory activity during motor learning. Preprint at https://doi.org/10.1101/2020.01.30.919894 (2020).
23. Adamantidis, A. et al. Optogenetics: 10 years after ChR2 in neurons–views from the community. *Nat. Neurosci.* **18**, 1202–1212 (2015).
24. Sadtler, P. T. et al. Neural constraints on learning. *Nature* **512**, 423–426 (2014).
25. Long, M. A. et al. Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* **89**, 1187–1193 (2016).
26. Pinto, L. et al. Task-dependent changes in the large-scale dynamics and necessity of cortical regions [Internet]. *Neuron* https://doi.org/10.1016/j.neuron.2019.08.025 (2019).
27. Kohn, A. et al. Principles of corticocortical communication: proposed schemes and design considerations [Internet]. *Trends Neurosci.* https://doi.org/10.1016/j.tins.2020.07.001 (2020).
28. Kang, B. & Druckmann, S. Approaches to inferring multi-regional interactions from simultaneous population recordings. *Curr. Opin. Neurobiol.* **65**, 108–119 (2020).
29. Vyas, S., O'Shea, D. J., Ryu, S. I. & Shenoy, K. V. Causal role of motor preparation during error-driven learning. *Neuron* **106**, 329–339 (2020).
30. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
31. Song, H. F., Yang, G. R. & Wang, X.-J. Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework. *PLoS Comput. Biol.* **12**, e1004792 (2016).
32. Remington, E. D., Narain, D., Hosseini, E. A. & Jazayeri, M. Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron* **98**, 1005–1019.e5 (2018).
33. Banino, A. et al. Vector-based navigation using grid-like representations in artificial agents. *Nature* **557**, 429–433 (2018).
34. Michaels, J. A., Dann, B. & Scherberger, H. Neural population dynamics during reaching are better explained by a dynamical system than representational tuning. *PLoS Comput. Biol.* **12**, e1005175 (2016).
35. Rajan, K., Harvey, C. D. & Tank, D. W. Recurrent network models of sequence generation and memory. *Neuron* **90**, 128–142 (2016).
36. Maheswaranathan, N., Williams, A. H., Golub, M. D., Ganguli, S. & Sussillo, D. Universality and individuality in neural dynamics across large populations of recurrent networks. In *Advances in Neural Information Processing Systems 32 (NIPS 2019)* (eds Wallach, H. et al.) 15629–15641 (Curran Associates, Inc., New York, 2019).
37. Kleinman, M., Chandrasekaran, C. & Kao J. C. Recurrent neural network models of multi-area computation underlying decision-making. Preprint at https://doi.org/10.1101/798553 (2020).
38. Pandarinath, C. et al. Latent factors and dynamics in motor cortex and their application to brain–machine interfaces. *J. Neurosci.* **38**, 9390–9401 (2018).
39. Kimmel, D. L. & Moore, T. Temporal patterning of saccadic eye movement signals. *J. Neurosci.* **27**, 7619–7630 (2007).
40. Yamins, D. L. K. et al. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl Acad. Sci. USA* **111**, 8619–8624 (2014).
41. Yamins, D. L. K. & DiCarlo J. J. Using goal-driven deep learning models to understand sensory cortex [Internet]. *Nature Neurosci.* 356–365 (2016) https://doi.org/10.1038/nn.4244

Acknowledgements

Author contributions

K.V.S. and J.C.K. wrote the manuscript.

Competing interests

K.V.S. serves on the Scientific Advisory Board of MIND-X Inc., Inscopix Inc., and Heal Inc., and is a consultant for Neuralink Corp. and CTRL-Labs division of Facebook Reality Labs. J.C.K. has no disclosures. These entities did not support this work.

Additional information

Correspondence and requests for materials should be addressed to K.V.S.

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

Reprints and permission information is available at http://www.nature.com/reprints

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