Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction

Ryan C. Williamson[1,2,3], Brent Doiron[1,4], Matthew A. Smith[1,5,6], Byron M. Yu[1,7,8]

1: Center for the Neural Basis of Cognition, Pittsburgh, PA, USA.
2: Department of Machine Learning, Carnegie Mellon University, Pittsburgh, PA, USA.
3: School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA.
4: Department of Mathematics, University of Pittsburgh, Pittsburgh, PA, USA.
5: Department of Ophthalmology, University of Pittsburgh, Pittsburgh, PA, USA.
6: Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA.
7: Department of Electrical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA.
8: Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA.

Highlights

- The interplay of neuronal recordings and network models is becoming ever stronger.

- Population activity structure provides common ground for incisive comparisons.

- Dimensionality reduction is used to identify population activity structure.

- This approach is used to study working memory, decision making, motor control, etc.

Abstract

A long-standing goal in neuroscience has been to bring together neuronal recordings and neural network modeling to understand brain function. Neuronal recordings can inform the development of network models, and network models can in turn provide predictions for subsequent experiments. Traditionally, neuronal recordings and network models have been related using single-neuron and pairwise spike train statistics. We review here recent studies that have begun to relate neuronal recordings and network models based on the multi-dimensional structure of neuronal population activity, as identified using dimensionality reduction. This approach has been used to study working memory, decision making, motor control, and more. Dimensionality reduction has provided common ground for incisive comparisons and tight interplay between neuronal recordings and network models.

Introduction

For decades, the fields of experimental neuroscience and neural network modeling proceeded largely in parallel. Whereas experimental neuroscience focused on understanding how the activities of individual neurons relate to sensory stimuli and behavior, the modeling community sought to understand theoretically how neural networks can give rise to brain function. In recent years, developments in neuronal recording technology have enabled the simultaneous recording of hundreds of neurons or more [1]. Concurrently, increases in computational power have enabled the simulation of large neural networks [2]. Together, these developments should enable experimental data to more stringently constrain network model design and network models to better predict neuronal activity for subsequent experiments [3].

When relating large-scale neuronal recordings with large-scale network models, it is not meaningful to correspond each recorded neuron to a neuron in the network model. This is because network models typically do not attempt to replicate the precise anatomical connectivity of the biological network from which the neurons are recorded. To date, comparisons between recordings and models have primarily relied on aggregate spike train statistics based on single neurons (e.g., distribution of firing rates [4], distribution of tuning preferences [5], Fano factor [6]) and pairs of neurons (e.g., spike time [7] and spike count correlations [8,9]). To go beyond single-neuron and pairwise statistics, recent studies have examined the multi-dimensional structure of neuronal population activity to uncover important insights into mechanisms underlying neuronal computation (e.g., [10–18]). This has motivated the inquiry of whether network models reproduce such population activity structure, in addition to single-neuron and pairwise statistics, raising the bar on what constitutes an agreement between a network model and neuronal recordings.

Population activity structure is typically characterized using dimensionality reduction [19], which provides a concise summary (i.e., a low-dimensional representation) of how a population of neurons covaries and how their activities unfold over time. The low-dimensional representation describes a neural process being carried out by the larger circuit from which the neurons were recorded (e.g., [20–23]). The same dimensionality reduction method can be applied to the recorded activity and to the network model activity, resulting in population activity structures that can be directly compared (Fig. 1). Importantly, the comparison of population activity structure does not require a one-to-one correspondence between each recorded neuron and a model neuron, and instead relies on correspondence at the level of the population.
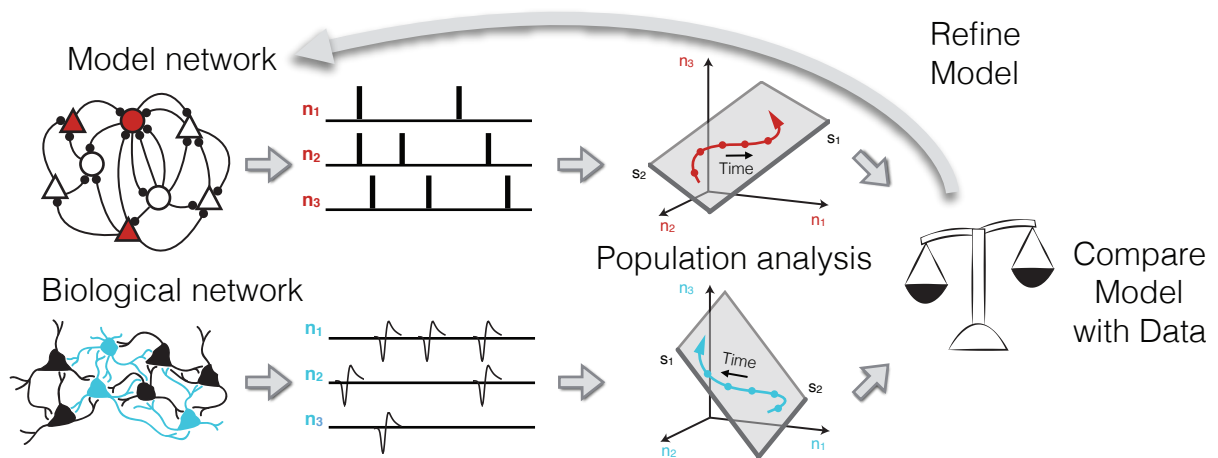
2

Fig 1. Relating biological and model networks using population analyses: Because a model network typically does not attempt to replicate the precise anatomical connectivity of a biological network, it is not meaningful to correspond each biological neuron with a model neuron. Instead, dimensionality reduction can be used to obtain a concise summary of the population activity from each network. This provides common ground for incisive comparisons between biological and model networks. Discrepancies in the population activity structure between biological and model networks can then help to refine model networks.

This approach has been adopted by recent studies to relate neuronal recordings and network models more closely than ever before to study working memory, decision making, motor control, and more. Although many studies have separately employed large-scale neuronal recordings, large-scale network models, and dimensionality reduction, this review focuses on studies that incorporate all three components. Below we describe these studies, organized by the aspect of population activity structure used to relate neuronal recordings and network models: population activity time courses, functionally-defined neuronal subspaces, and population-wide neuronal variability. These were chosen first because they represent the key ways in which dimensionality reduction has been used in the literature to relate population recordings and network models. More importantly, these three categories represent fundamental aspects of population activity structure – how it unfolds over time, how different types of information can be encoded in different subspaces, and how it varies from trial to trial.

Population activity time courses

Dynamical structures, such as point attractors, line attractors, and limit cycles, arising from network models have long been hypothesized to underlie the computational ability of biological networks of neurons [24–26]. Such dynamical structures have been implicated in decision making [27, 28], memory [29–31], oculomotor integration [32, 33], motor control [34], olfaction [35], and more. A fundamental question in systems neuroscience is whether these dynamical structures are actually used by the brain. Although single-neuron and pairwise metrics can be informative [30, 33], analyzing the activity of a population of neurons together has enabled deeper connections. In particular, the time course of the activity of a population of neurons can be summarized by low-dimensional neural trajectories [10, 20], as identified by dimensionality reduction. These neural trajectories can provide a signature of a particular

3

dynamical structure. For example, a point attractor shows convergent trajectories. A key ₆₇
advantage of this approach is that it does not require a one-to-one correspondence between each ₆₈
recorded neuron and a model neuron, but instead relies on a summary of the population ₆₉
activity time courses. ₇₀

   This approach was recently used to study how the brain flexibly controls the timing of ₇₁
behavior [17]. By applying dimensionality reduction to neuronal activity recorded from medial ₇₂
frontal cortex, Wang et al. found that population activity time courses for different time ₇₃
intervals followed a stereotypical path, but traversed that path at different speeds (Fig. 2A, ₇₄
top). To understand how a network of neurons can accomplish this, the authors trained a ₇₅
recurrent network model with 200 neurons to produce only the appropriate stimulus-behavior ₇₆
relationships. Wang et al. then applied dimensionality reduction to the activity from the ₇₇
network model. They surprisingly observed that the neural trajectories of the network model ₇₈
also followed a stereotypical path, even though the network model was not trained to reproduce ₇₉
the recorded activity (Fig. 2A, bottom). This population-level correspondence enabled by ₈₀
dimensionality reduction laid the foundation for them to then dissect the network model to ₈₁
understand the core neuronal mechanisms [36]. They found that the input to the network drove ₈₂
the network activity from one fixed point to another, where the transition speed was ₈₃
determined by the depth of the energy basin created by the input (Fig. 2A, bottom). ₈₄

   Other studies have also used this approach to understand how the time course of neuronal ₈₅
activity relates to computations underlying motor control [37–40] and decision ₈₆
making [13, 41, 42]. In each of these studies, a network model was constructed without ₈₇
referencing the recorded activity. Dimensionality reduction was applied to extract neural ₈₈
trajectories to obtain a correspondence between the neuronal recordings and network models. ₈₉
To study the neuronal mechanisms underlying the observed time courses, the network models ₉₀
were then dissected to reveal dynamical structures, such as fixed points or point ₉₁
attractors [13, 41, 42], line attractors [13], and oscillatory modes [37–40]. Whether or not these ₉₂
dynamical structures are indeed at play in real neuronal networks is still an open question. ₉₃
Nevertheless, these studies are beginning to demonstrate that it is at least fruitful to interpret ₉₄
neuronal activity in terms of these dynamical structures, a process facilitated by dimensionality ₉₅
reduction. ₉₆

Functionally-defined neuronal subspaces ₉₇

Recent studies have investigated how distinct types of information encoded by the *same* ₉₈
neuronal population can be parsed by downstream brain circuits [44–46]. An enticing proposal ₉₉
is that different types of information are encoded in different subspaces within the population ₁₀₀
activity space, where the subspaces are identified using dimensionality reduction. For example, ₁₀₁
Kaufman et al. [14] asked how it is possible for neurons in the motor cortex to be active during ₁₀₂
motor preparation, yet not generate an arm movement. They found that motor cortical activity ₁₀₃
during motor preparation resided outside of the activity subspace most related to muscle ₁₀₄
contractions. This allows the motor cortex to prepare arm movements without driving ₁₀₅
downstream circuits, a characteristic which can be implemented by a linear readout mechanism. ₁₀₆
This concept of functionally-defined neuronal subspaces has also been used in other studies of ₁₀₇
motor control [47, 48], decision making [21, 49], short-term memory [21, 50], learning [15], and ₁₀₈
stimulus encoding [51]. ₁₀₉

   To understand how a neuronal circuit can implement and exploit such functionally-defined ₁₁₀
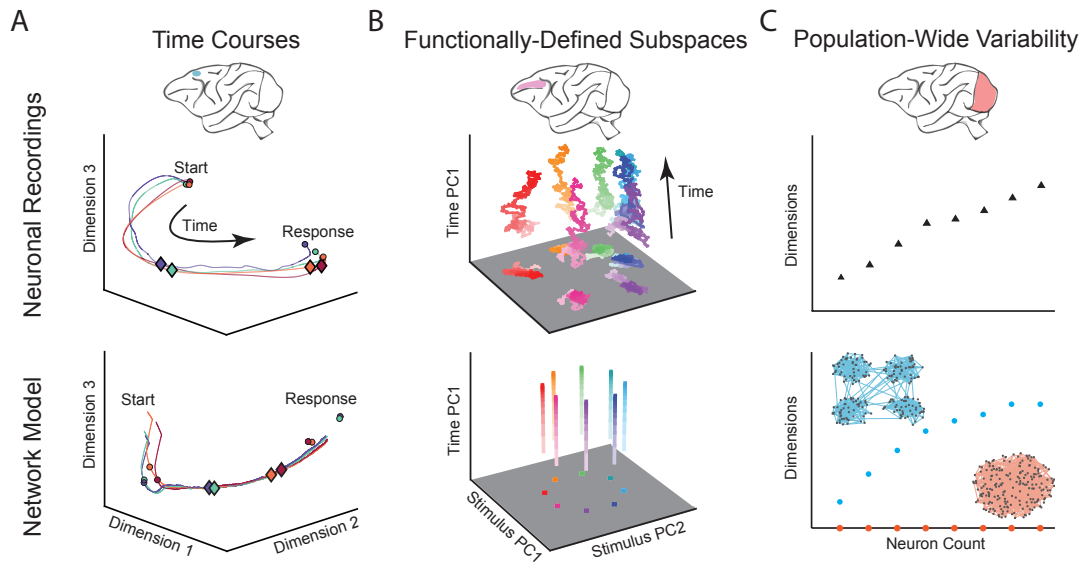
4

Fig 2. Examples of comparing neuronal recordings and network models using dimensionality reduction. A. (Top) Population activity time courses from medial frontal cortex during a time production task. Each trajectory represents a different produced time interval. Diamonds indicate 500 ms after the start of the time production interval. (Bottom) Population activity time courses for a recurrent neural network trained to perform the same time production task. Circles indicate fixed points of the network. Adapted with permission from [17]. B. (Top) Delay period activity from prefrontal cortex during a delayed saccade task. Each trajectory represents a different stimulus condition. The trajectories for different stimuli remains well-separated in a stimulus subspace throughout the delay period. (Bottom) Network model activity demonstrating similar subspace stability. Adapted with permission from [16]. C. (Top) Dimensionality of population-wide neuronal variability in primary visual cortex increases with the number of neurons recorded. (Bottom) A similar dimensionality trend is observed for a spiking network model with clustered excitatory connections (blue), but not for a model with unstructured connectivity (red). Adapted with permission from [43].

neuronal subspaces, one can construct a network model that reproduces the empirical observations, then dissect the network to study mechanisms. Mante et al. [13] applied dimensionality reduction to recordings in prefrontal cortex to find that motion and color of the visual stimulus were encoded in distinct subspaces. They then trained a recurrent network model with 100 neurons to produce only the appropriate stimulus-behavior relationships. When they applied the same dimensionality reduction method to the network model activity, they surprisingly found that the motion and color of the visual stimulus were also encoded in distinct subspaces, even though the network model was not trained to reproduce the recorded activity. This population-level correspondence between the network model and recordings was enabled by dimensionality reduction and went beyond comparisons based on individual neurons or pairs of neurons. Mante et al. then dissected the network model to uncover how the two types of information encoded by a single neural population can be selectively used to form a decision.

Dimensionality reduction has also revealed that, in some cases, standard network models do not reproduce the functionally-defined subspaces identified from neuronal recordings. For example, Murray et al. [16] applied dimensionality reduction to recordings in prefrontal cortex

5

during a working memory task to find that, even though firing rates of individual neurons   126
changed over time, there was a subspace in which the activity stably encoded the memorized   127
target location (Fig. 2B, top). They then applied the same analyses to activity from several   128
prominent network models and found that none of them reproduced both the time-varying   129
activity of individual neurons *and* the subspace in which the memory was stably encoded. This   130
provided the impetus to develop a new network model that did reproduce these features of the   131
recorded activity (Fig. 2B, bottom) (see also [52]). As another example, Elsayed et al. [53]   132
found that standard network models do not reproduce the empirical observation described   133
above that neuronal activity during movement preparation and movement execution lie in   134
orthogonal subspaces. Such insights obtained using dimensionality reduction can guide the   135
development of more sophisticated network models.   136

Population-wide neuronal variability   137

The previous sections focus largely on neuronal activity that is averaged across trials and on   138
firing rate-based network models. This naturally obscures the trial-to-trial variability that is a   139
fundamental feature of neuronal responses across the cortex [54], both at the level of single   140
neuron responses [55] as well as variability shared by the population [9, 56]. Theoretical and   141
experimental studies have focused on how the structure of that variability places limits on   142
information coding [57–60], and in turn influences our behavior. At the same time, a growing   143
body of work has demonstrated that variability can be thought of not only as noise to be   144
removed, but also a signature of ongoing decision processes and cognitive variables [61–63]. To   145
move beyond single-neuron and pairwise measurements of neuronal variability, recent studies   146
have begun to consider population-wide measures of neuronal variability [64–68], as enabled by   147
dimensionality reduction. Such measures allow one to i) assess whether the large number of   148
single-neuron and pairwise variability measurements can be succinctly summarized by a small   149
number of variables (e.g., the entire population increasing and decreasing its activity together   150
can be described by a single scalar variable), and ii) relate the population activity on individual   151
experimental trials to behavior [18, 20, 22, 23, 69–71].   152

In parallel with the growing interest in neuronal variability have been attempts to create   153
network models that exhibit variability matching recorded neurons. In particular, a class of   154
models has used the balance between excitation and inhibition as a way to generate variability   155
as an emergent property of network structure, rather than via an external variable   156
source [57, 72, 73]. In these models, the particular structure of the network has a large impact   157
on the population-wide variability that emerges. Using the lens of factor analysis, Williamson et   158
al [43] found that the dimensionality of spontaneous activity fluctuations in V1 neurons   159
increases with the number of recorded neurons (Fig. 2C, top). This was more consistent with   160
activity generated by networks with clustered excitatory connections [6] than networks with   161
unstructured connectivity [72] (Fig. 2C, bottom). The combination of population-wide   162
measures of variability (in this case, dimensionality) and the ability to manipulate model   163
network structures facilitated an understanding of how features of variability observed in   164
biological networks relate to network structure.   165

The approach of using dimensionality reduction to compare the population-wide variability   166
of neuronal recordings and network models has also been applied to study spontaneous versus   167
evoked activity [74], the activity of different classes of neurons [75], and the activity during   168
different behavioral conditions, such as attention [66, 67]. Dimensionality reduction has also   169

6

been used to analyze population activity from balanced network models to help identify the   170
crucial network architecture and synaptic timescales required to produce the low-dimensional   171
shared variability that is widely reported in neuronal recordings [67, 73]. Together these studies   172
demonstrate the power of combining dimensionality reduction and network models to   173
understand the mechanisms and effects of neuronal variability.   174

Conclusion   175

Dimensionality reduction has enabled incisive comparisons between biological and model   176
networks in terms of population activity time courses, functionally-defined neuronal subspaces,   177
and population-wide neural variability. Such comparisons result in either i) a correspondence   178
between the the neural recordings and the network model, in which case the model can be   179
dissected to understand underlying network mechanisms, or ii) discrepancies between the neural   180
recordings and standard network models, leading to the development of improved models. This   181
approach (cf. Fig. 1) has already provided important insight into the neuronal mechanisms   182
underlying working memory, decision making, and motor control, and is likely to become even   183
more important as the scale of neural recordings and network models grows.   184

An important consideration in network modeling is what aspects of neuronal recordings the   185
model should reproduce. We posit that the population activity structure (including population   186
activity time courses, functionally-defined neuronal subspaces, and population-wide neuronal   187
variability) will provide key signatures of how neurons work together to give rise to brain   188
function. Thus, if a network model is to provide a systems-level account of brain function, we   189
should require it to reproduce the population activity structure of neuronal recordings, in   190
addition to standard spike train statistics based on individual and pairs of neurons.   191

Most studies described here have used neuronal recordings to inform network models via   192
dimensionality reduction. An important future direction is to use network models and   193
dimensionality reduction to design new experiments and form predictions. For example, if one   194
day we can experimentally perturb neuronal activity in specified directions in the population   195
activity space [47], we can test whether driving the population activity in particular directions   196
leads to particular decisions or movements predicted by the network model. The hope is to   197
establish a virtuous cycle, where neuronal recordings and network models closely inform each   198
other through the common ground provided by dimensionality reduction.   199

Acknowledgements   200

7

References

1. Stevenson, I.H. and Kording, K.P. (2011). How advances in neural recording affect data analysis. Nature Neuroscience *14*, 139–142.

2. Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J.M., Diesmann, M., Morrison, A., Goodman, P.H., Harris, F.C., et al. (2007). Simulation of networks of spiking neurons: a review of tools and strategies. Journal of Computational Neuroscience *23*, 349–398.

3. Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. Current Opinion in Neurobiology *46*, 1–6.

4. Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. (2011). On the distribution of firing rates in networks of cortical neurons. Journal of Neuroscience *31*, 16217–16226.

5. Chariker, L., Shapley, R., and Young, L.S. (2016). Orientation selectivity from very sparse LGN inputs in a comprehensive model of macaque V1 cortex. Journal of Neuroscience *36*, 12368–12384.

6. Litwin-Kumar, A. and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. Nature Neuroscience *15*, 1498–1505.

7. Trousdale, J., Hu, Y., Shea-Brown, E., and Josić, K. (2012). Impact of network structure and cellular response on spike time correlations. PLoS computational biology *8*, e1002408.

8. Stringer, C., Pachitariu, M., Steinmetz, N.A., Okun, M., Bartho, P., Harris, K.D., Sahani, M., and Lesica, N.A. (2016). Inhibitory control of correlated intrinsic variability in cortical networks. eLife *5*, e19695.

9. Doiron, B., Litwin-Kumar, A., Rosenbaum, R., Ocker, G.K., and Josić, K. (2016). The mechanics of state-dependent neural correlations. Nature neuroscience *19*, 383–393.

10. Mazor, O. and Laurent, G. (2005). Transient Dynamics versus Fixed Points in Odor Representations by Locust Antennal Lobe Projection Neurons. Neuron *48*, 661–673.

11. Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56.

12. Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature *484*, 62–68.

13. Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

14. Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2014). Cortical activity in the null space: permitting preparation without movement. Nature Neuroscience *17*, 440–448.

15. Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., and Batista, A.P. (2014). Neural constraints on learning. Nature *512*, 423–426.

16. Murray, J.D., Bernacchia, A., Roy, N.A., Constantinidis, C., Romo, R., and Wang, X.J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. Proceedings of the National Academy of Sciences *114*, 394–399.

17. Wang, J., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. Nature Neuroscience *21*, 102–110.

18. Ruff, D.A., Ni, A.M., and Cohen, M.R. (2018). Cognition as a Window into Neuronal Population Space. Annual Review of Neuroscience *41*, 77–97.

19. Cunningham, J.P. and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nature Neuroscience *17*, 1500–1509.

20. Yu, B.M., Cunningham, J.P., Santhanam, G., Ryu, S.I., Shenoy, K.V., and Sahani, M. (2009). Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. Journal of Neurophysiology *102*, 614–635.

21. Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. eLife *5*, e10989.

22. Pandarinath, C., O'Shea, D.J., Collins, J., Jozefowicz, R., Stavisky, S.D., Kao, J.C., Trautmann, E.M., Kaufman, M.T., Ryu, S.I., Hochberg, L.R., et al. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. Nature Methods *15*, 805–815.

23. Williams, A.H., Kim, T.H., Wang, F., Vyas, S., Ryu, S.I., Shenoy, K.V., Schnitzer, M., Kolda, T.G., and Ganguli, S. (2018). Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. Neuron *98*, 1099–1115.e8.

24. Wilson, H.R. and Cowan, J.D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik *13*, 55–80.

25. Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences *79*, 2554–2558.

26. Brody, C.D., Romo, R., and Kepecs, A. (2003). Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. Current Opinion in Neurobiology *13*, 204–211.

27. Machens, C.K., Romo, R., and Brody, C.D. (2005). Flexible Control of Mutual Inhibition: A Neural Model of Two-Interval Discrimination. Science *307*, 1121–1124.

28. Wang, X.J. (2008). Decision Making in Recurrent Neuronal Circuits. Neuron *60*, 215–234.

29. Wang, X.J. (2001). Synaptic reverberation underlying mnemonic persistent activity. Trends in Neurosciences *24*, 455–463.

30. Wimmer, K., Nykamp, D.Q., Constantinidis, C., and Compte, A. (2014). Bump attractor   283
    dynamics in prefrontal cortex explains behavioral precision in spatial working memory.   284
    Nature Neuroscience *17*, 431–439.   285

31. Chaudhuri, R. and Fiete, I. (2016). Computational principles of memory. Nature   286
    Neuroscience *19*, 394–403.   287

32. Seung, H.S. (1996). How the brain keeps the eyes still. Proceedings of the National   288
    Academy of Sciences *93*, 13339–13344.   289

33. Miri, A., Daie, K., Arrenberg, A.B., Baier, H., Aksay, E., and Tank, D.W. (2011). Spatial   290
    gradients and multidimensional dynamics in a neural integrator circuit. Nature   291
    Neuroscience *14*, 1150–1159.   292

34. Shenoy, K.V., Sahani, M., and Churchland, M.M. (2013). Cortical Control of Arm   293
    Movements: A Dynamical Systems Perspective. Annual Review of Neuroscience *36*,   294
    337–359.   295

35. Rabinovich, M., Huerta, R., and Laurent, G. (2008). Transient Dynamics for Neural   296
    Processing. Science *321*, 48–50.   297

36. Sussillo, D. and Barak, O. (2013). Opening the Black Box: Low-Dimensional Dynamics   298
    in High-Dimensional Recurrent Neural Networks. Neural Computation *25*, 626–649.   299

37. Hennequin, G., Vogels, T., and Gerstner, W. (2014). Optimal Control of Transient   300
    Dynamics in Balanced Networks Supports Generation of Complex Movements. Neuron   301
    82, 1394 – 1406.   302

38. Sussillo, D., Churchland, M.M., Kaufman, M.T., and Shenoy, K.V. (2015). A neural   303
    network that finds a naturalistic solution for the production of muscle activity. Nature   304
    Neuroscience *18*, 1025–1033.   305

39. Michaels, J.A., Dann, B., and Scherberger, H. (2016). Neural Population Dynamics   306
    during Reaching Are Better Explained by a Dynamical System than Representational   307
    Tuning. PLoS Computational Biology *12*, e1005175.   308

40. Russo, A.A., Bittner, S.R., Perkins, S.M., Seely, J.S., London, B.M., Lara, A.H., Miri, A.,   309
    Marshall, N.J., Kohn, A., Jessell, T.M., et al. (2018). Motor Cortex Embeds Muscle-like   310
    Commands in an Untangled Population Response. Neuron *97*, 953–966.e8.   311

41. Carnevale, F., de Lafuente, V., Romo, R., Barak, O., and Parga, N. (2015). Dynamic   312
    Control of Response Criterion in Premotor Cortex during Perceptual Detection under   313
    Temporal Uncertainty. Neuron *86*, 1067–1077.   314

42. Chaisangmongkon, W., Swaminathan, S.K., Freedman, D.J., and Wang, X.J. (2017).   315
    Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential,   316
    Category-Based Decisions. Neuron *93*, 1504–1517.e4.   317

43. Williamson, R.C., Cowley, B.R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M.A.,   318
    and Yu, B.M. (2016). Scaling Properties of Dimensionality Reduction for Neural   319
    Populations and Network Models. PLoS Computational Biology *12*, e1005141.   320

10

44. Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for   321
    higher cognition. Current Opinion in Neurobiology *37*, 66–74.   322

45. Park, I.M., Meister, M.L.R., Huk, A.C., and Pillow, J.W. (2014). Encoding and decoding   323
    in parietal cortex during sensorimotor decision-making. Nature Neuroscience *17*,   324
    1395–1403.   325

46. Pagan, M. and Rust, N.C. (2014). Quantifying the signals contained in heterogeneous   326
    neural responses and determining their relationships with task performance. Journal of   327
    Neurophysiology *112*, 1584–1598.   328

47. Li, N., Daie, K., Svoboda, K., and Druckmann, S. (2016). Robust neuronal dynamics in   329
    premotor cortex during motor planning. Nature *532*, 459–464.   330

48. Miri, A., Warriner, C.L., Seely, J.S., Elsayed, G.F., Cunningham, J.P., Churchland,   331
    M.M., and Jessell, T.M. (2017). Behaviorally Selective Engagement of Short-Latency   332
    Effector Pathways by Motor Cortex. Neuron *95*, 683 – 696.e11.   333

49. Raposo, D., Kaufman, M.T., and Churchland, A.K. (2014). A category-free neural   334
    population supports evolving demands during decision-making. Nature Neuroscience *17*,   335
    1784–1792.   336

50. Daie, K., Goldman, M., and Aksay, E.F. (2015). Spatial Patterns of Persistent Neural   337
    Activity Vary with the Behavioral Context of Short-Term Memory. Neuron *85*, 847–860.   338

51. Cowley, B.R., Smith, M.A., Kohn, A., and Yu, B.M. (2016). Stimulus-Driven Population   339
    Activity Patterns in Macaque Primary Visual Cortex. PLoS Computational Biology *12*,   340
    e1005185.   341

52. Druckmann, S. and Chklovskii, D. (2012). Neuronal Circuits Underlying Persistent   342
    Representations Despite Time Varying Activity. Current Biology *22*, 2095–2103.   343

53. Elsayed, G.F., Lara, A.H., Kaufman, M.T., Churchland, M.M., and Cunningham, J.P.   344
    (2016). Reorganization between preparatory and movement population responses in   345
    motor cortex. Nature Communications *7*, 13239.   346

54. Renart, A. and Machens, C.K. (2014). Variability in neural activity and behavior.   347
    Current Opinion in Neurobiology *25*, 211–220.   348

55. Faisal, A.A., Selen, L.P., and Wolpert, D.M. (2008). Noise in the nervous system. Nature   349
    Reviews Neuroscience *9*, 292–303.   350

56. Cohen, M.R. and Kohn, A. (2011). Measuring and interpreting neuronal correlations.   351
    Nature Neuroscience *14*, 811–819.   352

57. Shadlen, M.N. and Newsome, W.T. (1998). The Variable Discharge of Cortical Neurons: 353
    Implications for Connectivity, Computation, and Information Coding. Journal of   354
    Neuroscience *18*, 3870–3896.   355

58. Abbott, L.F. and Dayan, P. (1999). The Effect of Correlated Variability on the Accuracy   356
    of a Population Code. Neural Computation *11*, 91–101.   357

11

59. Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nature Reviews Neuroscience 7, 358–366.

60. Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nature Neuroscience 17, 1410–1417.

61. Cohen, M.R. and Maunsell, J.H.R. (2009). Attention improves performance primarily by reducing interneuronal correlations. Nature Neuroscience 12, 1594–1600.

62. Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2009). Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. Neuron 63, 879–888.

63. Nienborg, H., R. Cohen, M., and Cumming, B.G. (2012). Decision-related activity in sensory neurons: correlations among neurons and with behavior. Annual Review of Neuroscience 35, 463–483.

64. Ecker, A., Berens, P., Cotton, R.J., Subramaniyan, M., Denfield, G., Cadwell, C., Smirnakis, S., Bethge, M., and Tolias, A. (2014). State Dependence of Noise Correlations in Macaque Primary Visual Cortex. Neuron 82, 235–248.

65. Rabinowitz, N.C., Goris, R.L., Cohen, M., and Simoncelli, E.P. (2015). Attention stabilizes the shared gain of V4 populations. eLife 4, e08998.

66. Kanashiro, T., Ocker, G.K., Cohen, M.R., and Doiron, B. (2017). Attentional modulation of neuronal variability in circuit models of cortex. eLife 6, e23978.

67. Huang, C., Ruff, D., Pyle, R., Rosenbaum, R., Cohen, M., and Doiron, B.D. (2017). Circuit-based models of shared variability in cortical networks. bioRxiv .

68. Lin, I.C., Okun, M., Carandini, M., and Harris, K.D. (2015). The Nature of Shared Cortical Variability. Neuron 87, 644–656.

69. Cohen, M.R. and Maunsell, J.H. (2010). A neuronal population measure of attention predicts behavioral performance on individual trials. Journal of Neuroscience 30, 15241–15253.

70. Kiani, R., Cueva, C., Reppas, J., and Newsome, W. (2014). Dynamics of Neural Population Responses in Prefrontal Cortex Indicate Changes of Mind on Single Trials. Current Biology 24, 1542–1547.

71. Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2015). Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex. eLife 4, e04677.

72. Van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science 274, 1724–1726.

73. Rosenbaum, R., Smith, M.A., Kohn, A., Rubin, J.E., and Doiron, B. (2017). The spatial structure of correlated neuronal variability. Nature Neuroscience 20, 107–114.

74. Mazzucato, L., Fontanini, A., and La Camera, G. (2016). Stimuli Reduce the Dimensionality of Cortical Activity. Frontiers in Systems Neuroscience 10.

12

75. Bittner, S.R., Williamson, R.C., Snyder, A.C., Litwin-Kumar, A., Doiron, B., Chase, S.M., Smith, M.A., and Yu, B.M. (2017). Population activity structure of excitatory and inhibitory neurons. PLoS One *12*, e0181773.

Mante et al. [13]*: In a context-dependent decision-making task, the authors found that color and motion information were encoded in distinct subspaces of PFC population activity. A recurrent network model trained to perform the same task revealed a network-level mechanism of how the two types of information can be selectively used to form a decision.

Murray et al. [16]**: The authors analyzed population activity recorded in prefrontal cortex and identified a low-dimensional subspace in which stimulus information was reliably encoded, despite the fact that individual neurons showed substantial time-varying activity. They then found that standard models did not reproduce this empirical observation, and proceeded to develop a "stable subspace" model that did reproduce this observation.

Wang et al. [17]**: Recording from medial frontal cortex during a timing task, the authors found that population activity time courses followed a stereotypical path, but traversed the path at different speeds based on the duration of the timing interval. They found similar trends in a recurrent network model trained to perform the task and showed that speed of traversal was determined by the network inputs.

Sussillo et al. [38]*: The authors trained a recurrent network model to produce muscle activity patterns observed in an arm reaching task. The model activity surprisingly showed rotational dynamics that mimicked those observed empirically in M1 population recordings.

Chaisangmongkan et al. [42]*: The authors found that PFC and LIP neurons show mixed selectivity during a delayed match-to-category task, and that the neural trajectories extracted using dimensionality reduction are interpretable during each epoch of the task. They then constructed a recurrent network model to understand the network principles that govern the activity time courses during this task.

Williamson et al. [43]**: This study compared the population activity structure of V1 recordings and spiking network models while varying the number of neurons and trials analyzed. The scaling trends of the V1 recordings better resembled a model with clustered excitatory connections than one with unstructured connectivity.

Elsayed et al. [53]*: This study found that M1 population activity during movement preparation and movement execution resides in orthogonal subspaces. Standard network models did not reproduce this empirical observation.

Mazzucato et al. [74]*: Comparing gustatory cortex recordings and spiking network models, the authors examined how the dimensionality of population activity grows with population size during spontaneous and evoked activity. They then developed a theoretical upper bound on dimensionality based on the level of pairwise correlations.

14