Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

The quest for interpretable models of neural population activity

Matthew R Whiteway[1] and Daniel A Butts[2]

Many aspects of brain function arise from the coordinated activity of large populations of neurons. Recent developments in neural recording technologies are providing unprecedented access to the activity of such populations during increasingly complex experimental contexts; however, extracting scientific insights from such recordings requires the concurrent development of analytical tools that relate this population activity to system-level function. This is a primary motivation for latent variable models, which seek to provide a low-dimensional description of population activity that can be related to experimentally controlled variables, as well as uncontrolled variables such as internal states (e.g. attention and arousal) and elements of behavior. While deriving an understanding of function from traditional latent variable methods relies on low-dimensional visualizations, new approaches are targeting more interpretable descriptions of the components underlying system-level function.

Addresses
[1] Zuckerman Mind Brain Behavior Institute, Jerome L Greene Science Center, Columbia University, 3227 Broadway, 5th Floor, Quad D, New York, NY 10027, USA
[2] Department of Biology and Program in Neuroscience and Cognitive Science, University of Maryland, 1210 Biology-Psychology Bldg. #144, College Park, MD 20742, USA

Corresponding author: Butts, Daniel A (dab@umd.edu)

Current Opinion in Neurobiology 2019, **58**:86–93

This review comes from a themed issue on **Computational neuroscience**

Edited by **Máté Lengyel** and **Brent Doiron**

https://doi.org/10.1016/j.conb.2019.07.004

Introduction

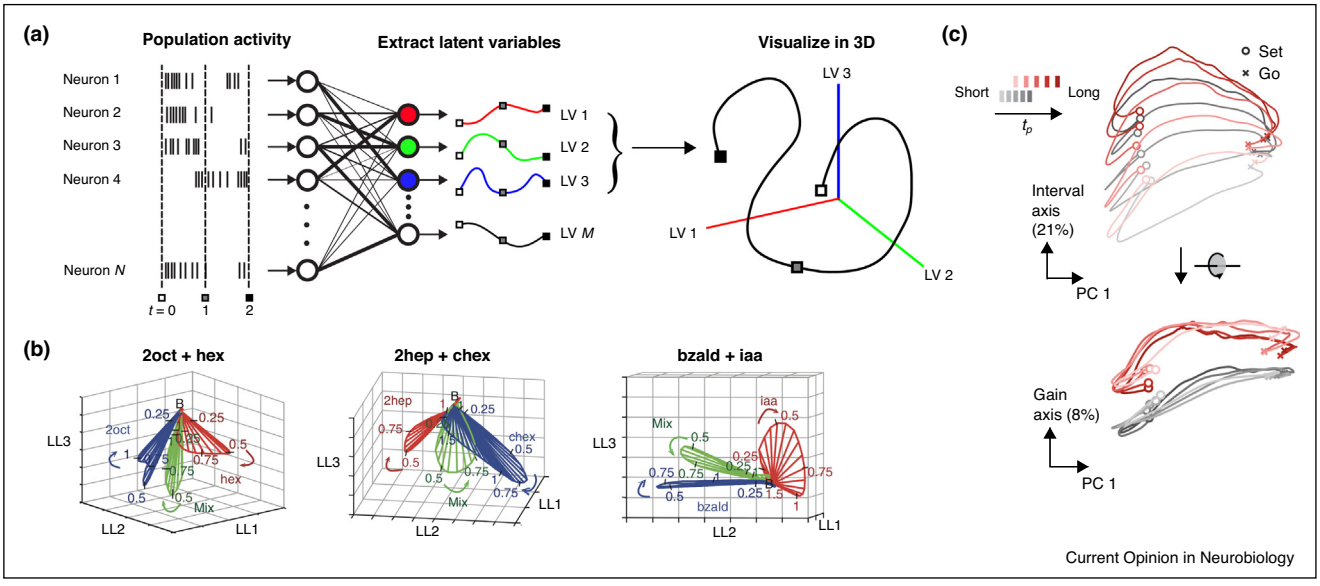
Most of the tasks performed by the brain are implemented by large networks of interconnected neural populations. One consequence of the resulting distributed population-level computations is that is that the activity of a given neuron often reflects multiple aspects of these computations [1], and can thus be difficult to understand in isolation. Instead, an understanding of how neural activity is connected to brain function likely requires relating such function to neural activity at the population level [2–7].

Establishing this relationship first relies on obtaining large-scale recordings of neural activity, which has recently been enabled through the rapid development of new recording technologies [8–10]. These technologies have revealed complex population activity patterns, which in turn present a second challenge: developing computational models that relate such complex activity to the system-level functions of interest, such as sensory processing, decision making, or motor control. As the number of simultaneously recorded neurons increases and experimental paradigms become more complex (e.g. awake animals performing naturalistic behaviors), it becomes increasingly challenging to develop computational models that can describe the population activity while still providing a meaningful interpretation of how it relates to system-level function [4,11].

We suggest that this is the essential goal of latent variable (LV) models in neuroscience, which infer a small number of unobserved — or 'latent' — variables that represent the structure of the high-dimensional neural activity (Figure 1a). In their simplest application, LV models are a form of dimensionality reduction, providing a means to visualize the population activity via its low-dimensional representation [2]. However, such visualization generally depends on finding meaningful representations with three or fewer LVs, which is often insufficient for much of the neural data now collected.

In this review we highlight three emerging conceptual approaches to LV modeling that move beyond visualization while still providing insight into the computations that underlie the function of the observed neural system. The first approach identifies LVs simply as the dimensions of population activity that correlate with experimentally controlled or 'compelled' variables, such as sensory stimuli or motor output, respectively. The second approach places explicit mathematical assumptions on the LVs and their interactions, which has been used to derive LVs that relate to processes in the brain not generally controlled by the experimenter (e.g. attention and arousal). The third approach models the dynamics of LVs, such that the computations of interest are performed through the evolution of a dynamical system that is learned from neural data.

Figure 1



Using controlled experimental variables to identity relevant dimensions in neural population activity.
(a) Schematic demonstrating linear dimensionality reduction. Activity from *N* neurons (*left*) is projected down onto *M* dimensions, or latent variables (LVs), such that each LV is a linear combination of the single-neuron activities. The population activity at each point in time (*middle*) thus has a corresponding set of LV values, which can be visualized in three (or less) dimensions (*right*). **(b)** If more than three LVs are needed to accurately describe population activity, dimensionality reduction can be performed on a subset of experimental trials to highlight distinctions in population activity between the conditions in each trial. For example, the activity of olfactory projection neurons traces out odor-specific trajectories in the latent space, but this space is not low-dimensional if many odors are considered. However, differences in odor-driven trajectories can be visualized using unsupervised methods performed on trials corresponding to individual pairs of odors Reproduced from Ref. [12]. **(c)** When experimental trials contain combinations of controlled variables, demixed PCA (dPCA) [25••] can be used to identify LVs related to particular task variables. Here, a variant of dPCA was applied to understand recorded population activity in a time interval production task [31], where subjects had to saccade after a length of time given by a previously presented interval: either reproducing the interval, or 1.5x the presented interval (considered 'gains' of 1 or 1.5 here). *Top*: dPCA was used to identify the LV most closely related to interval, and a second LV ('PC1') that explains the most remaining variance, which clearly orders the different recordings by interval. *Bottom*: dPCA used to identify a 'gain' axis, and here the population activity is clearly segregated by the two gain conditions. Thus, this demonstrates that both of these aspects of the task were represented in the population activity. Adapted from Ref. [31].

Using controlled experimental variables to identify relevant dimensions in population activity

Traditional approaches for finding a small number of LVs that explain high-dimensional population activity are 'unsupervised', meaning they preserve the structure of the population activity without reference to variables controlled during the experiment. Unsupervised methods such as Principal Component Analysis (PCA) and Factor Analysis (FA) take advantage of the fact that the trial-averaged activity of the recorded neurons (see Box 1) will usually capture the (typically few) variables being explored in the experiment [2,3], and thus the identified LVs can be used to visualize low-dimensional structure in the population activity that is all but invisible on a single-neuron basis (Figure 1a). The resulting low-dimensional description has thus been meaningfully related to experimentally observed variables in a variety of systems, such as odor identity in locust olfaction [12,13], orientation tuning in macaque visual cortex [14], locomotion state in *Caenorhabditis elegans* [15,16], reach direction in macaque motor cortex [17••,18], and head direction in rodent hippocampus [19].

Using unsupervised methods to establish a relationship between neural activity and experimentally controlled variables becomes less straightforward when the first few dimensions of the population activity do not adequately capture the experimental variables of interest. This may occur, for example, when the experimental design itself explores a higher-dimensional space, such as when sensory stimulation itself is high-dimensional (e.g. Refs. [12,20,21]). One approach that retains the intuitive appeal of visualization while simultaneously reducing the complexity of the visualization is to probe pairwise distinctions between individual conditions by applying these traditional methods to subsets of experimental trials (Figure 1b). Of course, such visualization requires a relatively simple experimental design so that all combinations of relevant experimental variables can be assessed.

Box 1 Trial averaging

A key limitation of unsupervised dimensionality reduction methods such as PCA and Factor Analysis is their inability to distinguish between activity driven by experimentally-controlled variables versus activity that is different across repeated trials. Such trial-to-trial variability can often dominate the overall variance in the data — particularly for spiking data that have intrinsic spike-count variability — and thus pull unsupervised LVs away from experimentally relevant dimensions (and into a higher dimensional space). To remove trial-specific fluctuations that are not directly related to experimentally-controlled variables, the activity of individual neurons is often averaged over repeated trials, which removes such variability. Furthermore, because single-trial information is discarded, one benefit of this approach is that population analyses can be performed on neurons that were not simultaneously recorded, and much larger populations can be assembled through serially performed experiments. However, as described below, trial-to-trial variability might be driven by experimentally unconstrained — but nevertheless relevant — variables, and recently large-scale simultaneous recordings [8,22] as well as new statistical approaches [17••,23] can offer an alternative to trial-averaging when single-trial analyses are desired.

Targeting relevant axes of population activity that are related to experimental variables is more difficult when individual trials combine such variables; for example, each trial of a decision-making experiment may involve different combinations of sensory stimuli and motor outputs. Because each of these variables will drive variability in neural activity, methods like PCA will find LVs that mix their effects. Likewise, when variables are manipulated together on each trial, it is not possible to target subsets of trials (as in Figure 1b). In order to disentangle the separate sources of variability, supervised dimensionality reduction methods such as demixed PCA (dPCA) [24,25••] target axes that account for variance associated with each experimental variable. This approach has been useful for understanding how well experimental variables are represented within a neural population across a wide range of experimental paradigms, including motor control [26,27], context-dependent sensory integration [28], delayed match-to-category tasks [29••] and the neural representation of timing [30–32] (Figure 1c).

Relating population activity to uncontrolled experimental variables

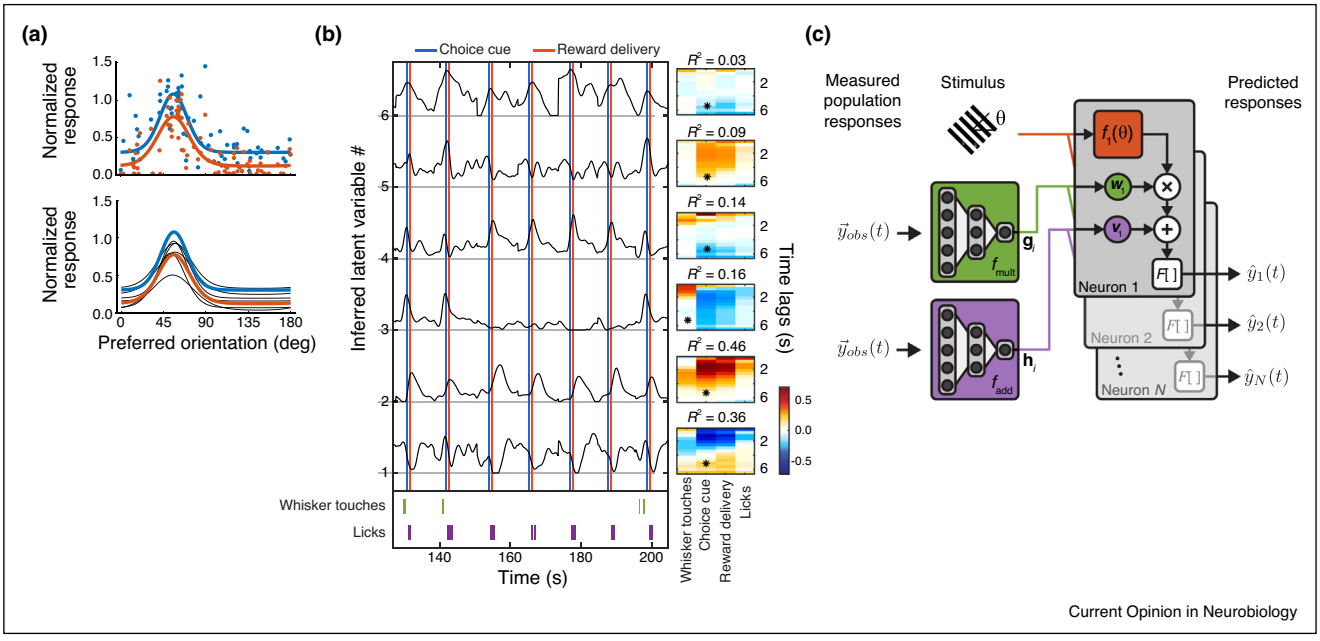
While trial-averaging implicitly removes fluctuations unrelated to experimentally controlled variables (see Box 1), in many experiments the trial-to-trial fluctuations represent a significant amount of the neural activity during the experiment [33–35]. Furthermore, recent large-scale neural recordings demonstrated that trial-to-trial variability (which we will also refer to simply as 'variability') is typically shared among large populations of neurons [36–38] (Figure 2a), and modulated by a variety of factors including task context [39], cortical state [40,41], arousal [42,43•,44], attention [43•,45,46] and motor activity [22,35,42]. Many of these factors are different from those considered in the first section in that they cannot be precisely experimentally controlled — or in some cases, even directly observed — but likely play a critical role in how the recorded neurons process information.

First of all, supervised methods can be used to relate single-trial neural activity to experimentally-uncontrolled variables, which rely on making explicit models of how these observed variables are mapped to neural activity — typically via a generalized linear model (GLM, see Box 2). For example, the summed population activity [36] and local-field potential [33], both of which can be thought of as a proxy for experimentally uncontrolled inputs to the recorded neurons, have been used to successfully predict a significant fraction of single-trial fluctuations in neural activity. Likewise, motor outputs of the animal, such as pupil diameter [22,35,42], running speed [42], and even spontaneous facial and limb movements [22,35] have also been used to describe variability in single neuron responses. The drawback to such supervised approaches for identifying sources of variability is that they are limited by what can be experimentally observed, and thus how directly these observations are related to the underlying neural activity. For example, an internal (unobserved) variable such as arousal can be correlated with — but not equivalent to — pupil diameter and running speed [47]. The resulting supervised model that attempts to relate neural activity and arousal will thus be limited by the strength of the correlation between arousal state and these observed motor outputs.

The fact that much trial-to-trial variability is shared across neurons suggests that it might also be amenable to unsupervised LV approaches, without the need for trial-averaging or information from experimental observables. Indeed, recent studies have found that unsupervised LV models can account for a large fraction of single-trial variability [37,40,43•,48,49•]. However — as with the mixed experimental conditions described in the previous section — unsupervised LV models will by default mix different sources of variability together in the latent space. Furthermore, approaches such as dPCA cannot be applied in cases where the uncontrolled sources of variability are continuously varying (e.g. arousal) because they rely on both trial-averaging and discrete experimental conditions.

One means to disentangle different sources of variability (both controlled and uncontrolled) is to impose mathematical assumptions on the model that allow the LVs to assume distinct functional and/or computational roles [50]. For example, one assumption is to constrain LVs to be non-negative (or 'rectified'), in order to capture the typically spike-based composition of neural inputs [51,52]. The application of this assumption to LV models of activity from the primary somatosensory (barrel) cortex during a whisker-based decision-making task resulted in

Figure 2



Finding relevant dimensions in single-trial population activity.
(a) Trial-to-trial variability is often shared across neural populations, which results in observable shifts in population-level computation. *Top*: Responses from a neural population in macaque primary visual cortex, ordered by their preferred orientation (horizontal axis) are plotted for two trials in response a drifting grating. *Bottom:* This variability across the population is summarized by a fitted 'population tuning curve' for many repeats, demonstrating that the population responses generally increase or decrease together, indicating that single-trial variability is shared across the recorded population. Data available at CRCNS.org from Smith and Kohn (https://doi.org/10.6080/K0NC5Z4X). **(b)** One approach to finding more interpretable LVs is to impose a non-negativity constraint on LVs. This was used to model neural activity in mouse barrel cortex during a decision-making task [52], and results in clear relationships between specific LVs to different trial variables observed during the experiment: the auditory cue that signals the animal to make its choice (blue vertical lines), the onset of reward delivery when the animal makes the correct choice (red vertical lines), the timing of whisker touches against the pole (*bottom*, green), and the timing of licks (*bottom*, purple). *Right:* Once the variables were identified, a GLM was used to demonstrate their relationship to each experimental observable as a function of time-lag. Asterisks label the variable that best predicts the LV based on mean-squared error. Adapted from Ref. [52]. **(c)** Another example of introducing mathematical assumptions on the form of the LVs: the Generalized Affine Model (GAM) [49•] explicitly models multiplicative and additive interactions between LVs and the stimulus response of each neuron. The multiplicative LV g_i and additive LV h_i are inferred from the population activity using neural networks (*green* and *purple* networks, respectively). Adapted from Ref. [49•].

Box 2 Supervised models for single-trial activity

The Generalized Linear Model (GLM) relates neural activity $y(t)$ at time t to a vector of experimental observables $\mathbf{x}(t)$, which could represent, for example, pixel intensities in a visual stimulus, as well as uncontrolled (but observed) variables such as the components of the local field potential [33] or pupil size [22,35]. This relationship is expressed through a weighting of the experimental observables with parameters $\mathbf{w}$, and the result is often passed through a static nonlinearity to produce an estimate of the firing rate $y(t) = f[\mathbf{w} \cdot \mathbf{x}(t)]$. After choosing an appropriate noise distribution (e.g. Gaussian or Poisson), the models are fit by finding the $\mathbf{w}$ that maximizes the likelihood that the model generated the data [53], using optimization approaches such as gradient ascent. While not explicitly a latent variable model (since all variables used are 'observed'), this general framework is similar to that underlying the approaches described below which fit more complex, nonlinear functions.

the segregation of LVs that were driven by whisker contacts versus reward cues, as well as other LVs unrelated to both (Figure 2b).

Another common mathematical assumption is to constrain some LVs to act multiplicatively as a gain signal, for example to modulate stimulus processing [22,37,38,43•,48,49•,54,55] (Figure 2c). Multiplicative model structures have been fruitfully applied to the study of attention in visual cortex, where the inferred LV associated with a multiplicative gain is correlated with attentional state and task performance [43•,46,54]. Such models supported the development of competing hypotheses regarding the origins of the structure of attention-induced correlated variability in neural populations, which led to targeted experiments that could dissociate

the computational mechanisms underlying these different hypotheses [46].

The key feature of this approach to LV modeling that we wish to highlight is that they provide an alternative to visualization by explicitly incorporating a description of the system-level computations. As a result, such models can provide insights even in the case where the relevant population activity is high-dimensional. For example, this is common in sensory systems where sensory-driven activity is by itself high-dimensional [20,21], which will tend to prevent visualization of what is likely a lower-dimensional LV space (representing internal variables) that interacts with sensory-driven responses.

Evaluating the assumed mathematical form of the model (i.e. as a valid description of the true computations performed by the neural system), however, requires — at a minimum — more general nonlinear models of population activity to compare performance to. Such more general models have recently become possible through advances in machine-learning algorithms (Figure 2c), which can be used to fit LV models that can express a range of nonlinear structures [49•,56]. For example, the application of these general nonlinear models to population recordings in anesthetized primary visual cortex have recently validated the assumptions of multiplicative interactions made in previous studies, while finding, in contrast, that multiplicative interactions failed to better describe activity in prefrontal cortex [49•]. Comparisons between structured and more general nonlinear models thus suggest a means to test hypotheses of how LVs interact within and between different populations, which will be an important element of describing system-level function.

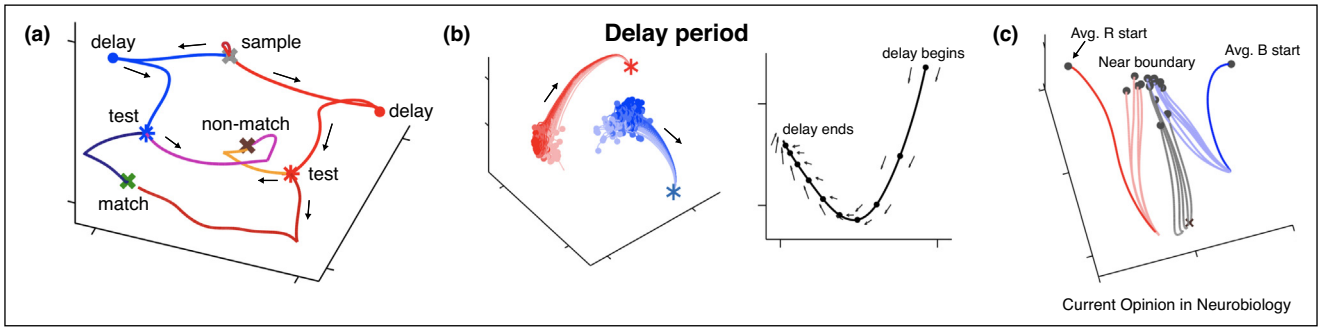
Characterizing computation in dynamical systems

The sensory computations described above predict neural activity largely from experimental variables without regards to the activity within the system itself. However, some systems-level functions, such as perceptual decision-making, appear to require dynamical computations, which integrate inputs and previous neural activity within an area over time [57]. While most brain areas are generally thought to have internal dynamics, the extent to which dynamics play a critical role in the underlying system computation remains an open question in many brain areas [58].

Much of the early work in fitting dynamical models to neural data focused on linear dynamics [59–63], where LVs at a given time point are given by a linear combination of LVs at previous time points. However, in many cases linear dynamics will not be able to capture the richness exhibited by neural data in awake and behaving animals. As a result, recent work has focused on extending these models to learn nonlinear dynamics directly from neural data, which includes switching linear dynamics [64–66], locally linear dynamics [67], parametric approximations [14,68], and the use of recurrent neural networks (RNN) [17••,57].

A parallel line of work has recently emerged which focuses on training both animals and RNNs to perform the same task (rather than fitting the RNN to neural data), and then relates recorded neural activity to that of the artificial neurons – through the use of targeted dimensionality reduction methods like dPCA. This approach has led to new interpretations of complex, high-dimensional neural activity in the context of a diverse range of tasks

Figure 3



Characterizing computation in dynamical systems.
(a) A schematic of the dynamical landscape discovered by an RNN trained on a delayed match-to-category task [29••]. Here, subjects had to indicate whether a second 'test' stimulus, presented with a second-long delay, matched the category of a previously presented 'sample' stimulus. **(b)** The dynamical landscape of the RNN could be studied in detail using standard dynamical systems analyses [69••], resulting in a compelling perspective of how 'working memory' might be generated in the brain. *Left*: trajectories in the 'delay period'; for each category, the starting conditions (dots) result in trajectory for 'red' and 'blue' for a variety of initial conditions, with faded colors showing states closer to the category borders. *Right*: Perturbations from a given trajectory shows that the dynamics force the system state is forced back to a given 'tunnel', gradually slowing to maintain a distinct position for when the 'test' stimulus is presented. **(c)** The behavior of the network during the delay period also demonstrates how misclassifications occur for "difficult" stimuli near the category borders. Noise can perturb the population activity into the wrong tunnel (R to B or B to R), or result in the system going to an incorrect fixed point (grey). All panels adapted from Ref. [29••].

such as context-dependent sensory integration [28], delayed match-to-category tasks [29••] and the neural representation of timing [30,31].

Regardless of whether a model is trained by fitting neural data or by learning a task, new methods must be developed to understand how the resulting nonlinear dynamical system might implement a given computation through dynamical features such as fixed points, line attractors, and limit cycles. For example, [69••] analyzed RNNs that were trained on a variety of tasks by first finding fixed points of the nonlinear dynamics, then linearizing the dynamics around these points, which generated descriptions of model dynamics within the localized regions that shaped the dynamics. This approach has been used, for example, to map out the dynamical landscape of an RNN trained on a delayed match-to-category task [29••], offering a compelling picture of how neural activity might represent previously presented stimuli, as well as suggesting a mechanism for incorrect perceptual judgments (Figure 3). Such interpretability is also implicit in constrained models of the nonlinear dynamics mentioned above. Thus, the increasing ability to understand computations performed by nonlinear dynamical systems [69••], along with the new ability to fit these models directly to neural data [14,17••,64–68], is paving the way towards understanding system-level functions via dynamical-system-based approaches.

Conclusions

In this review we have considered studies across a range of brain areas and experimental paradigms, many of which have found low-dimensional structure in either trial-averaged responses or single-trial variability. Although these results suggest low-dimensionality to be a hallmark of neural activity, such low-dimensional structure might be an artifact of limited data, or the relative simplicity of the experimental design itself [3]. Indeed, with the increasing use of large-scale recordings and more complex task design, it may be that the picture of population activity imparted by current latent variable approaches may not be so simple after all. In order for LV models to continue to offer insights into neural function they must evolve beyond just visualization tools. These models must continue to incorporate additional structure related to specific computations [70], cell types [71], and interacting brain regions [72,73], among others. By doing so, LV models will become tools not just for extracting low-dimensional structure, but for more generally describing system-level neural function.

Conflict of interest statement

Nothing declared.

Acknowledgements

We thank Kenneth Kay for useful discussions and feedback on the manuscript. This work was supported by National Science FoundationIIS-1350990 (DAB), and National Science Foundation NeuroNex AwardDBI-1707398 and the Gatsby Charitable Foundation (MRW).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

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

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

3. Gao P, Ganguli S: **On simplicity and complexity in the brave new world of large-scale neuroscience**. *Curr Opin Neurobiol* 2015, **32**:148-155.

4. Paninski L, Cunningham JP: **Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience**. *Curr Opin Neurobiol* 2018, **50**:232-241.

5. Saxena S, Cunningham JP: **Towards the neural population doctrine**. *Curr Opin Neurobiol* 2019, **55**:103-111.

6. Williamson RC, Doiron B, Smith MA, Yu BM: **Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction**. *Curr Opin Neurobiol* 2019, **55**:40-47.

7. Yamins DLK, DiCarlo JJ: **Using goal-driven deep learning models to understand sensory cortex**. *Nat Neurosci* 2016, **19**:356-365.

8. Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydn Ç *et al.*: **Fully integrated silicon probes for high-density recording of neural activity**. *Nature* 2017, **551**:232-236.

9. Ahrens MB, Orger MB, Robson DN, Li JM, Keller PJ: **Whole-brain functional imaging at cellular resolution using light-sheet microscopy**. *Nat Methods* 2013, **10**:413-420.

10. Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V *et al.*: **Ultrasensitive fluorescent proteins for imaging neuronal activity**. *Nature* 2013, **499**:295-300.

11. Stevenson IH, Körding KP: **How advances in neural recording affect data analysis**. *Nat Neurosci* 2011, **14**:139-142.

12. Saha D, Leong K, Li C, Peterson S, Siegel G, Raman B: **A spatiotemporal coding mechanism for background-invariant odor recognition**. *Nat Neurosci* 2013, **16**:1830-1839.

13. Stopfer M, Jayaraman V, Laurent G: **Intensity versus identity coding in an olfactory system**. *Neuron* 2003, **39**:991-1004.

14. Zhao Y, Memming Park II: *Interpretable Nonlinear Dynamic Modeling of Neural Trajectories*. NeurIPS; 2016.

15. Briggman KL, Abarbanel HDI, Kristan WB: **Optical imaging of neuronal populations during decision-making**. *Science* 2005, **307**:896-901.

16. Kato S, Kaplan HS, Schrödel T, Skora S, Lindsay TH, Yemini E, Lockery S, Zimmer M: **Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans***. *Cell* 2015, **163**:656-669.

17. •• Pandarinath C, O'Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, Trautmann EM, Kaufman MT, Ryu SI, Hochberg LR *et al.*: **Inferring single-trial neural population dynamics using sequential auto-encoders**. *Nat Methods* 2018, **15**:805-815.
The authors develop LFADS, a powerful dynamical system model that can be fit directly to neural activity. They show that LFADS is able to produce highly structured dynamical trajectories on single trials, an advance over prior models that required trial-averaged activity to learn dynamics.

18. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV: **Neural population dynamics during reaching**. *Nature* 2012, **487**:51-56.

19. Chaudhuri R, Gercek B, Pandey B, Peyrache A, Fiete IR: **The population dynamics of a canonical cognitive circuit**. *bioRxiv* 2019:516021.

20. Cowley BR, Smith MA, Kohn A, Yu BM: **Stimulus-driven population activity patterns in macaque primary visual cortex**. *PLoS Comput Biol* 2016, **12**:e1005185.

21. Stringer C, Pachitariu M, Steinmetz NA, Carandini M, Harris KD: **High-dimensional geometry of population responses in visual cortex**. *Nature* 2019, **571**:361-365 http://dx.doi.org/10.1038/s41586-019-1346-5.

22. Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD: **Spontaneous behaviors drive multidimensional, brainwide activity**. *Science* 2019, **364**:255.

23. Gallego JA, Perich MG, Miller LE, Solla SA: **Neural manifolds for the control of movement**. *Neuron* 2017, **94**:978-984.

24. Aoi M, Pillow JW: *Model-based Targeted Dimensionality Reduction for Neuronal Population Data*. 2018:6690-6699.

25. •• Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi X-L, Romo R, Uchida N, Machens CK: **Demixed principal component analysis of neural population data**. *eLife* 2016, **5**.

The authors develop demixed Principal Component Analysis as a supervised dimensionality reduction method that finds dimensions in population activity space related to experimentally-defined variables, aiding in the interpretation of complex and heterogeneous single-neuron responses.

26. Michaels JA, Dann B, Scherberger H: **Neural population dynamics during reaching are better explained by a dynamical system than representational tuning**. *PLoS Comput Biol* 2016, **12**:e1005175.

27. Gallego JA, Perich MG, Naufel SN, Ethier C, Solla SA, Miller LE: **Cortical population activity within a preserved neural manifold underlies multiple motor behaviors**. *Nat Commun* 2018, **9**:4233.

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

29. •• Chaisangmongkon W, Swaminathan SK, Freedman DJ, Wang X-J: **Computing by robust transience: how the fronto-parietal network performs sequential, category-based decisions**. *Neuron* 2017, **93**:1504-1517.e4.

The authors show how a recurrent neural network trained on a delayed match-to-category task uses robust transient trajectories to keep a sample category in working memory, and that this model reproduces a range of features seen in neural responses from lateral intraparietal cortex (LIP) and prefrontal cortex (PFC).

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

31. Remington ED, Narain D, Hosseini EA, Jazayeri M: **Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics**. *Neuron* 2018, **98**:1005-1019.e5.

32. Murakami M, Shteingart H, Loewenstein Y, Mainen ZF: **Distinct sources of deterministic and stochastic components of action timing decisions in rodent frontal cortex**. *Neuron* 2017, **94**:908-919.e7.

33. Cui Y, Liu LD, McFarland JM, Pack CC, Butts DA: **Inferring cortical variability from local field potentials**. *J Neurosci* 2016, **36**:4121-4135.

34. McFarland JM, Cumming BG, Butts DA: **Variability and correlations in primary visual cortical neurons driven by fixational eye movements**. *J Neurosci* 2016, **36**:6225-6241.

35. Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK: **Single-trial neural dynamics are dominated by richly varied movements**. *bioRxiv* 2019:308288 http://dx.doi.org/10.1101/308288.

36. Okun M, Steinmetz NA, Cossell L, Iacaruso MF, Ko H, Bartho P, Moore T, Hofer SB, Mrsic-Flogel TD, Carandini M *et al.*: **Diverse coupling of neurons to populations in sensory cortex**. *Nature* 2015, **521**:511-515.

37. Lin I-C, Okun M, Carandini M, Harris KD: **The nature of shared cortical variability**. *Neuron* 2015, **87**:644-656.

38. Arandia-Romero I, Tanabe S, Drugowitsch J, Kohn A, Moreno-Bote R: **Multiplicative and additive modulation of neuronal tuning with population activity affects encoded information**. *Neuron* 2016, **89**:1305-1316.

39. Bondy AG, Haefner RM, Cumming BG: **Feedback determines the structure of correlated variability in primary visual cortex**. *Nat Neurosci* 2018, **21**:598-606.

40. Ecker AS, Berens P, Cotton RJ, Subramaniyan M, Denfield GH, Cadwell CR, Smirnakis SM, Bethge M, Tolias AS: **State dependence of noise correlations in macaque primary visual cortex**. *Neuron* 2014, **82**:235-248.

41. Pachitariu M, Lyamzin DR, Sahani M, Lesica NA: **State-dependent population coding in primary auditory cortex**. *J Neurosci* 2015, **35**:2058-2073.

42. Vinck M, Batista-Brito R, Knoblich U, Cardin JA: **Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding**. *Neuron* 2015, **86**:740-754.

43. • Rabinowitz NC, Goris RLT, Cohen MR, Simoncelli EP: **Attention stabilizes the shared gain of V4 populations**. *eLife* 2015, **4**:e08998.

The authors develop a statistical model of neural population activity that incorporates several types of gain signals. They find that trial-to-trial fluctuations in gain decrease in attention-directed conditions, providing an explanation for previously observed reductions in correlated variability among pairs of neurons under attention.

44. Goris RLT, Ziemba CM, Movshon JA, Simoncelli EP: **Slow gain fluctuations limit benefits of temporal integration in visual cortex**. *J Vis* 2018, **18**:8.

45. Ni AM, Ruff DA, Alberts JJ, Symmonds J, Cohen MR: **Learning and attention reveal a general relationship between population activity and behavior**. *Science* 2018, **359**:463-465.

46. Denfield GH, Ecker AS, Shinn TJ, Bethge M, Tolias AS: **Attentional fluctuations induce shared variability in macaque primary visual cortex**. *Nat Commun* 2018, **9**:2654.

47. Larsen RS, Waters J: **Neuromodulatory correlates of pupil dilation**. *Front Neural Circuits* 2018, **12**:21.

48. Goris RLT, Movshon JA, Simoncelli EP: **Partitioning neuronal variability**. *Nat Neurosci* 2014, **17**:858-865.

49. • Whiteway MR, Socha K, Bonin V, Butts DA: *Characterizing the Nonlinear Structure of Shared Variability in Cortical Neuron Populations Using Neural Networks*. NBDT; 2019 https://nbdt.scholasticahq.com.

The authors show that in primary visual cortex a nonlinear LV model that incorporates a multiplicative gain term describes neural population activity better than a competing LV model that implements an arbitrary nonlinearity, but this is not the case in prefrontal cortex. This demonstrates the potential for structured LV models to inform our understanding of neural computations in different brain regions.

50. Linderman SW, Gershman SJ: **Using computational theory to constrain statistical models of neural data**. *Curr Opin Neurobiol* 2017, **46**:14-24.

51. Onken A, Liu JK, Karunasekara PPCR, Delis I, Gollisch T, Panzeri S: **Using matrix and tensor factorizations for the single-trial analysis of population spike trains**. *PLoS Comput Biol* 2016, **12**:e1005189.

52. Whiteway MR, Butts DA: **Revealing unobserved factors underlying cortical activity with a rectified latent variable model applied to neural population recordings**. *J Neurophysiol* 2017, **117**:919-936.

53. Paninski L: **Maximum likelihood estimation of cascade point-process neural encoding models**. *Network* 2004, **15**:243-262.

54. Ecker AS, Denfield GH, Bethge M, Tolias AS: **On the structure of neuronal population activity under fluctuations in attentional state**. *J Neurosci* 2016, **36**:1775-1789.

55. Goris RLT, Ziemba CM, Movshon JA, Simoncelli EP: **Slow gain fluctuations limit benefits of temporal integration in visual cortex**. *J Vis* 2018, **18**:8-13.

56. Benjamin AS, Fernandes HL, Tomlinson T, Ramkumar P, VerSteeg C, Chowdhury RH, Miller LE, Körding KP: **Modern**

machine learning as a benchmark for fitting neural responses. *Front Comput Neurosci* 2018, **12**:56.

57. Sussillo D: **Neural circuits as computational dynamical systems**. *Curr Opin Neurobiol* 2014, **25**:156-163.

58. Seely JS, Kaufman MT, Ryu SI, Shenoy KV, Cunningham JP, Churchland MM: **Tensor analysis reveals distinct population structure that parallels the different computational roles of areas M1 and V1**. *PLoS Comput Biol* 2016, **12**:e1005164.

59. Macke JH, Buesing L, Cunningham JP, Yu BM, Shenoy KV, Sahani M: *Empirical Models of Spiking in Neural Populations*. NeurIPS; 2011.

60. Pachitariu M, Petreska B, Sahani M: *Recurrent Linear Models of Simultaneously-recorded Neural Populations*. 2013:3138-3146.

61. Archer EW, Köster U, Pillow JW, Macke JH: *Low-dimensional Models of Neural Population Activity in Sensory Cortical Circuits*. 2014:343-351.

62. Smith AC, Brown EN: **Estimating a state-space model from point process observations**. *Neural Comput* 2003, **15**:965-991.

63. Yu BM, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, Sahani M: **Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity**. *J Neurophysiol* 2009, **102**:614-635.

64. Petreska B, Yu BM, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, Sahani M: *Dynamical Segmentation of Single Trials From Population Neural Data*. 2011:756-764.

65. Wei Z, Inagaki H, Li N, Svoboda K, Druckmann S: **An orderly single-trial organization of population dynamics in premotor cortex predicts behavioral variability**. *Nat Commun* 2019, **10**:216.

66. Nassar J, Linderman SW, Bugallo M, Park IM: *Tree-structured Recurrent Switching Linear Dynamical Systems for Multi-scale Modeling*. 2019.

67. Hernandez D, Moretti AK, Wei Z, Saxena S, Cunningham J, Paninski L: *A Novel Variational Family for Hidden Nonlinear Markov Models*. 2019.

68. Duncker L, Bohner G, Boussard J, Sahani M: *Learning Interpretable Continuous-time Models of Latent Stochastic Dynamical Systems*. 2019:1726-1734.

69. •• Sussillo D, Barak O: **Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks**. *Neural Comput* 2013, **25**:626-649.
The authors show how the nonlinear dynamics of a recurrent neural network can be understood by considering a locally linear approximation to the dynamics around each fixed point, allowing them to extract the computational strategies employed by trained networks.

70. Latimer KW, Yates JL, Meister MLR, Huk AC, Pillow JW: **Single-trial spike trains in parietal cortex reveal discrete steps during decision-making**. *Science* 2015, **349**:184-187.

71. Bittner SR, Williamson RC, Snyder AC, Litwin-Kumar A, Doiron B, Chase SM, Smith MA, Yu BM: **Population activity structure of excitatory and inhibitory neurons**. *PLoS One* 2017, **12**: e0181773.

72. Semedo JD, Zandvakili A, Machens CK, Yu BM, Kohn A: **Cortical areas interact through a communication subspace**. *Neuron* 2019, **102**:249-259 http://dx.doi.org/10.1016/j.neuron.2019.01.026.

73. Perich MG, Gallego JA, Miller LE: **A neural population mechanism for rapid learning**. *Neuron* 2018, **100**:964-976.e7.