from the environment to blend in and match other objects, as exemplified by decorator crabs and caddis fly larvae.

How is camouflage connected to animal cognition? Cognitive processes also influence what makes an effective camouflage, beyond sensory processing (such as visual detection). As the brain may interpret stimuli differently, it may affect predator behavior and thus have consequences on camouflage efficacy. Predators have been shown to be worse at finding camouflaged prey when prey populations are polymorphic in appearance. This is because under some conditions predators concentrate on prey types that they have recent experience with, forming 'search images' for these and thus overlooking the rare morphs. As a result, negative frequency-dependent selection can maintain polymorphic prey and fluctuations in morph frequency. Learning and cognitive processes may also have a major effect on the value of different camouflage strategies. Predators learn some types of camouflage more quickly than others, especially those involving high contrast patterns. The value of a given type of camouflage thus depends not just on initial detection, but also on predator experience and cognition.

Where can I find out more?

Bond, A.B., and Kamil, A.C. (2002). Visual predators select for crypticity and polymorphism in virtual prey. Nature *415*, 609–613.

Diamond, J., and Bond, A.B. (2013). Concealing Coloration in Animals. Harvard University Press, Massachusetts.

Hanlon, R.T. (2007). Cephalopod dynamic camouflage. Curr. Biol. *17*, 400–404.

Lovell, P.G., Ruxton, G.D., Langridge, K.V., and Spencer, K.A. (2013). Egg-laying substrate selection for optimal camouflage by quail. Curr. Biol. *23*, 260–264.

Skelhorn, J., and Rowe, C. (2016). Cognition and the evolution of camouflage. Proc. R. Soc. B. *283*, 20152890.

Skelhorn, J., Rowland, H.M., Speed, M.P., and Ruxton, G.D. (2010). Masquerade: camouflage without crypsis. Science *327*, 51.

Stevens, M. (2016). Cheats and Deceits: How Animals and Plants Exploit and Mislead. Oxford University Press, Oxford.

Stevens, M., and Merilaita, S. (2011). Animal Camouflage: Mechanisms and Function. Cambridge University Press, Cambridge.

Stuart-Fox, D., and Moussalli, A. (2009). Camouflage, communication and thermoregulation: lessons from colour changing organisms. Phil. Trans. R. Soc. B. *364*, 463–470.

Troscianko, J., Wilson-Aggrawal, J., and Stevens, M. (2016). Camouflage predicts survival in ground-nesting birds. Sci. Rep. *6*, 19966.

Centre for Ecology & Conservation, University of Exeter, Penryn Campus, Penryn, TR10 9FE, UK.
*E-mail: martin.stevens@exeter.ac.uk

Primer

Dimensionality reduction in neuroscience

Rich Pang[1], Benjamin J. Lansdell[2], and Adrienne L. Fairhall[3,4,5,*]

The nervous system extracts information from its environment and distributes and processes that information to inform and drive behaviour. In this task, the nervous system faces a type of data analysis problem, for, while a visual scene may be overflowing with information, reaching for the television remote before us requires extraction of only a relatively small fraction of that information. We could care about an almost infinite number of visual stimulus patterns, but we don't: we distinguish two actors' faces with ease but two different images of television static with significant difficulty. Equally, we could respond with an almost infinite number of movements, but we don't: the motions executed to pick up the remote are highly stereotyped and related to many other grasping motions. If we were to look at what was going on inside the brain during this task, we would find populations of neurons whose electrical activity was highly structured and correlated with the images on the screen and the action of localizing and picking up the remote.

Describing a complex signal, such as a visual scene or a pattern of neural activity, in terms of just a few summarizing features is called *dimensionality reduction*. The core notion of dimensionality reduction is long established in neuroscience. For example, in characterizing the response of a neuron in primary visual cortex (V1), Hubel and Wiesel observed that an object's motion orientation modulated the firing rate of the cell. This allowed them to describe the firing rate as a function of this one variable, rather than of the intensities of all of the pixels in the visual scene. Conversely, dimensionality reduction can be applied to patterns of multi-neuronal activity. We do just that when we map a visual stimulus to just one neuron's firing rate rather than a possibly complex multi-neuron response; or when, for example, we evaluate the effects of attention in terms of changes in the power in a certain frequency band in the local field potential.

As this illustrates, one of the goals of neuroscience is to find interpretable descriptions of what the brain represents and computes. Choosing to describe a V1 neuron's response in terms of the orientation of a moving bar is somewhat arbitrary, however, as the firing rates of V1 cells can be modulated by many other visual features. Further, thinking of the brain's output in terms of the firing rate of individual neurons or the power of the summed electrical signal in a certain frequency band is also an arbitrary choice of representation of neural activity that may not reflect the brain's natural computational 'units'. In general, we ought to seek representations, both of the stimulus and of brain activity, that are concise, complete, and informative about the workings of the nervous system, and yet which are not biased by an experimenter's arbitrary choice. Considering this task from the perspective of dimensionality reduction provides an entry point into principled mathematical techniques that let us discover these representations directly from experimental data, a key step to developing rich yet comprehensible models for brain function.

Single neuron coding

A tenet of sensory neuroscience is that, within a rich and varying world, neurons have evolved to respond to a small set of behaviourally meaningful inputs and to represent them efficiently. And indeed, it is often observed that many sensory neurons' responses can be characterized as depending only on a small set of features of an external stimulus.

An example of such dimensionality reduction is color vision. Light hitting the eye has intensity in a wide range of frequencies. While a spectrophotometer would provide a complete description of the light beam in terms of its power spectrum across all frequencies, our retina has only

three kinds of color sensor, the L, M and S cone types (corresponding to long, medium, and short wavelengths). All we can know about the incoming light is given to us by the activation of those three sensors: because of the unique frequency absorption properties of each cone type, the activation of a given cone type is a function of a weighted sum of the light's intensities at different frequencies. Thus, our color perception is a three-dimensional representation of the original, infinite-dimensional spectrogram that specifies the light's intensity at every frequency.

Further into the visual system, a neuron's response is often a function of only a small set of visual inputs, or features. These features are identified in a given stimulus through a remarkably simple procedure known as *linear filtering*, which consists of simply weighting and summing of the components of a signal according to a given set of weights known as the filter. This is a general procedure that can be applied to any stimulus representation, be it color-spectral components, light intensities in an image, time-varying intensities in a movie, and so forth. Linear filtering produces a single value that expresses the *similarity* of the image to the filter — the extent to which the stimulus feature is present in the image. A geometric illustration makes clear how filtering accomplishes this task (Figure 1).

For example, some retinal ganglion cells (RGC) are excited by, or positively weight, the image intensities at the 'center' of a visual stimulus and are suppressed by, or negatively weight, intensities in the surrounding region. Together, these weights define the *filter*, or a stimulus *feature* that drives the neuron. The RGC's firing can then be predicted by taking an input image, weighting the value of each pixel in the image by the filter values and summing the result. Thus, just as the cone activations *reduce* the full spectrum to three components, here the RGC's activation is reduced from being a function of the full image (specified by its intensity at each pixel) to being a function of a single number that represents a measure of the image's similarity to that neuron's selected feature. More generally one might consider a neuron that is selective for a *sequence* of images, or a short movie. For example, an 'ON' RGC which responds to a particular spot becoming brighter over time can be understood with an appropriate spatio-temporal filter. That is, the neuron would weight the intensities of all the pixels at all recent time points — for example, there would be 2000 weights for 20 frames of a 10 x 10 grayscale image — and the sum of the weighted intensities over both space and time would determine the probability of the neuron's emitting a spike.

It is possible that the neuron's response is sensitive to more than one feature of the stimulus. For example, the RGC might be sensitive not only to the brightening of the spot but also to the speed of the change. In this case, the response could depend on the outputs of multiple filters, and the neuron's response would depend on the similarities to these multiple features. As long as there are many fewer features than there are, in this case, pixels in the movie, this feature representation — the set of similarity values — is a much more compact way to describe the input, and ideally captures everything about the input that is relevant to the response of the neuron.

Generally, the feature or features that a neuron is selective for are not known *a priori*. Dimensionality reduction methods identify relevant features directly from experimental data. The key idea is simple: one presents the system with many random examples of complex stimuli (images, movie segments, and so on) and notes which stimuli make the neuron spike and which do not. One can then use these samples to characterize what is particular to the cases that caused the neuron to respond.

The simplest statistic to look at is the average of the spike-triggering stimulus examples (called the spike-triggered average). In many cases this can lead to accurate spike prediction, for example in an ON retinal ganglion cell that responds primarily to upward deflections in light level. If there are multiple relevant features, they can be found

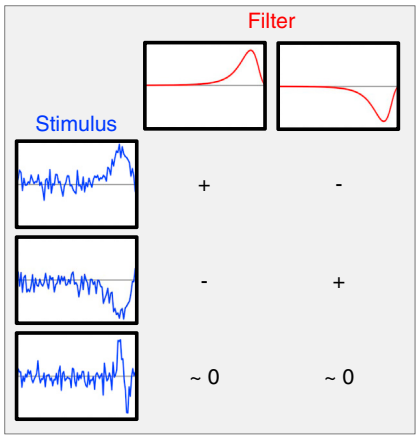

Figure 1. Linear filters detect the presence of specific features.
Linearly filtering a stimulus yields a single number that quantifies how similar the stimulus is to the filter. If the filter shape is a positive deflection, of a dot's luminance over time, for example (upper red trace), then stimuli that resemble positive deflections (upper blue trace) will get filtered to positive values, whereas stimuli that resemble negative deflections (middle blue trace) will get filtered to negative values. The opposite is true for a negative deflection filter. If the stimulus has approximately equal positive and negative deflections (lower blue trace), then filtering it with either a positive or negative deflection filter will yield a value of approximately 0.

using a variety of techniques. One straightforward approach is to analyze the *covariance* of the spike-triggering stimuli (Figure 2A, B) in order to find additional relevant features. This is especially useful in cases when the spike-triggered average alone is not very informative; for example for the ON–OFF retinal ganglion cells, which are triggered either by an upward or a downward change in light level. The spike-triggering stimuli average to almost zero, but computing the covariance of these stimuli allows one to find a set of stimulus features that capture both the upward and downward variations, even, for example, if they have different rates of change. V1 responses have also been found to be best fit by models that include a number of features, where the additional features allow one to account for properties like phase invariance in complex cells, and components that lead to suppression. Multiple features can also be found using methods that use alternate statistical properties like entropy and

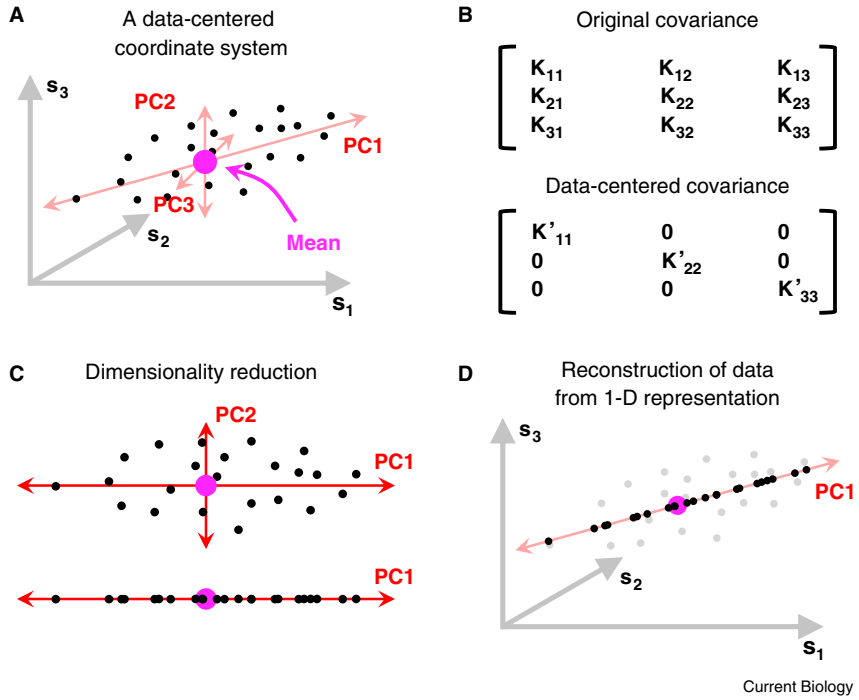

Figure 2. Reducing a dataset's dimensionality by computing its mean and covariance.
Using PCA to reduce the dimensionality of a dataset. (A) One can conceptualize a dataset as a cloud of points. Each point represents an observation (for example, a spike-triggering stimulus or a multi-neuron firing response) and the number of axes is equal to the number of measurements per observation. This data cloud can be characterized by various structural properties, including its mean (pink) and its covariance. The principal components (PCs, red) describe the directions along which the data cloud varies the most. (B) The PCs are computed by finding the *eigenvectors* of the original covariance matrix (top). When viewed in the reference frame of the PCs, all off-diagonal covariances become zero (bottom). (C) To reduce the dimensionality of a dataset, one can view the data in a coordinate system defined only by a few PCs. (D) One can reconstruct the data in the original coordinate system given the data in the reduced coordinate system. Information will always be lost upon doing this, but using the PCs as the reduced coordinates helps to preserve as much relevant information as possible.

mutual information to characterize the spike-triggering stimuli.

An alternative, data-efficient strategy for finding a reasonable *single* feature is to assume that the firing rate is a known function of the stimulus' similarity to that feature and to search for the feature that best predicts the observed spike train. This can be done by finding the feature that maximises the probability of the observed spike train given the feature and the stimulus, following the principle of maximum likelihood. An example is the *generalized linear model* (GLM), in which the firing rate is computed by (1) filtering the recent stimulus to yield a single number indicating how similar that stimulus was to the sought-after feature; and (2) evaluating a chosen nonlinear function (such as an exponential) of this similarity value (Figure 3).

This model can be easily extended to include influences apart from the stimulus, for example the dependence of the firing rate on the recent spiking activity, thereby allowing for the influence of the neuron's refractory period or intrinsic dynamics, the influences of other neurons, and other factors such as context or behavior. Such a model has been used to show how spike rates in retinal ganglion cells are modulated not only by visual stimuli but also by other cells in the network, and that more information about the stimulus can be extracted from the network activity if one accounts for the *interactions* of the cells in the network, rather than their individual activity alone. GLMs have also been used, for example, to identify what features of birdsong cause strong responses in auditory neurons in the zebra finch. While we have focused on linear feature extraction, some methods also consider nonlinear transformations to find features. For example, a sensory neuron in the vibrissa system might be sensitive to the vibrissa's phase in the whisk cycle, which is a nonlinear function of the position.

Given these examples, it becomes clear that thinking of complex stimuli in terms of low-dimensional descriptions can provide insight into the function of different parts of the nervous system. By identifying these descriptions directly from the data, we can learn concise ways of describing the stimuli that make neurons fire, and we can do so in a way that is less influenced by experimental bias. Such a characterization of what a neuron encodes is the first step towards understanding the mechanisms by which it or the network of neurons surrounding it transforms and processes complicated inputs from the natural world.

Multi-neuronal recordings and population codes

In the approaches outlined above, dimensionality reduction was applied to stimulus samples to find features of the input that drive selected single neurons to fire. Given the increasing prevalence of multineuronal recordings, it is natural to ask whether such single-neuron models are sufficient for understanding the neural codes for stimuli and behavior. It is unlikely that every neuron fires independently, in which case each neuron's activity could contribute to a vast number of possible patterns. Rather, neurons may fire in coordinated ways that can be described in terms of a smaller number of population-level features. Given an experimental recording, can we determine how distributed patterns of activity are involved in encoding sensory or behavioral information? How do we identify these patterns from the recorded responses of individual neurons? Understanding the nature of this population code can help to constrain theories about how neural activity patterns are generated and how they underlie computation.

Let's start by thinking about a neural state as characterized by the

firing rates of all of the recorded neurons in some interval of time. This representation has *N* numbers, as many as there are neurons, but we would like to express the neural state as a weighted sum over a small set of activation patterns. To do so, we define a pattern to be an activation profile over the *N* original neurons; some neurons in a given pattern may be highly active and others not at all. We can then ask *which* activation patterns best account for the data, or equivalently, capture the largest amount of the variance in the data. Geometrically this is equivalent to finding a new set of coordinates in which to plot the instantaneous activity that are learned from the data, where the new coordinates describe the activation of patterns, rather than of individual neurons (Figure 2C, D). As in our analysis of spike-triggering *stimuli*, we can do this by computing the *N* x *N* covariance matrix of the firing rates (Figure 2A, B), which tells us which neurons tend to fire together and which tend to fire independently. Analysis of this matrix yields a set of activation patterns, each specified by activity levels for each neuron, and ranked by the variance explained by each. This technique is called principal component analysis (PCA), and cases in which only a few patterns capture a large amount of the variance are indicative of low-dimensional structure.

In the insect antennal lobe, for example, different odorants may activate different neurons to varying degrees (Figure 4A). To tease out how the *population* encodes and differentiates odors, one can use PCA to discover a small set of activation patterns used by the network, regardless of which particular odorant was presented. The resulting activation patterns then define neural *ensembles.* One can now express the activity of the network in terms of the time-varying activations of the ensembles. The time course of the ensemble response to two different odors clearly shows how the neural representation of the two stimuli differed (Figure 4).

In PCA, all snapshots of the original neural state are treated independently, regardless of when they occurred. Our example from the insect antennal lobe, however, suggests that the neural state often evolves smoothly over time. Is it possible to use this knowledge to help our search for relevant neural ensembles? *Gaussian Process Factor Analysis* (GPFA) seeks to do just that. Here, one considers the covariance not only of simultaneously active neurons but also of individual neurons as their firing changes over time. This adds complexity but allows the analysis to include an assumption that firing rates usually change not instantly but rather over some characteristic timescale. This analysis was used, for example, to analyse data from a large number of neurons in primate motor cortex as a monkey made reaches in response to visual targets. When the dimensionality reduction was performed, it became easy to identify distinct neural activity patterns corresponding to distinct parts of the task, such as the target onset, the 'go' cue, and the movement onset. This suggests that the low-dimensional representation of the neural activity may be an important component of the neural computations used to complete the task.

While these methods, PCA and GPFA, identify features of the simultaneous firing of many neurons, one might more generally consider *spatiotemporal* features — distributed patterns of firing that play out over time. This is similar to the identification of important features of a short movie stimulus, as described in the previous section, but now applied to the neural activity. As a familiar example, Fourier analysis would decompose this 'neural movie' into features corresponding to different frequency components. In line with our prevailing theme, however, one would generally like to find features of the activity that are learned directly from the data. For instance, olfactory neurons often show characteristic temporal responses that look quite different from sine waves. So why not represent the neural activity in a way that directly incorporates these characteristic responses? Practically, this can be done by performing PCA on iterations of the neural movie, which may well reveal stereotyped representations of specific odors in the insect antennal lobe. On larger scales, different brain regions show responses that unfold over a distribution of timescales. Such spatiotemporal features can be identified with methods such as dynamic mode decomposition (DMD), or approaches that search for patterns in the sequences of time lags between activations in different areas.

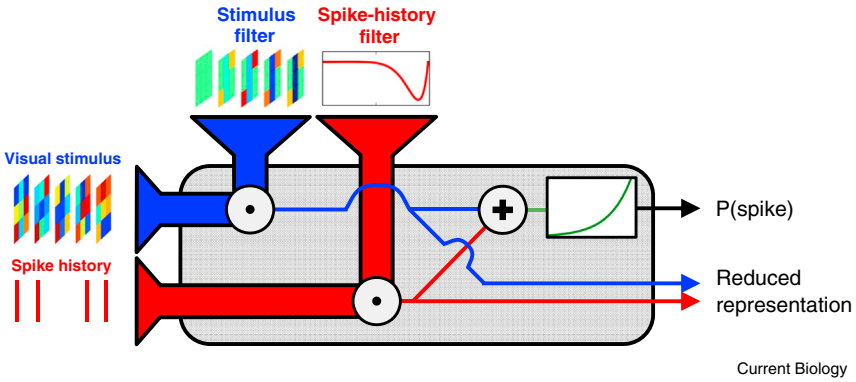

Figure 3. Dimensionality reduction using a generalized linear model (GLM).
Using a GLM one can reduce the dimensionality of a full set of inputs to a neuron (in this case a movie and the neuron's previous spikes) by finding the filters that when applied to the inputs best predict the neuron's spiking probability.

Dual dimensionality reduction

During natural behavior, both the environmental inputs experienced by the animal and the accompanying neural activity vary outside the experimenter's control. Ideally one would like to simultaneously extract the features of the environment that are correlated with neural activity,

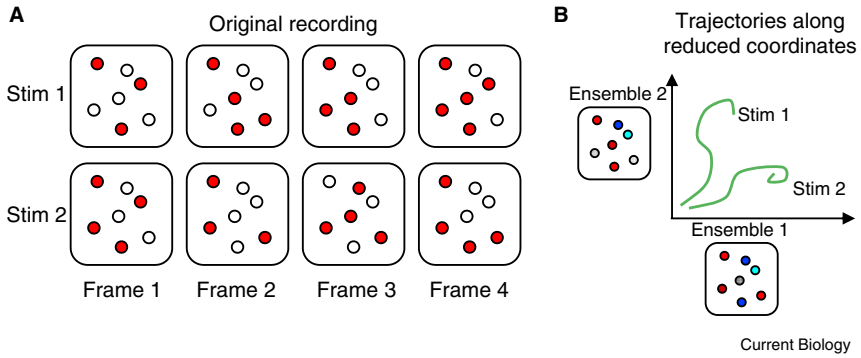

Figure 4. Reducing the dimensionality of a multi-neuron dataset to visualize stimulus-dependent responses.
Stimulus dependence can be hard to identify when looking at a high-dimensional dataset of many neurons recorded over many time points. (A) Diagram of a time-varying neural population recording one might obtain after presenting two different stimuli. The circles represent neurons, with red fill indicating their activity. (B) The time-varying activities of two ensembles (activation patterns) of neurons identified using PCA. The colors of the neurons in each ensemble represent the positive (red) or negative (blue) contribution of each neuron to the ensemble. In this two-dimensional view, the separability of the population responses to the two stimuli becomes more obvious.

and reduced representations of neural activity that 'encode' those stimulus features. Such 'two-sided' dimensionality reduction can be accomplished through methods known as canonical correlation. This has been used, for example, to test hypotheses about how the timing of the activation of a moth's power muscles determines variations in its flight dynamics, although the successful application of such techniques in neuroscience is still in its infancy.

Discussion

If the dynamics of a multi-neuron recording can be largely accounted for by only a small number of features relative to the number of neurons and/or timepoints, what might this tell us about potential models or mechanisms for the neural activity? First, it suggests that neurons do not fire independently, but rather act together in a coordinated manner. This may simply be redundancy: many neurons may act coherently, perhaps to mitigate the effects of discretization due to spiking, and noise at the single neuron level. Examining how these neurons act coherently, for example which neurons participate in which low-dimensional features, may teach us how the neurons are connected anatomically and functionally. Further, the nonlinear dynamics of neural networks can create stable patterns of activity called attractors, which can serve as low-dimensional 'building blocks' for a particular computation. Identifying low-dimensional features in brain activity not only lends credence to the hypothesis that the brain might perform these kinds of computations but can also help reveal what activity patterns correspond to these building blocks. In theory, multiple attractors can exist in the same network; it is possible that different computations use different building blocks, allowing a single network to perform multiple functions, selected by task.

The world around us, complex as it is, is relatively low-dimensional: the familiar visual scenes made up of textures, faces, buildings, and other objects are highly structured and are but a miniscule subset of all possible images, and the physics of the world strongly constrains the sequences of actions that can occur. As it is commonly believed that the developed brain contains an internal model of the environment that it expresses through its structure and activity, it is natural to think that this model should be similarly highly structured, and that the dimensionality reduction characterizing the brain's activity might be related to intrinsic properties of natural sensory stimuli and motor output. The methods discussed here provide a route, if imperfect, to deciphering the coding structure of evoked and spontaneous activity.

FURTHER READING

Aljadeff, Y., Lansdell, B.J., Fairhall, A.L., and Kleinfeld, D. (2016). Analysis of neuronal spike trains, deconstructed. Neuron, in press.
Brunton, B.W., Johnson, L.A., Ojemann, J.G., and Kutz, J.N. (2016). Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. J. Neurosci. Meth. *258*, 1–15.
Calabrese, A., Schumacher, J.W., Schneider, D.M., Paninski, L., and Woolley, S.M.N. (2011). A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. PLoS One *6*, e16104.
Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509.
Fairhall, A.L., Burlingame, C.A., Puchalla, J., Harris, R., and Berry II, M.J. (2006). Feature selectivity in retinal ganglion cells. J. Neurophysiol. *96*, 2724–2738.
Fitzgerald, J.D., Rowekamp, R.J., Sincich, L.C., and Sharpee, T.O. (2011). Second order dimensionality reduction using minimum and maximum mutual information models. PLoS Comput. Biol. *7*, e1002249.
Levi, R., Varona, P., Arshavsky, Y.I., Rabinovich, M.I., and Selverston, A.I. (2005). The role of sensory network dynamics in generating a motor program. J. Neurosci. *25*, 9807–9815.
Mazor, O., and Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. Neuron *48*, 661–673.
Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., and Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature *454*, 995–999.
Sharpee, T., Rust, N.C., and Bialek, W. (2004). Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput. *16*, 223–250.
Shlizerman, E., Riffell, J., and Kutz, J.N. (2014). Data-driven inference of network connectivity for modeling the dynamics of neural codes in the insect antennal lobe. Front. Comp. Neurosci. *8*, 1–14.
Simoncelli, E.P, Pillow, J., Paninski, L. and Schwartz, O. (2004) Characterization of neural responses with stochastic stimuli. InThe Cognitive Neurosciences III, MIT Press, 327–338.
Stopfer, M., Jayaraman, V., and Laurent, G. (2003). Intensity versus identity coding in an olfactory system. Neuron *39*, 991–1004.
Tafreshi, A.K., Nasrabadi, A.M., and Omidvarnia, A.H (2008). Epileptic seizure detection using empirical mode decomposition. IEEE. Int. Symp. Signal Process. Inf. Technol., 238–242.
Yuste, R. (2015). From the neuron doctrine to neural networks. Nat. Rev. Neurosci. *16*, 487–497.

[1]Neuroscience Graduate Program, University of Washington, Box 357270, T-471 Health Sciences Ctr, Seattle, WA 98195, USA. [2]Department of Applied Mathematics, University of Washington, Lewis Hall #202, Box 353925, Seattle, WA 98195, USA. [3]Department of Physiology and Biophysics, University of Washington, 1705 NE Pacific Street, Box 357290, Seattle, WA 98195, USA. [4]WRF UW Institute for Neuroengineering, University of Washington, Box Seattle, WA 98195, USA. [5]Center for Sensorimotor Neural Engineering, University of Washington, Box 37, 1414 NE 42nd St., Suite 204, Seattle, WA 98105, USA.
*E-mail: fairhall@uw.edu