affecting exploration of novel stimuli (**Fig. 1e**, center left). Photostimulation of the mHB terminals in the IPN decreased exploration of novel social and nonsocial stimuli without affecting exploration of familiar stimuli (**Fig. 1f**, center and right). Next, the authors photostimulated the VTA dopaminergic terminals in the IPN. As in the phenomenon of *jamais vu*, this manipulation mimicked the novelty signal, resulting in increased exploration of a familiar mouse (**Fig. 1e**, center right). Interestingly, the photostimulation of dopamine terminals did not affect exploration of inanimate objects. Thus, the novelty signaling pathway may differ for social and nonsocial signals. The authors suggest that different subtypes of VTA dopaminergic neurons may mediate novelty responses to social and nonsocial stimuli.

It is tempting to conclude that novelty is simply the absence of memory-based familiarity. Yet a number of studies have provided evidence that the processing of novelty information and familiarity information can be functionally dissociated in the forebrain medial temporal lobe memory system. A study using c-Fos expression methods combined with structural equation modeling found evidence that, in rats presented with familiar objects, caudal perirhinal cortex activated the entorhinal-to-hippocampal field CA1 pathway, also known as the temporo-ammonic pathway[13]. When rats were presented with novel objects, perirhinal cortex activated the entorhinal-to-dentate gyrus pathway, also known as the perforant pathway. Another c-Fos study showed that exploration of a novel environment increased activation in the hippocampus, the prelimbic prefrontal cortex and the dopaminergic reward circuit[14]. Exploration of a familiar environment, however, increased activation in the amygdala. A better understanding of how the midbrain circuits interact with the forebrain circuits could help explain the human prevalence differences between *déjà vu* and *jamais vu*. Future work could elucidate other neural bases of neuropsychiatric disorders by explaining dysregulation of novelty and familiarity processing, depersonalization, derealization and other symptoms that involve detachment from familiar surroundings.

In this elegant series of experiments, Molas *et al.* have elucidated the mechanisms and circuitry by which novelty transitions to familiarity. A primary contribution of their work is the demonstration that novelty and familiarity are signaled by different pathways, partially overlapping in the IPN, to support novelty preference. These findings may explain why *déjà vu* and *jamais vu* contribute differently to symptom profiles of neuropsychiatric disorders. More importantly, the findings of Molas *et al.* have profound implications for understanding and treating neuropsychiatric disorders in which processing of novelty and familiarity are compromised.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. O'Connor, A.R. & Moulin, C.J.A. *Curr. Psychiatry Rep.* **12**, 165–173 (2010).
2. Ellis, H.D., Luauté, J.P. & Retterstøl, N. *Psychopathology* **27**, 117–120 (1994).
3. Sengoku, A., Toichi, M. & Murai, T. *Psychiatry Clin. Neurosci.* **51**, 23–26 (1997).
4. Molas, S. *et al. Nat. Neurosci.* **20**, 1260–1268 (2017).
5. Fantz, R.L. *Science* **146**, 668–670 (1964).
6. Feinberg, L.M., Allen, T.A., Ly, D. & Fortin, N.J. *Neurobiol. Learn. Mem.* **97**, 7–16 (2012).
7. Ho, J.W. *et al. J. Neurosci.* **35**, 13323–13335 (2015).
8. Kogan, J.H., Frankland, P.W. & Silva, A.J. *Hippocampus* **10**, 47–56 (2000).
9. Heimer-McGinn, V.R., Poeta, D.L., Aghi, K., Udawatta, M. & Burwell, R.D. *J. Neurosci.* **37**, 4819–4829 (2017).
10. Templer, V.L. & Hampton, R.R. *Curr. Biol.* **23**, R801–R806 (2013).
11. Bunzeck, N. & Düzel, E. *Neuron* **51**, 369–379 (2006).
12. Chou, M.Y. *et al. Science* **352**, 87–90 (2016).
13. Albasser, M.M., Poirier, G.L. & Aggleton, J.P. *Eur. J. Neurosci.* **31**, 134–147 (2010).
14. Bourgeois, J.P. *et al. Cereb. Cortex* **22**, 1007–1015 (2012).

Is population activity more than the sum of its parts?

Jonathan W Pillow & Mikio C Aoi

A study introduces innovative ways to test whether neural population activity exhibits structure above and beyond that of its basic components.

Suppose a fancy new analysis method reveals an (apparently) surprising form of population-level organization in your large-scale neural data set. How can you tell if the observed pattern is truly surprising? Is it the hallmark of a population-level mechanism that reveals the circuit's true function, or is it merely an expected byproduct of things we already knew about neurons contained in the population? To put it bluntly: when are findings of population-level structure 'new science' and when are they merely old knowledge dressed up in new clothes? In this month's issue of *Nature Neuroscience*, Elsayed and Cunningham propose new methods for resolving this question[1].

Jonathan W. Pillow and Mikio C. Aoi are in the Princeton Neuroscience Institute and Department of Psychology, Princeton University, Princeton, New Jersey, USA.
e-mail: pillow@princeton.edu or maoi@princeton.edu

Their main contribution is to formalize the notion of primary (or already known) features of a neural population so that claims of surprising population structure can be tested against them. To make this concrete, consider, for example, the recent claim that a neural population exhibits 'rotational dynamics'[2], a contention we'll return to later. Elsayed and Cunningham show that standard shuffling methods do not, in fact, preserve the full set of primary features of a neural population; to address this problem, they introduce two methods for sampling from a properly defined null model, allowing claims of novel population-level structure to be put to the appropriate test.

The starting point for the population-level analyses in question is a collection of peristimulus time histograms, or time-varying firing rates, from multiple neurons across time and across multiple experimental conditions. We can think of these data as living in a 3D tensor (or array) with axes denoting time, neuron and condition (**Fig. 1**). Every entry in the tensor is a number indicating the firing rate of a particular neuron at a single time bin for a particular condition.

What would it mean for this dataset to contain meaningful population structure above and beyond its primary features? Elsayed and Cunningham propose that we should consider as primary the means and correlations along each side of the tensor: temporal correlations, neuronal correlations and conditional correlations. Temporal correlations reflect the fact that, before we say anything about population-level structure, neural firing rates are typically smooth in time. Neuronal correlations, the

second primary feature, reflect differences in individual neural tuning: neurons with similar tuning curves will be strongly correlated over time and conditions, whereas neurons with dissimilar tuning may exhibit zero or even negative correlation (in which one neuron's firing rate goes up whenever another's goes down). Lastly, conditional correlations reflect similarity between the patterns of population activity observed during different experimental conditions or tasks. For example, in a reaching study, two subtly different rightward movements may generate similar patterns of population activity, producing high conditional correlations; a leftward reach that recruits an entirely different set of neurons would be weakly correlated with the previous two conditions.

The null model defined by Elsayed and Cunningham (**Fig. 1**) is a distribution over data tensors that preserves nothing beyond these three sets of primary features, each described by a mean and covariance matrix (over time bins, neurons and conditions, respectively). Population-level phenomena that arise from this null model are, in the parlance of Elsayed and Cunningham, "expected byproducts" of the population's primary features.

Prior to this paper, the standard approach to testing for population-level structure involved forming a null distribution by permuting each neuron's responses across experimental conditions. Elsayed and Cunningham show that this shuffling procedure preserves only one of those three correlations (e.g., temporal, which is standard), but that, notably, it fails to preserve correlations across the other two (neurons and conditions). This failure can elevate nonsignificant effects to an appearance of significance (that is, it can create false positives) or inflate weak population-level effects so they appear more significant than they are.

Elsayed and Cunningham introduce two methods for creating the correct null distribution. The first is corrected Fisher randomization, a direct extension of conventional shuffling methods in which the reshuffled data are weighted so that all three sets of correlations are preserved. The second method is to sample from the tensor maximum entropy distribution, a distribution that preserves primary (mean and covariance) features of the original dataset but is otherwise maximally unstructured. Samples from this distribution are data tensors that exhibit the same temporal smoothness and the same pattern of marginal correlations across neurons and conditions, but that are otherwise as random as possible. The logic of Elsayed and Cunningham's proposed approach, following the standard logic of statistical hypothesis testing, is to sample datasets from the null distribution using either corrected Fisher randomization or tensor maximum entropy, apply the same population-level analysis to each sample (for example, fit it with a rotational dynamical system) and test whether the result for the original dataset (for example, the percent of variance explained) is an outlier compared to the samples from the null.

It is worth noting that Elsayed and Cunningham's derivation of the tensor maximum entropy distribution is a mathematical achievement in its own right, above and beyond the neuroscience questions considered in their paper. This result, which amounts to a particular Gaussian distribution constrained by marginal covariance matrices (not the one you would expect if you are familiar with the tensor normal distribution) has not, to the best of our knowledge, appeared previously in the statistics literature.

To focus solely on the statistical issues, however, is to miss part of the larger scientific story surrounding Elsayed and Cunningham's paper. A debate has raged for several years about claims put forth by Churchland *et al.*[2], who argued that population activity in motor cortex exhibits rotational dynamics during (nonrotational) reaching movements. This result electrified the field, but also sparked controversy over whether the findings were truly surprising. Critics grumbled (although never in print) that the finding of rotational dynamics was a trivial consequence of the neurons' smooth firing rates; in high-dimensional spaces, they argued, any set of smooth trajectories will look rotational.

The results of Elsayed and Cunningham can be viewed as a direct response to this criticism, vindicating the original claims of Churchland *et al.*[2]: the fits of a dynamical system explain significantly more variance than expected under a null model that preserves correlations across time, neurons and conditions[1]. To understand this result intuitively, consider the following: neural responses generated by a rotational dynamical system exhibit consistent phase relationships between neurons; these relationships are preserved across different reach conditions, allowing them to be well fit by a single linear dynamics matrix. Samples from the null model, by contrast, although smooth and matching covariances over neurons and conditions, have scrambled phases, since consistent phases involve relationships between rates over multiple time bins and conditions, and such relationships are beyond the purview of the null model.

Elsayed and Cunningham's paper can be seen as a natural extension of previous work seeking to quantify whether high-level features of neural activity are expected from known low-level features. Previous work on multineuron spike patterns, for example, used maximum entropy models to ask whether the observed distribution of spike patterns could be explained by the simpler set of correlations between pairs of neurons, neglecting higher

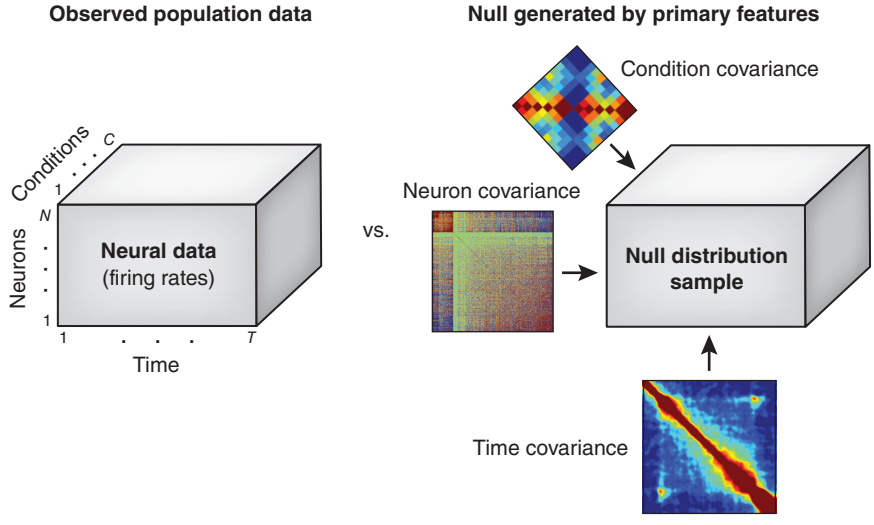

Figure 1 Schematic of testing approach for population-level structure. Left: multineuron firing rate data from multiple conditions can be arranged in a 3D tensor, with dimensions T (number of time bins) by N (number of neurons) by C (number of experimental conditions). Right: a properly designed null distribution preserves the correlations found in the original data over time, neurons and conditions, each described by a covariance matrix: a $T \times T$ temporal covariance matrix, an $N \times N$ neuronal covariance matrix and a $C \times C$ condition covariance matrix. These primary features define a null distribution that can be sampled by either corrected Fisher randomization or tensor maximum entropy; these samples can be used to rigorously test claims about population-level structure.

order interactions[3–5]. However, it bears mentioning that calling a feature an expected byproduct is not the same as calling it uninteresting or unimportant, a point that Elsayed and Cunningham also take care to emphasize. Physics is rife with examples in which lower order features are sufficient to explain fundamental large-scale phenomena such as heat, states of matter or the behavior of artificial neural networks.

Considered broadly, the recent rise of advanced imaging and electrical recording technologies has driven a broad-based demand for new analysis methods of neural population recordings. Central to this need is the desire of researchers to go beyond a purely representational view of neuronal activity and move toward a characterization of the dynamics underlying neural computations. Such demands have spurred a creative explosion of new analysis methods focused particularly on dynamical systems[2,6–8] and dimensionality reduction[9–11] for high-dimensional neural activity. Elsayed and Cunningham's approach will therefore provide a much-needed check on the overexuberant application of such methods and on claims about the patterns they reveal.

There are several directions in which the methods proposed by Elsayed and Cunningham could be extended to address richer datasets or analysis methods. First, their proposed methodology is designed for rate codes and thus does not apply to phase coding or other forms of temporal coding. Second, they use trial-averaged responses (although this is not a fundamental constraint on the method) and thus do not consider noise correlations or patterns of single-trial variability that might reflect population-level dynamics. The proposed methods are, nonetheless, applicable to a wide range of dataset types that are being generated in massive volumes, and the basic approach can be extended to incorporate constraints beyond the primary features the authors consider (for example, ref. 12). Ultimately, the ability to precisely quantify structure arising from low-level features will be crucial to settling future debates over what neural population structure is, and whether or not to believe it when you see it.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Elsayed, G.F. & Cunningham, J.P. *Nat. Neurosci.* **20**, 1310–1318 (2017).
2. Churchland, M.M. *et al. Nature* **487**, 51–56 (2012).
3. Schneidman, E., Berry, M.J., II, Segev, R. & Bialek, W. *Nature* **440**, 1007–1012 (2006).
4. Shlens, J. *et al. J. Neurosci.* **26**, 8254–8266 (2006).
5. Ohiorhenuan, I.E. *et al. Nature* **466**, 617–621 (2010).
6. Buesing, L., Macke, J. & Sahani, M. *Adv. Neural. Inf. Process. Syst.* **25,** 1691–1699 (2012).
7. Gao, Y., Archer, E.W., Paninski, L. & Cunningham, J.P. *Adv. Neural. Inf. Process. Syst.* **29,** 163–171 (2016).
8. Zhao, Y. & Park, I.M. *Neural Comput.* **29**, 1293–1316 (2017).
9. Mante, V., Sussillo, D., Shenoy, K.V. & Newsome, W.T. *Nature* **503**, 78–84 (2013).
10. Cunningham, J.P. & Yu, B.M. *Nat. Neurosci.* **17**, 1500–1509 (2014).
11. Kobak, D. *et al. eLife* **5**, e10989 (2016).
12. Loaiza-Ganem, G., Gao, Y. & Cunningham, J.P. in *International Conference on Learning Representations*. Preprint available at https://arxiv.org/abs/1701.03504 (2017).