

Highlights:

- A demonstration for how neural activity can be read out with minimal information loss
- The read-out linearly weighs neural responses
- More reliable neurons receive greater weight
- The read-out does not need to be adjusted when correlations between neurons change

Title: Linking neural responses to behavior with information-preserving population vectors.

Abstract: All systems for processing signals, both artificial and within animals, must obey fundamental statistical laws for how information can be processed. We discuss here recent results using information theory that provide a blueprint for building circuits where signals can be read-out without information loss. Many properties that are necessary to build information-preserving circuits are actually observed in real neurons, at least approximately. One such property is the logistic nonlinearity that relates inputs to neural response probability. Such nonlinearities are common in neural and intracellular networks. With this nonlinearity type, there is a linear combination of neural responses that is guaranteed to preserve Shannon information contained in the response of a neural population, no matter how many neurons it contains. This read-out measure is related to a classic quantity known as the population vector that has been quite successful in relating neural responses to animal behavior in a wide variety of cases. Nevertheless, the population vector did not withstand the scrutiny of detailed information-theoretical analyses that showed that it discards substantial amounts of information contained in the responses of a neural population. We discuss recent theoretical results showing how to modify the population vector expression to make it “information-preserving”, and what is necessary in terms of neural circuit organization to allow for lossless information transfer. Implementing these strategies within artificial systems is likely to increase their efficiency, especially for brain-machine interfaces.

Distributed computation is thought to be one of the hallmarks of the brain computation where signals are processed in parallel across many, often quite variable, neural responses. On short times, responses of individual neurons are binary, because neurons respond to stimulation by producing all-or-nothing events in the voltage traces across their membranes termed “spikes”. However, the representational capacity of neural circuits grows quickly with the number of neurons. Given N neurons, even for a short time interval where each neuron can produce at most a single spike, one can construct 2^N binary responses patterns using 1s and 0s to represent spikes and no spikes from each neuron. This quickly yields astronomical numbers of possible response patterns, even for modestly sized populations. However, to take advantage of this large response capacity, one needs to be able to take into account the identity of neurons when reading out their responses. Otherwise, the combinatorial expansion in the number of possible response patterns will be lost. Supporting

this idea, studies show that one needs to keep track of neuronal identity, in ways that go beyond simply specifying the preferred stimulus pattern for each neuron, in order to avoid substantial information loss [1**,2*]. These information-theoretic studies, however, do not specifically describe how the information can be read out.

The counter-point to this line of argument is that such combinatorial read-outs are too difficult to implement in the nervous system. Indeed, practice shows that many aspects of animal's behavior, including arm and eye movements, can be accurately predicted using a relatively simple readout of neural activity known as the population vector [3-7]. In the population vector, each neuron "votes" for its preferred stimulus by the number of spikes it produces

$$\vec{V}\{r_1, \dots, r_N\} = \sum_i \vec{w}^{(i)}(2r_i - 1), \quad (1)$$

Here, r_i is a binary variable that takes values 1 or 0 when the i th neuron produces a spike; vector $\vec{w}^{(i)}$ corresponds to the preferred stimulus or movement pattern for the i th neuron, the quantity also known as the neuron's receptive field (RF). According to the population vector expression, spikes from the i th neuron shift it towards the preferred stimulus for that neuron; the absence of spikes from the neuron shifts the population average in the opposite direction. Thus, both spikes and their absences carry information [1**,8]. The population vector has a simple construction and in many cases makes it possible to accurately predict animal behavior [3,4,6,9]. However, its key assumption -- that spikes from neurons with the same preferred stimulus can be simply added -- has been explicitly refuted by the above mentioned studies of similarly tuned neurons [1**,2*]. In both of these studies, the authors found that the population count discarded a substantial portion of information that is available in the full population response.

Recent theoretical work [10,11*] shows how this dilemma can be resolved provided neural responses are structured in a certain way. It turns out that it is possible to modify the population vector such that it retains the computational simplicity but at the same time is guaranteed not to lose information contained in the responses of a neural population, such those described in [1**,2*]. The key modification is to introduce an additional weighting factor β_i that characterizes the reliability of the i th neuron:

$$\vec{T}\{r_1, \dots, r_N\} = \sum_i \beta_i \vec{w}^{(i)}(2r_i - 1) \quad . \quad (2)$$

We will refer to $\vec{T}$ as the sufficient population vector, because it represents a “sufficient statistic” [12] for neural responses, provided certain conditions are met that are detailed below. Compared to the original population vector, the sufficient population vector $\vec{T}$ weighs neural responses by both their RFs $\vec{w}^{(i)}$ and the reliability factors β_i .

What are these reliability factors β_i and why are they sufficient to capture full information contained in neural responses compared to, for example, other parameters such as the maximum spike rate or the width of orientation tuning curves? Parameters β_i can be computed from the knowledge of the peaks of tuning curves and their widths (Figure 1), but have a simpler interpretation in the framework of the so-called linear-nonlinear (LN) model. For neurons whose preferred stimulus pattern can be described by a single parameter, such as orientation, there is one-to-one correspondence between the peak and width parameters of a tuning curve and parameters of the LN models (Figure 1). However, LN models can also be used in cases where RFs of neurons are more complex and cannot be described with just one parameter, as in the case of neurons whose tuning changes depending on the time interval before the spike. In its simplest form, the LN model describes the instantaneous spiking rate $r_i(\vec{s})$ for the i th neuron in response to input $\vec{s}$ as:

$$r_i(\vec{s}) = \frac{R_{max}}{1 + \exp(-2\beta_i (\vec{s} \cdot \vec{w}^{(i)} - \alpha_i))} \quad (3)$$

Here, R_{max} is the maximal spike rate, parameters α_i and β_i describe the threshold and steepness of the nonlinearity. The steepness parameter β_i can be incorporated into the definition of the RF, but we keep RFs normalized to unit length ($|\vec{w}^{(i)}| = 1$) for two reasons. First, with this choice, Eq. (1) corresponds to the original definition of the population vector [3-7]. Second, with this choice, parameters β_i in Eqs. (2) and (3) directly reflect neuronal reliability. The second property arises because larger β values correspond to steeper nonlinearities (Figure 1), and this in turn corresponds to fewer inputs for which the responses fluctuate between spiking and no spiking on repeated trials.

The linearity of the argument of the logistic function defined in Eq. (3) will be important for many of the information-preserving properties discussed below. This linearity effectively allows only for two independent parameters α and β to describe tuning around the preferred stimulus. In particular, this constraints the peak

and the lowest value of the tuning curve relative to each other (Fig. 1C). For different values of threshold α , increasing the steepness parameter β simultaneously decreases the lowest (background) value of the tuning curve and increases the difference between the smallest and largest value (gain). Notably, the resultant inverse correlation between the background and gain values of the tuning curve agrees with experimental observations on tuning curves in the primary visual cortex (V1) [13**]. The study [13**] also reported a lack of correlation between average spike rate across the neural population and the Fisher information. This phenomenon is also consistent with the logistic model of neural nonlinearity, at least qualitatively as one might be cautious when directly comparing results using Fisher and Shannon mutual information. In the logistic model, increase in the spike rate can result from either a smaller threshold α or from a smaller steepness value β . Smaller thresholds typically increase the Shannon mutual information, whereas smaller steepness values decrease it. Thus, similarly to Fisher information analysis [13**], within the logistic model also, increasing the overall population spike count can be accompanied by increased information from neurons whose thresholds are lowered and decreased information from neurons whose tuning becomes less steep. Thus the observed variation in neural tuning curves [13**] provides additional support for the logistic neuronal tuning model.

We now illustrate how one can capture full information contained in the responses structured according to Eq. (3) using the sufficient population vector, and why the procedure works even in cases where the original population vector does not do so. Figure 2A illustrates the case where all neurons in a population have the same RFs (and therefore preferred orientation). In this case, the original population vector reduces to a “population count”, i.e. the number of spikes produced by the population as a whole. As reported previously [1**,2*], population count losses significant amount of information in this situation. Importantly, this information loss is avoided when using the sufficient population vector. This is despite the fact that the sufficient population vector is also reduced to in this case to its scalar magnitude, which may be termed the sufficient population count. Just like the population count, the sufficient population count takes discrete values. But with the sufficient population count, neuronal identity is coded into different discrete values. For example, joint spiking from neurons 1 and 2 that have, respectively, β_1 and β_2 values, corresponds to $\beta_1 + \beta_2$ value of the sufficient population count. The response pattern where neuron 1 spikes but neuron 2 does not spike corresponds to the $\beta_1 - \beta_2$ value, whereas the pattern where neuron 1 does not spike but neuron 2 spikes maps onto the $\beta_2 - \beta_1$

value. Finally, silence from both neurons is represented by the $-\beta_1 - \beta_2$ value. Thus, four possible responses across the four neurons are represented by four different values of the magnitude of the sufficient population vector.

These results generalize to the more common case where neurons in the population have different RFs (Figure 2B). Here, both the original and sufficient population vectors are vectors of the same dimensionality. The dimensionality of these vectors also matches that of RFs and of the input space where neural RFs are defined. Both of the population vectors take a discrete set of values. But this set is such that the sufficient population vector is guaranteed to capture full information, whereas the original population vector does not always do so. For example, if there are neurons in the population that have different tuning around the same preferred stimulus, then the original population vector will lose information (Figure 2B). This loss does not happen for the sufficient population vector.

The mathematical reason that the sufficient population vector captures full information under these diverse circumstances is that when the neuronal nonlinearity is described by a logistic function, as is the case in Eq. (3), then the probability of neural responses across the population can be written in the form where the sufficient population vector $\vec{T}$ is the only term that couples neural responses $\vec{r}$ to inputs $\vec{s}$ [10,11*]:

$$P(\vec{r}|\vec{s}) = A(\vec{s})H(\vec{r})e^{-\vec{T}\cdot\vec{s}}, \quad (4)$$

where $A(\vec{s})$ and $H(\vec{r})$ are some functions of the input and neural responses, respectively. This makes $\vec{T}$ the “sufficient statistic” [12] for neural responses. Monitoring neural responses using any other measure that does not distinguish distinct values of $\vec{T}$ will lead to information loss. This includes the original population vector.

It is worth pointing out that the sufficient population vector can preserve information even in the case where the number of discrete values that it can take is less than 2^N . For example, if two neurons have the same RF and the same β_i values, then the two response patterns where just one of these neurons spikes will be mapped onto the same value of the sufficient population vector. A larger population may have many of such redundant sub-groups, and so the sufficient population vector will take much fewer number of values than the 2^N number of different response patterns that the population can produce. Despite this compression in the

representation values, the sufficient population vector continues to preserve the information in this case (Figure 2B). The original population vector continues to lose information because it also maps responses of neurons with different β_i values onto the same value. Only in the case where all neurons with the same RFs also have the same β_i values do the original and sufficient population vectors fully match, with both capturing full information (Figure 2C).

The last example is intriguing because here neurons have different tuning curves width, yet their responses can be summed without information loss. To understand why these differences between tuning curves can be ignored, we recall that there is a one-to-one correspondence between the parameters of LN models and the tuning curve widths. The differences in tuning curves in Figure 2C were due to differences in threshold values α_i not the steepness parameters β_i . While changing thresholds affects the overall amount of information conveyed by the neural population [14-21], the knowledge of thresholds is not needed for reading-out these neuronal responses without loss. These results illustrate one of the properties of the sufficient population vector, namely that it does not explicitly depend on neuronal thresholds. It depends on thresholds only indirectly, through the observed neural responses, whose probabilities are affected by threshold settings. These analyses illustrate how one can separate which aspects of neuronal tuning diversity [22*,23-24] affect the efficiency of encoding stimuli into neural responses and which need to be taken into account when reading out neural responses.

These results remain valid in the presence of certain types of correlated variability across neurons (the so-called noise correlation reviewed in [9,25]) on encoding and read-out. Depending on their structure, noise correlations can either increase or decrease the overall information provided by the neural responses [9,24,26-31]. In particular, the differential noise correlations [28*], while often small, can disproportionately limit the amount of information that the neural population can transmit. However, it turns out that one can continue to capture full information – whether it is increased or decreased by noise correlations – with the same expression for the sufficient population vector (Fig. 2D). This result holds true provided the noise correlations do not change with the stimulus (although differences between neuronal pairs are permissible). The main technical requirement is that the probability of responses across the population can still be written in a form similar to Eq. (4) [32]:

$$P(\vec{r}|\vec{s}) = A_j(\vec{s})H(\vec{r})e^{-\vec{r}\cdot\vec{s}+\sum J_{ij}r_i r_j}. \quad (4)$$

where $A_j(\vec{s})$ is now a different function of the stimulus that depends on the noise correlation matrix J_{ij} . In Figure 2D we show the results of model simulations with noise correlations that vary between neuronal pairs according to how close are their preferred input orientations, $J_{ij} = \frac{1+\cos(\varphi_i-\varphi_j)}{10\sqrt{N}}$. These correlations approximate differential correlations for similarly tuned neurons [28]. We note that in the presence of correlation, tuning curves can have complicated shapes but information preservation is still possible.

For the sufficient population vector to be valid, the nonlinearity in the LN model has to be a logistic function [11]. This type of nonlinearity provides good approximation for many types of neural response [13,22*,33,34], but of course not all. One therefore might be curious how the sufficient population works with real neurons responses. Figure 3A shows the results for neurons from the primary visual cortex (V1) whose responses were recorded when presenting natural stimuli [35]. We first estimated for each neuron its preferred orientation and nonlinearity by finding the first maximally informative dimension (MID) [36] for relating the presented set of natural stimuli to the observed responses of that neuron. The steepness parameters β (and thresholds α) were estimated by fitting the nonlinearity of the MID model using logistic regression. The MID features themselves were used to estimate the preferred orientation for each neurons. The corresponding orientation values were consistent with values estimated directly using moving gratings [37]. To account for experimental uncertainties in the orientation values and to ensure that population vectors (either the standard or the information-preserving) were not assigned distinct value due to differences in preferred orientation values that were smaller than their error-bars, we used a coarse-grained set of orientations. Specifically, starting with the set of orientation values estimated from the MIDs, we computed the d' values for each pair of orientation values and then merged them in a greedy, pairwise manner starting with the pair of values with the smallest d' value. Once the selectivity of each neuron is represented by a preferred orientation (plus the tuning around it as described by parameters α and β), the population vectors, both the standard and the sufficient, become two-dimensional vectors (Figure 3B,C). One can then compare the full amount of information provided by the neural responses with the information provided by the standard and sufficient population vectors. One finds that just like in the model neural populations, the sufficient population vector (but not the standard

population vector) captured all the information (Figure 3A) provided by the responses of simultaneously recorded neurons.

It is worth pointing out that the Bernoulli noise model considered here corresponds to a short-time interval approximation to commonly used probabilistic model of neural population responses based on a set of conditionally independent Poisson neurons [32]:

$$P(\vec{r}|\vec{s}) = \prod_i \frac{e^{-f_i(\vec{s})}}{r_i!} f_i(\vec{s})^{r_i} = \prod_i \frac{e^{-f_i(\vec{s}) + r_i \log[f_i(\vec{s})]}}{r_i!}, \quad (5)$$

where r_i is generalized to being a nonnegative number and the function $f_i(\vec{s})$ is the tuning curve of neuron i , such that the average number of spikes produced by neuron i to stimulus s is $f_i(s)$. Although Eq. (5) still describes an exponential family, as is the case for the logistic nonlinearity model, for general choices of the set of tuning curves $\{f_i(\vec{s})\}$, there is no lower dimensional sufficient statistic for the response $\vec{r}$ other than $\vec{r}$ itself. However, if the tuning curves of the Poisson model can be described as log affine in $\vec{s}$ for all neurons:

$$f_i(\vec{s}) = \exp[-\beta_i (\vec{s} \cdot \vec{w}^{(i)} - \alpha_i)], \quad (6)$$

Then the sufficient population vector we defined will also be a sufficient statistic for this family. For short time bins where r_i is effectively binary, Poisson models can be approximated as binary spiking probability using just the first two terms of the Poisson distribution: $P(r_i | \vec{s}) \propto f_i(\vec{s})^{r_i}$, and $r_i = 0, 1$. After normalization, and taking into account (6), one obtains the original logistic spiking model:

$$P(r_i = 1|\vec{s}) = \frac{1}{1 + \exp(-\beta_i (\vec{s} \cdot \vec{w}^{(i)} - \alpha_i))}, \quad (7)$$

as a short time approximation of a Poisson model. In Eq. (7) the probability is evaluated over bins $\Delta t = \frac{1}{R_{max}}$, which corresponds to spike rate provided in Eq. (3).

Overall, these analyses simultaneously explain why the original population vector has been so successful in linking neural activity to behavior [3,4,6,9] and its deficiencies revealed by detailed analyses [1**,2*]. The population vector has a similar enough expression to the sufficient population vector to work well under many circumstances. At the same detailed analyses show that its expression is not general enough to

avoid information loss. The empirical linear decoder [22*] addressed this issue by adjusting the weights to increase the accuracy of the decoding. The sufficient population vector of Eq. (2) shows what the weights should be using first principles of information theory, obviating the need for optimization. The fact that the read-out remains a linear function of neural responses should make for an easy implementation within brain-machine interfaces. At the same time it is important to point out the nonlinear relationship between the input vector and the sufficient population vector. Because of limitations on the neural spike rates, the sufficient population vector typically represents, on average, a compressed function of the input vector, making this decoder different from the optimal linear decoder proposed previously [6].

Further improvements may come from extending the population vector analysis to cases where individual neurons are sensitive to multiple input components. With successive stages of neural processing, neural responses acquire more complex dependencies on input than is afforded by the LN model from Eq. (3). Cortical neurons in particular often exhibit sensitivity to multiple input components [38-41], some of which excite and/or suppress the neural responses. Extending the population vector approaches to apply to these more complex neural responses is likely to be key for effective decoding from cortical neurons.

Figure 1. The relationship between receptive field (RF) and tuning curve descriptions of the neural response. (A) Three representative model nonlinearities that describe neural response as a saturating function of stimulus component along RF. The black and cyan lines have the same threshold (midpoints) α but different steepness values β . The magenta lines has the same steepness value as the black lines but a different threshold value. Inset shows an example orientation selective RF. (B) Corresponding tuning curves from (A) but as a function of stimulus angle. Thin horizontal lines at the top of the tuning curves illustrate the width of these tuning curves computed as the inverse of the square root of the second derivative of the log of the tuning curve at the peak. Equations show the expressions for the tuning curve width and peak as a function of the LN model parameters. Both the threshold and the steepness value of the LN model affect them. (C) Inverse relationship between the background (lowest) value of the tuning curve and the tuning curve modulation (the difference between the peak and the lowest values of the tuning curve). Each curve corresponds to a different value of the threshold, with steepness values varying between 0 and 2 along each curve.

Figure 2. The sufficient vector captures all information from diverse neural populations and with correlated variability across neurons. In all panels, we compare information transmitted by a population response (black line) with information transmitted by the sufficient population vector (red) and the standard population vector (dashed gray). Neural populations tuned to the same (**A**, **C**) or different (**B**, **D**) preferred stimuli. In (**A**) differences in neural tuning curves are due to differences in steepness values, whereas in (**C**) they are due to differences in thresholds. Dotted lines show the information values obtained by binning response variables. Dotted lines overlap with solid curves. (**D**) same as (**B**) but with noise correlations. Insets show example population tuning curves.

Figure 3. The sufficient population vector captures all the information provided by responses of simultaneously recorded V1 neurons. (**A**) Curves for the information provided by neural responses (black line) and that captured by the sufficient population vector (red circles) overlap. The standard population vector (grey) and population count (blue) transmit less information. Error bars are standard deviations. The values for sufficient (**B**) and standard (**C**) population vector for $n=7$ neuronal population, along a subset of different trials of natural scenes. Points of the same color show vector values on the repeated presentations of the same stimulus. Different colors correspond to different stimuli. Crosses show the mean value of the sufficient population vector for each stimulus. Note that there are many overlapping points at (0,0) corresponding to the absence of spiking. Compared to the sufficient population vector, there is more spread in the values of the original population vector across repeats of the same stimuli.

Acknowledgments: This research was supported by the Rose Hill Foundation (JB), and by the following grants to TS from the National Science Foundation: Career Award IIS-1254123, CRCNS IIS-1724421, and Ideas Lab IOS-1556388.

References.

- 1**. Osborne LC, Palmer SE, Lisberger SG, Bialek W: **The neural basis for combinatorial coding in a cortical population response.** *J Neurosci* 2008, **28**:13522-13531.
This paper demonstrates that it is important to keep track of neuronal identities even among neurons that have the same preferred stimuli and that tuning width around the preferred stimulus plays an important role. The expression for the sufficient population vector provides a specific expression for how differences in tuning width should be taken into account.
- 2*. Reich DS, Mechler F, Victor JD: **Independent and redundant information in nearby cortical neurons.** *Science* 2001, **294**:2566-2568.
This paper also demonstrates that not keeping track of neuronal identities among similarly tuned neurons leads to significant information loss.
- 3*. Georgopoulos AP, Schwartz AB, Kettner RE: **Neuronal population coding of movement direction.** *Science* 1986, **233**:1416-1419.
This is a classic paper introducing the population vector as a way of linking neural activity to behavior.
4. Hohl SS, Chaisanguanthum KS, Lisberger SG: **Sensory population decoding for visually guided movements.** *Neuron* 2013, **79**:167-179.
5. Lewis JE, Kristan WB, Jr.: **A neuronal network for computing population vectors in the leech.** *Nature* 1998, **391**:76-79.
6. Salinas E, Abbott LF: **Vector reconstruction from firing rates.** *J. Comput. Neurosci.* 1994, **1**:89-107.
7. Seelig JD, Jayaraman V: **Neural dynamics for landmark orientation and angular path integration.** *Nature* 2015, **521**:186-191.
8. Schneidman E, Puchalla JL, Segev R, Harris RA, Bialek W, Berry MJ, 2nd: **Synergy from silence in a combinatorial neural code.** *J Neurosci* 2011, **31**:15732-15741.
9. Shamir M: **Emerging principles of population coding: in search for the neural code.** *Curr Opin Neurobiol* 2014, **25**:140-148.
10. Berkowitz JA, Sharpee TO: **Decoding neural responses with minimal information loss.** *Bioarxiv* 2018:<https://doi.org/10.1101/273854>
- 11*. Berkowitz JA, Sharpee TO: **Quantifying information conveyed by large neural populations.** *Neural Comp.* 2019.
This study derives the information-preserving population vector.
12. Cover TM, Thomas JA: *Elements of Information Theory.* New York: Wiley-Interscience; 1991.
- 13**. 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.
This study reports inverse correlation between multiplicative and additive modulation of V1 tuning curves. This observation is consistent with the properties of logistic nonlinearity from Eq. (3) when evaluated in terms of tuning curves.
14. Bethge M, Rotermund D, Pawelzik K: **Optimal neural rate coding leads to bimodal firing rate distributions.** *Network* 2003, **14**:303-319.
15. Nikitin AP, Stocks NG, Morse RP, McDonnell MD: **Neural population coding is optimized by discrete tuning curves.** *Phys Rev Lett* 2009, **103**:138101.
16. McDonnell MD, Stocks NG, Pearce CEM, Abbott D: **Point singularities and suprathreshold stochastic resonance in optimal coding.** *arXiv:cond-mat/0409528* 2005.
17. McDonnell MD, G. SN, Pearce CEM, Abbott D: **Optimal information transmission in nonlinear arrays through suprathreshold stochastic resonance.** *Physics Letters* 2006, **352**:183-189.
18. Sharpee TO: **Optimizing Neural Information Capacity through Discretization.** *Neuron* 2017, **94**:954-960.
19. Kastner DB, Baccus SA, Sharpee TO: **Critical and maximally informative encoding between neural populations in the retina.** *Proc Natl Acad Sci U S A* 2015, **112**:2533-2538.
20. Gjorgjieva J, Sompolinsky H, Meister M: **Benefits of pathway splitting in sensory coding.** *J Neurosci* 2014, **34**:12127-12144.
21. Brinkman BA, Weber AI, Rieke F, Shea-Brown E: **How Do Efficient Coding Strategies Depend on Origins of Noise in Neural Circuits?** *PLoS Comput Biol* 2016, **12**:e1005150.
- 22*. Graf AB, Kohn A, Jazayeri M, Movshon JA: **Decoding the activity of neuronal populations in macaque primary visual cortex.** *Nat Neurosci* 2011, **14**:239-245.

This study demonstrates that adjusting the weights of a linear decoder can improve its accuracy.

23. Chelaru MI, Dragoi V: **Efficient coding in heterogeneous neuronal populations.** *Proc Natl Acad Sci U S A* 2008, **105**:16344-16349.
 24. Shamir M, Sompolinsky H: **Implications of neuronal diversity on population coding.** *Neural Comput* 2006, **18**:1951-1986.
 25. Averbeck BB, Latham PE, Pouget A: **Neural correlations, population coding and computation.** *Nat Rev Neurosci* 2006, **7**:358-366.
 26. Abbott LF, Dayan P: **The effect of correlated variability on the accuracy of a population code.** *Neural Comput* 1999, **11**:91-101.
 27. Ecker AS, Berens P, Tolias AS, Bethge M: **The effect of noise correlations in populations of diversely tuned neurons.** *J Neurosci* 2011, **31**:14272-14283.
 - 28*. Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A: **Information-limiting correlations.** *Nat Neurosci* 2014, **17**:1410-1417.
- This study points out that differential noise correlations disproportionately limit Fisher information that neural responses provide about stimuli.
29. Shamir M, Sompolinsky H: **Nonlinear population codes.** *Neural Comput* 2004, **16**:1105-1136.
 30. Zohary E, Shadlen MN, Newsome WT: **Correlated neuronal discharge rate and its implications for psychophysical performance.** *Nature* 1994, **370**:140-143.
 31. Fitzgerald JD, Sharpee TO: **Maximally informative pairwise interactions in networks.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2009, **80**:031914.
 32. Ma WJ, Beck JM, Latham PE, Pouget A: **Bayesian inference with probabilistic population codes.** *Nat Neurosci* 2006, **9**:1432-1438.
 33. Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, Newsome WT, Clark AM, Hosseini P, Scott BB, et al.: **Stimulus onset quenches neural variability: a widespread cortical phenomenon.** *Nat Neurosci* 2010, **13**:369-378.
 34. Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, Tolias AS: **A fast and simple population code for orientation in primate V1.** *J Neurosci* 2012, **32**:10618-10626.
 35. Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD: **Adaptive filtering enhances information transmission in visual cortex.** *Nature* 2006, **439**:936-942.
 36. Sharpee T, Rust NC, Bialek W: **Analyzing neural responses to natural signals: maximally informative dimensions.** *Neural Comput* 2004, **16**:223-250.
 37. Sharpee TO, Miller KD, Stryker MP: **On the importance of static nonlinearity in estimating spatiotemporal neural filters with natural stimuli.** *J Neurophysiol* 2008, **99**:2496-2509.
 38. Atencio CA, Sharpee TO, Schreiner CE: **Receptive field dimensionality increases from the auditory midbrain to cortex.** *J Neurophysiol* 2012, **107**:2594-2603.
 39. Atencio CA, Sharpee TO: **Multidimensional receptive field processing by cat primary auditory cortical neurons.** *Neuroscience* 2017, **359**:130-141.
 40. Sharpee TO, Atencio CA, Schreiner CE: **Hierarchical representations in the auditory cortex.** *Curr Opin Neurobiol* 2011, **21**:761-767.
 41. Rust NC, Schwartz O, Movshon JA, Simoncelli EP: **Spatiotemporal elements of macaque V1 receptive fields.** *Neuron* 2005, **46**:945-956.





