bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Bridging neuronal correlations and dimensionality reduction

Akash Umakantha[1,2,*], Rudina Morina[3,*], Benjamin R. Cowley[2,4,*], Adam C. Snyder[3,5,6,7], Matthew A. Smith[1,8,†] and Byron M. Yu[1,3,8,†]

[1]Carnegie Mellon Neuroscience Institute, Pittsburgh, PA
[2]Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA
[3]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA
[4]Princeton Neuroscience Institute, Princeton University, Princeton, NJ
[5]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY
[6]Department of Neuroscience, University of Rochester, Rochester, NY
[7]Center for Visual Science, University of Rochester, Rochester, NY
[8]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA
[*,†]Denote equal contributions

Abstract

Two commonly used approaches to study interactions among neurons are spike count correlation, which describes pairs of neurons, and dimensionality reduction, applied to a population of neurons. While both approaches have been used to study trial-to-trial correlated neuronal variability, they are often used in isolation and have not been directly related. We first established concrete mathematical and empirical relationships between pairwise correlation and metrics of population-wide covariability based on dimensionality reduction. Applying these insights to macaque V4 population recordings, we found that the previously reported decrease in mean pairwise correlation associated with attention stemmed from three distinct changes in population-wide covariability. Overall, our work builds the intuition and formalism to bridge between pairwise correlation and population-wide covariability and presents a cautionary tale about the inferences one can make about population activity by using a single statistic, whether it be mean pairwise correlation or dimensionality.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

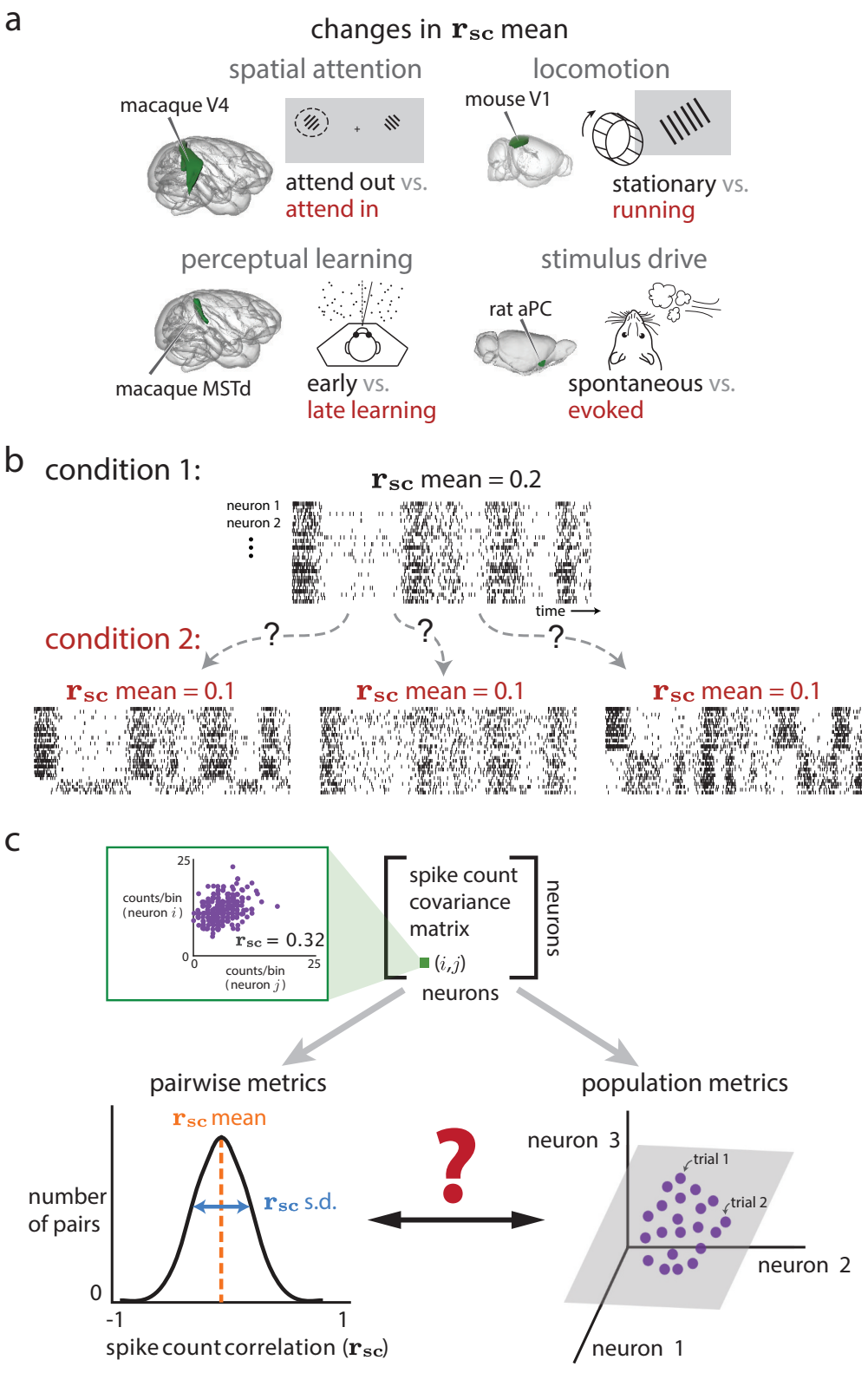
Introduction

A neuron can respond differently to repeated presentations of the same stimulus. These variable responses are often correlated across pairs of neurons from trial to trial, measured using spike count correlations (r_{sc}, also referred to as noise correlation; Cohen and Kohn, 2011). Studies have reported changes in spike count correlation across various experimental manipulations and cognitive phenomena, including attention (Cohen and Maunsell, 2009; Mitchell et al., 2009; Herrero et al., 2013; Gregoriou et al., 2014; Ruff and Cohen, 2014b; Snyder et al., 2018), learning (Gu et al., 2011; Jeanne et al., 2013; Ni et al., 2018), task difficulty (Ruff and Cohen, 2014a), locomotion (Erisken et al., 2014), stimulus drive (Maynard et al., 1999; Kohn and Smith, 2005; Smith and Kohn, 2008; Miura et al., 2012; Ponce-Alvarez et al., 2013; Ruff and Cohen, 2016b), decision-making (Nienborg et al., 2012), task context (Bondy et al., 2018), anesthesia (Ecker et al., 2010), adaptation (Adibi et al., 2013), and more (Fig. 1*a*). Spike count correlation also depends on timescales of activity (Bair et al., 2001; Kohn and Smith, 2005; Smith and Kohn, 2008; Mitchell et al., 2009; Runyan et al., 2017), neuromodulation (Herrero et al., 2013; Minces et al., 2017), and properties of the neurons themselves, including their physical distance from one another (Lee et al., 1998; Smith and Kohn, 2008; Smith and Sommer, 2013; Ecker et al., 2014; Solomon et al., 2015; Rosenbaum et al., 2017), tuning preferences (Lee et al., 1998; Romo et al., 2003; Kohn and Smith, 2005; Huang and Lisberger, 2009), and neuron type (Qi and Constantinidis, 2012; Snyder et al., 2016). Theoretical work has posited that changes in correlations affect neuronal computations and sensory information coding (Zohary et al., 1994; Shadlen and Newsome, 1998; Abbott and Dayan, 1999; Averbeck et al., 2006; Moreno-Bote et al., 2014; Sharpee and Berkowitz, 2019; Rumyantsev et al., 2020; Bartolo et al., 2020). Given such widespread empirical observations and theoretical insight, spike count correlation has been and remains instrumental in our current understanding of how neurons interact.

Most studies compute the average spike count correlation over pairs of recorded neurons for different experimental conditions, periods of time, neuron types, etc. A decrease in this mean correlation is commonly attributed to a reduction in the size (or gain) of shared co-fluctuations (Shadlen and Newsome, 1998; Rabinowitz et al., 2015; Lin et al., 2015; Ecker et al., 2016; Huang et al., 2019; Ruff et al., 2019b), e.g., a decrease in the strength of "common input" that drives each neuron in the population. However, other distinct changes at the level of the entire neuronal population can manifest as the same decrease in mean pairwise correlation (Fig. 1*b*). For example, a common input that drives the activity of all neurons up and down together could be altered to drive some neurons up and other neurons down. Alternatively, that first common input signal might remain the same, but a second input could be introduced that drives some neurons up and others down. It is difficult to differentiate these distinct possibilities using a single summary statistic, such as mean spike count correlation.

Distinguishing among these changes to the population-wide covariability might be possible by considering additional statistics that measure how the entire population of neurons co-fluctuates together. In particular, one may use dimensionality reduction to compute statistics that characterize multiple distinct features of population-wide covariability (Cunningham and Yu, 2014). Dimensionality reduction has been used to investigate decision-making (Harvey et al., 2012; Mante et al., 2013; Kiani et al., 2014; Kaufman et al., 2015), motor control (Churchland et al., 2012; Gallego et al., 2017), learning (Sadtler et al., 2014; Ni et al., 2018; Vyas et al., 2018), sensory coding (Mazor and Laurent, 2005; Pang et al., 2016), spatial attention (Cohen and Maunsell, 2010; Rabinowitz et al., 2015; Snyder et al., 2018; Huang et al., 2019), interactions between brain areas (Perich et al., 2018; Ruff and Cohen, 2019a; Ames and Churchland, 2019; Semedo et al., 2019; Veuthey et al., 2020), and network models (Williamson et al., 2016; Mazzucato et al., 2016; Recanatesi et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1 ***(previous page)*****: How do spike count correlations between pairs of neurons (i.e., pairwise metrics) relate to how the entire population co-fluctuates (i.e., population metrics)?** **a**. Four example experiments in which mean spike count correlation (r_{sc} mean) has been observed to change between experimental conditions. These include spatial attention (macaque visual area V4; Cohen and Maunsell, 2009; Mitchell et al., 2009; Gregoriou et al., 2014; Luo and Maunsell, 2015; Snyder et al., 2018), perceptual learning (macaque dorsal medial superior temporal area; Gu et al., 2011), locomotion (mouse visual area V1; Erisken et al., 2014), and stimulus drive (rat anterior piriform cortex; Miura et al., 2012). **b**. The same change in r_{sc} mean (from 0.2 to 0.1 between conditions 1 and 2) could correspond to multiple distinct changes in the activity of the population of neurons. Condition 2, left: a decrease in r_{sc} mean could correspond to some neurons becoming anti-correlated with others in the population; in this case, some neurons that were previously positively correlated are now anti-correlated with the rest of the population (bottom rows of raster plot). Condition 2, middle: a decrease in r_{sc} mean could correspond to a decrease in how strongly neurons co-fluctuate together; in this case, neurons covary as in condition 1 but each neuron does not co-fluctuate with other neurons as strongly. Condition 2, right: a decrease in r_{sc} mean could correspond to the introduction of another 'mode' of covariation (i.e., an increase in the dimensionality of population activity); in this case, neurons in the top half of the raster covary as in condition 1, but neurons in the bottom half of the raster covary in a manner independent from those in the top half. **c**. Pairwise (r_{sc}) and population (dimensionality reduction) metrics both arise from the same spike count covariance matrix, but the precise relationship between these two sets of metrics remains unknown. Top row: Each element of the spike count covariance matrix corresponds to the covariance across responses to repeated presentations of the same stimulus for two simultaneously-recorded neurons (e.g., neurons i and j, left inset). Bottom row: Pairwise metrics (left) typically summarize the distribution of spike count correlation with the mean (r_{sc} mean); in this work, we propose additionally reporting the standard deviation (r_{sc} s.d.). Population metrics (right) of the spike count covariance matrix are identified by applying dimensionality reduction to the population activity (e.g., gray plane depicts a low-dimensional space describing how neurons covary). By understanding the relationship between pairwise and population metrics, we can better interpret how changes in pairwise statistics (e.g., experiments in **a**) correspond to changes in population metrics, and vice-versa.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2019), among others. As with mean spike count correlation, the statistics computed from dimensionality reduction can also change with attention (Rabinowitz et al., 2015; Huang et al., 2019), stimulus drive (Churchland et al., 2010; Cowley et al., 2016; Snyder et al., 2018), motor output (Gallego et al., 2018), learning (Athalye et al., 2017), and anesthesia (Ecker et al., 2014). However, unlike mean spike count correlation (henceforth referred to as a "pairwise metric") which averages across pairs of neurons, the statistics computed from dimensionality reduction (henceforth referred to as "population metrics") consider the structure of population-wide covariability (Fig. 1*c*). An example of a commonly reported population metric is dimensionality (Yu et al., 2009; Rabinowitz et al., 2015; Cowley et al., 2016; Williamson et al., 2016; Mazzucato et al., 2016; Gao and Ganguli, 2015; Gallego et al., 2017; Stringer et al., 2019a; Recanatesi et al., 2019). Dimensionality is used to assess whether the number of population co-fluctuation patterns (possibly reflecting the number of common inputs) changes across experimental conditions (Fig. 1*b*, condition 1 versus condition 2, right panel). Thus, population metrics could help to distinguish among the distinct ways in which population-wide covariability can change, especially those that lead to the same change in mean spike count correlation (Fig. 1*b*).

Both pairwise and population metrics aim to characterize how neurons covary, and both can be computed from the *same* spike count covariance matrix (Fig. 1*c*). Still, studies rarely report both, and the relationship between the two is not known. In this study, we establish the relationship between pairwise metrics and population metrics both analytically and empirically using simulations. We find that changes in mean spike count correlation could correspond to several distinct changes in population metrics including: 1) the strength of shared variability (e.g., the strength of a common input), 2) whether neurons co-fluctuate together or in opposition (e.g., how similarly a common input drives each neuron in the population), or 3) the dimensionality (e.g., the number of common inputs). Furthermore, we show that a rarely-reported statistic–the standard deviation of spike count correlation–provides complementary information to the mean spike count correlation about how a population of neurons co-fluctuates. Applying this understanding to recordings in area V4 of macaque visual cortex, we found that the previously-reported decrease in mean spike count correlation with attention stems from multiple distinct changes in population-wide covariability. Overall, our results demonstrate that common ground exists between the literatures of spike count correlation and dimensionality reduction and provides a cautionary tale for attempting to draw conclusions about how a population of neurons covaries using one, or a small number of, statistics. Our framework builds the intuition and formalism to navigate between the two approaches, allowing for a more interpretable and richer description of the interactions among neurons.

Results

Defining pairwise and population metrics

We first define the metrics that we will use to summarize 1) the distribution of spike count correlations (i.e., pairwise metrics) and 2) dimensionality reduction of a population covariance matrix (i.e., population metrics). For pairwise metrics, we consider the mean and standard deviation (s.d.) of r_{sc} across all pairs of neurons, which summarize the r_{sc} distribution (Fig. 1*c*, bottom left panel). For population metrics, we consider loading similarity, percent shared variance (abbreviated to %sv), and dimensionality (described below and in more detail in Methods). These metrics each describe some aspect of population-wide covariability and thus represent natural, multivariate extensions of r_{sc}.

To illustrate these three population metrics, consider the activity of a population of neurons over time (Fig. 2*a*, spike rasters). If the activity of all neurons goes up and down together, we would find

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

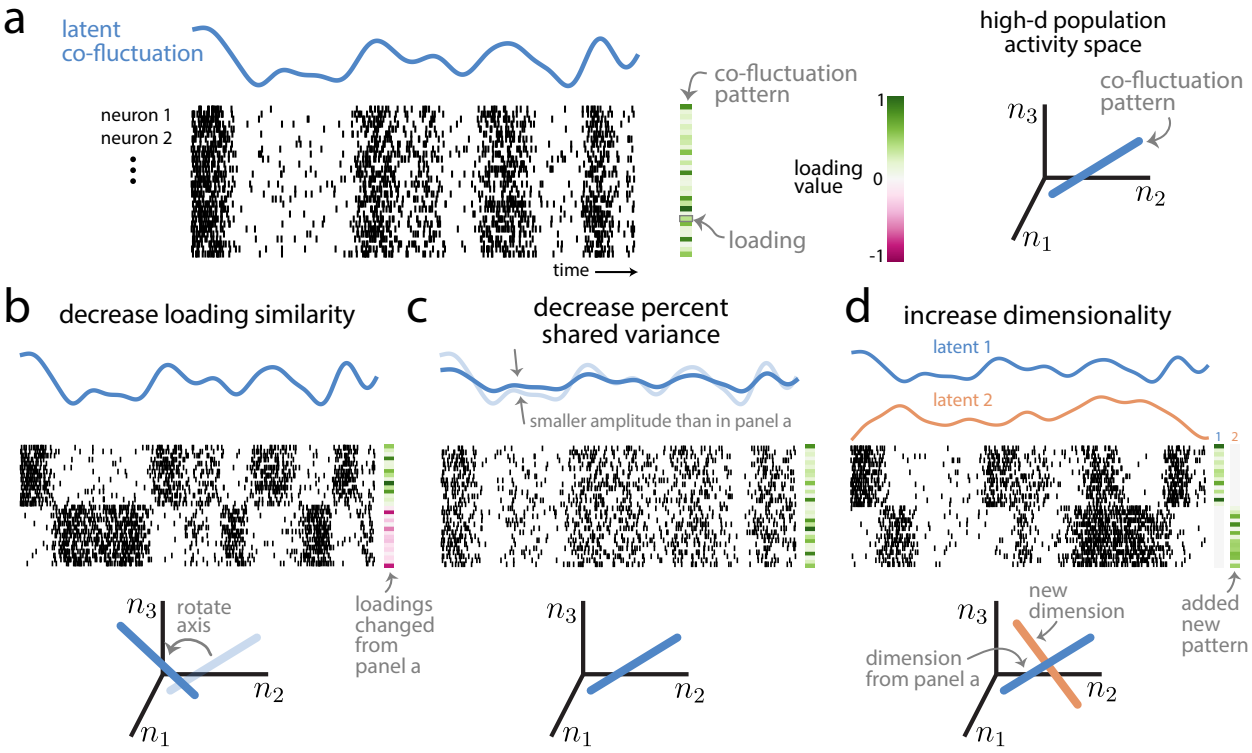


Figure 2: Intuition about population metrics: loading similarity, percent shared variance (%sv), and dimensionality. a. Population activity (population raster, where each row is the spike train for one neuron over time) is characterized by a latent co-fluctuation (blue) and a co-fluctuation pattern made up of loadings (green rectangles). Each neuron's underlying firing rate is a product of the latent and that neuron's loading (which may either be positive or negative). One may also view population activity through the lens of the population activity space (right plot), where each axis represents the activity of one neuron (n_1, n_2, n_3 represent neuron 1, neuron 2, and neuron 3). In this space, a co-fluctuation pattern corresponds to an axis whose orientation depends on the pattern's loadings (right plot, blue line). **b**. Population activity with a lower loading similarity than in panel **a**. The loadings have both positive and negative values (i.e., dissimilar loadings), leading to neurons that are anti-correlated (compare top rows with bottom rows of population raster). Changing the loading similarity will rotate a pattern's axis in the population activity space (bottom plot, 'rotate axis'). **c**. Population activity with a lower %sv than in panel **a**. The latent co-fluctuation shows smaller amplitude changes over time than in panel **a**, which leads to a lower %sv. Changing %sv leads to no changes of the co-fluctuation pattern (bottom plot, axis is same as that in panel **a**). **d**. Population activity with a dimensionality of 2, compared to a dimensinality of 1 in panel **a**. Adding a new dimension leads to a new latent co-fluctuation (orange line) and a new co-fluctuation pattern ('new dimension'). Each neuron's underlying firing rate is expressed as a weighted combination of the latents, where the weights correspond the neuron's loadings in each co-fluctuation pattern. Here, each dimension corresponds to a distinct subset of neurons (top rows vs. bottom rows); in general, this need not be the case, as each neuron typically has nonzero weights for both dimensions. In the population activity space (bottom plot), the activity varies along the two axes (i.e., a 2-d plane) defined by the two co-fluctuation patterns.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the pairwise spike count correlations between all pairs of neurons to be positive. A more succinct way to characterize this population activity is to identify a single time-varying *latent co-fluctuation* that is shared by all neurons (Fig. 2*a*, blue line). The way in which neurons are coupled to this latent co-fluctuation is indicated by a *loading* for each neuron. In this example, because the latent co-fluctuation describes each neuron's activity going up and down together, the loadings have the same sign (Fig. 2*a*, green rectangles). We refer to the latent co-fluctuation's corresponding set of loadings as a *co-fluctuation pattern*. A co-fluctuation pattern can be represented as a direction in the population activity space, where each coordinate axis corresponds to the activity of one neuron (Fig. 2*a*, right panel).

The first population metric is *loading similarity*, a value between 0 and 1 that describes to what extent the loadings differ across neurons within a co-fluctuation pattern. A loading similarity close to 1 indicates that the loadings have the same sign and are of similar magnitude (Fig. 2*a*, green rectangles). A loading similarity close to 0 indicates that many of the loadings differ, either in magnitude, sign, or both (Fig. 2*b*, green and pink squares). In this case, some neurons may have positive loadings and co-fluctuate in the same direction as the latent co-fluctuation (Fig. 2*b*, top rows of neurons show high firing rates when blue line is high and low firing rates when blue line is low), whereas other neurons may have negative loadings and co-fluctuate in the opposite direction as the latent co-fluctuation (Fig. 2*b*, bottom rows of neurons show low firing rates when blue line is high and high firing rates when blue line is low). One can view changing the loading similarity as rotating the direction of a co-fluctuation pattern in population activity space (Fig. 2*b*, bottom plot).

The second population metric is *percent shared variance* or %sv, which measures the percentage of spike count variance explained by the latent co-fluctuation. This percentage is computed per neuron, then averaged across all neurons in the population (Williamson et al., 2016). A %sv close to 100% indicates that the activity of each neuron is tightly coupled to the latent co-fluctuation, with a small portion of variance that is independent to each neuron (Fig. 2*a*). A %sv close to 0% indicates that neurons fluctuate almost independently of each other and their activity weakly adheres to the time course of the latent co-fluctuation (Fig. 2*c*). By changing %sv, one does not change the co-fluctuation pattern in population activity space (Fig. 2, blue lines are the same in panels *a* and *c*) but rather the strength of the latent co-fluctuation (Fig. 2*c*, blue line has smaller amplitude than in panel *a*).

The third population metric is *dimensionality*. We define dimensionality as the number of co-fluctuation patterns (or dimensions) needed to explain the shared variability among neurons (see Methods). The variable activity of neurons may depend on multiple common inputs, e.g., top-down signals like attention and arousal (Rabinowitz et al., 2015; Cowley et al., 2020) or spontaneous and uninstructed behaviors (Stringer et al., 2019b; Musall et al., 2019). Furthermore, these common inputs may differ in how they modulate neurons. This may result in two or more dimensions of the population activity (Fig. 2*d*, blue and orange latent co-fluctuations). For illustrative purposes, each dimension might correspond to a single group of tightly-coupled neurons (Fig. 2*d*, neurons in top rows have non-zero loadings for pattern 1, whereas neurons in bottom rows have non-zero loadings for pattern 2). However, in general, each neuron can have non-zero loadings for multiple patterns. In population activity space, adding a new dimension adds a new axis along which neurons covary (Fig. 2*d*, orange line). We use the term *dimension* to refer either to a latent co-fluctuation or its corresponding co-fluctuation pattern, depending on context.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Varying population metrics to assess changes in pairwise metrics.

Given that both pairwise and population metrics are computed from the same spike count covariance matrix (Fig. 1*c*), a connection should exist between the two. We establish this connection by deriving mathematical relationships and carrying out simulations. In simulations, we assessed how systematically changing one of the population metrics (e.g., increasing loading similarity, Fig. 3*a*), changes the spike count covariance matrix (Fig. 3*b*), and the corresponding r_{sc} distribution (Fig. 3*c*), which we summarized using its mean and standard deviation (Fig. 3*d*). The covariance matrix was parameterized in a way that allowed us to create covariance matrices with specified population metrics (see Methods). Thus, our simulation procedure does not simulate neuronal activity, but rather creates covariance matrices which are consistent with the specified population metrics.

Loading similarity has opposing effects on r_{sc} mean and s.d.

We first asked how the loading similarity of a single co-fluctuation pattern (i.e., one dimension) affected r_{sc} mean and s.d. Intuitively, a high loading similarity indicates that the activity of all neurons increases and decreases together (Fig. 2*a*), resulting in values of r_{sc} that are all positive and similar in value. Indeed, in simulations, we found that high loading similarity corresponded to large r_{sc} mean and r_{sc} s.d. close to 0 (Fig. 3*e*, green dots near horizontal axis). On the other hand, a low loading similarity indicates that when some neurons increase their activity, others decrease their activity (Fig. 2*b*), resulting in some positive r_{sc} values (for pairs that change their activity in the same direction) and some negative r_{sc} values (for pairs that change their activity in opposition). In simulations, a low loading similarity indeed corresponded to an r_{sc} mean close to 0 and a large r_{sc} s.d. (Fig. 3*e*, blue dots near vertical axis). By varying the loading similarity, we surprisingly observed an arc-like trend in the r_{sc} mean versus r_{sc} s.d. plot (Fig. 3*e*). In Supplementary Math Note A, we derive the analytical relationship between loading similarity and r_{sc}. In Supplementary Math Note B, we show mathematically why the r_{sc} mean versus r_{sc} s.d. relationship follows a circular arc.

Decreasing %sv reduces r_{sc} mean and s.d.

We next asked how %sv, which measures the percentage of each neuron's variance that is shared with other neurons in the population, is related to r_{sc} mean and s.d. Intuitively, one might expect %sv and r_{sc} mean to be closely related because r_{sc} measures the degree to which the activity of two neurons is shared (Cohen and Kohn, 2011). We investigated this in simulations and found that how closely %sv and r_{sc} mean were related depended on the loading similarity. When loading similarity was high (Fig. 3*f*, green dots), there was a direct relationship between %sv and r_{sc} mean (specifically, %sv equals r_{sc} mean). However, when loading similarity was low (Fig. 3*f*, blue dots), the relationship between %sv and r_{sc} mean was less direct. Namely, r_{sc} mean remained close to zero regardless of %sv. This illustrates that r_{sc} mean and %sv are not the same. It is possible for a population of neurons with high %sv (e.g., Fig. 3*f*, blue dots in outer arc) to have smaller r_{sc} mean than a population with lower %sv (e.g., Fig. 3*f*, green dots in inner arc).

These relationships that we have shown through simulation can be captured mathematically. First, if we have knowledge of the loading weights in the co-fluctuation pattern, the r_{sc} between a pair of neurons can be expressed in terms of the %sv and loading values of the two neurons (Supplementary Math Note A):

$$\rho_{ij} = \sqrt{\phi_i \phi_j}\ \mathrm{sign}(w_i w_j) \tag{1}$$

where ρ_{ij} is the r_{sc} between neurons i and j, ϕ_i and ϕ_j are the %sv of each neuron (expressed as a proportion per neuron, in contrast to %sv in Fig. 3*f* which shows the average %sv across all neurons), and w_i and w_j are the loadings of the neurons in the co-fluctuation pattern. The r_{sc}

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mean is the average of ρ_{ij} values across all neuron pairs. From equation (1), we observe that when loading similarity is high (i.e., most loading weights have the same sign), %sv and r_{sc} mean are directly related (i.e., $\rho_{ij} = \sqrt{\phi_i \phi_j}$). However, when loading similarity is low (i.e., some loading weights are positive and others are negative), r_{sc} mean is small regardless of %sv because some pairs have $\text{sign}(w_i w_j) = +1$ and others have $\text{sign}(w_i w_j) = -1$.

Second, if we have information about the r_{sc} s.d. (instead of loading weights), we can establish the following relationship between %sv, r_{sc} mean, and r_{sc} s.d. (Supplementary Math Note B):

$$\%\text{sv} \approx \sqrt{(r_{sc}\ \text{mean})^2 + (r_{sc}\ \text{s.d.})^2}$$

In other words, in the r_{sc} mean versus r_{sc} s.d. plot, %sv is reflected in the distance of a point from the origin (Fig. 3*f*). This relationship holds regardless of the loading similarity. The intuition is that the %sv corresponds to the magnitude of r_{sc} values (i.e., the $|\rho_{ij}|$ from equation (1)).

These findings highlight the pitfalls of considering a single statistic (e.g., r_{sc} mean) on its own and the benefits of considering multiple statistics (e.g., both r_{sc} mean and s.d.) when trying to draw conclusions about how neurons covary. By considering r_{sc} mean and s.d. together, one can gain insight into the loading similarity (Fig. 3*e*) and the %sv (Fig. 3*f*) of a neuronal population. Thus far, we have only considered the specific case where activity co-fluctuates along a single dimension in the firing rate space. We next considered how pairwise metrics change in the more general case where neuronal activity co-fluctuates along multiple dimensions.

Adding more dimensions tends to reduce r_{sc} mean and s.d.

We sought to assess how dimensionality (i.e., the number of co-fluctuation patterns) is related to pairwise metrics. In simulations, we increased the number of co-fluctuation patterns (compare Fig. 2*a* to *d*; see Methods), while sweeping loading similarity and fixing the total %sv. We found that increasing dimensionality tended to reduce r_{sc} mean and s.d. (Fig. 3*g*, dots for larger dimensionalities lay closer to the origin than dots for smaller dimensionalities).

It seems counterintuitive that adding a new way in which neurons covary reduces the magnitude of r_{sc}. The intuition is that if multiple distinct (i.e., orthogonal) dimensions exist, then a neuron pair interacts in opposing ways along different dimensions. For example, consider two neurons with loadings of the same sign in one co-fluctuation pattern, and opposite sign in the second pattern. If only the first dimension exists, the two neurons would go up and down together and be positively correlated. If only the second dimension exists, the two neurons would co-fluctuate in opposition and be negatively correlated. When both dimensions exist, the positive correlation from the first dimension and the negative correlation from the second dimension offset, and the resulting correlation between the neurons would be smaller than if only the first dimension were present. We formalize the above intuition in Supplementary Math Note C. We also show analytically that increasing dimensionality tends to move points closer to the origin in the r_{sc} mean versus r_{sc} s.d. plot (i.e., decrease r_{sc} mean and s.d.; Supplementary Math Note D).

However, we note that an increase in dimensionality does not imply that *both* r_{sc} mean and r_{sc} s.d. necessarily decrease. For example, in the case where the first co-fluctuation pattern has high loading similarity, adding more dimensions means it is less likely for r_{sc} s.d. to be 0 (Fig. 3*g*, compare dot closest to horizontal axis for '1 dim.' to that for '2 dims.'). The intuition is that if the first co-fluctuation pattern has a loading similarity of 1, the loading weights for all neurons are the same and thus r_{sc} values between all pairs are the same, resulting in r_{sc} s.d. of 0. Adding an orthogonal dimension to this pattern necessarily means adding a pattern with low loading similarity (Supplementary Math Note E), making it less likely for r_{sc} across all pairs to be the same. Therefore, r_{sc} s.d. is unlikely to be 0 for two dimensions (Fig. 3*g*, the smallest r_{sc} s.d. for '2 dims.' is around 0.2). Still, in Figure 3*g* the dots for '2 dims.' are closer to the origin than the

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

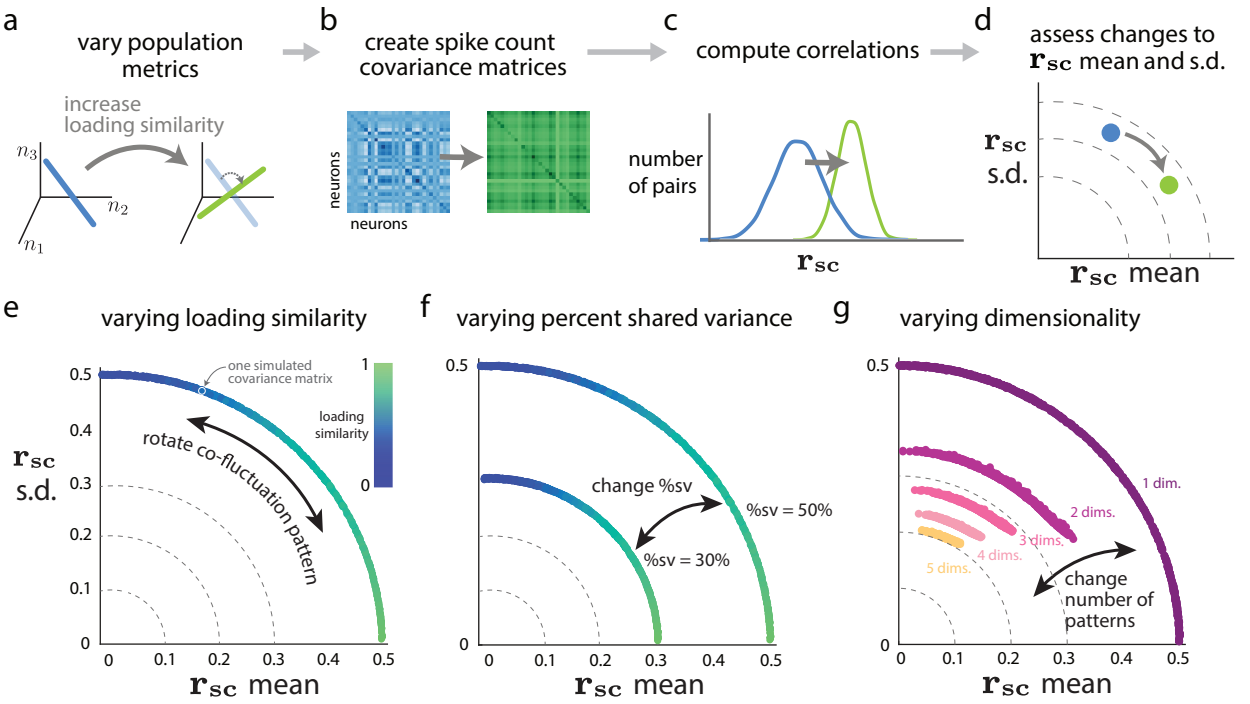


Figure 3: Relationship between population metrics and pairwise metrics. Panels **a-d** describe the simulation procedure to assess how systematic changes in population metrics lead to changes in pairwise metrics. **a**. We first systematically varied one of the population metrics while keeping the others fixed. For example, we can increase the loading similarity from a low value (left, blue) to a high value (right, green), while keeping %sv and dimensionality fixed. **b**. Then, we constructed covariance matrices corresponding to each value of the population metric in panel **a** (see Methods), without generating synthetic data. **c**. For each covariance matrix from panel **b**, we directly computed the correlations (i.e., the r_{sc} distributions). **d**. We computed r_{sc} mean and r_{sc} s.d. from the r_{sc} distributions in panel **c** and then assessed how the change in a given population metric from panel **a** changed pairwise metrics. In this case, the increase in loading similarity increased r_{sc} mean and decreased r_{sc} s.d. (blue dot to green dot). **e**. Varying loading similarity with a fixed %sv of 50% and dimensionality of 1. Each dot corresponds to the r_{sc} mean and r_{sc} s.d. of one simulated covariance matrix with specified population metrics (dots are close together and appear to form a continuum). The color of each dot corresponds to the loading similarity (see Methods), where a value of 1 indicates that all loading weights have the same value. **f**. Varying %sv. The same setting as in panel **e**, except we consider two different values of percent shared variance (50% and 30%). **g**. Varying dimensionality (i.e., number of co-fluctuation patterns) while sweeping loading similarity between 0 and 1 and keeping %sv fixed at 50%. In this simulation, the relative strengths of each dimension uniform across dimensions (i.e., flat eigenspectra; see Methods).

dots for '1 dim', implying that even if r_{sc} s.d. increases with an increase in dimensionality, the r_{sc} mean must decrease to a larger extent (Supplementary Math Note D).

The relative strength of each dimension impacts pairwise metrics.

In the previous simulation (Fig. 3*g*), we assumed that each dimension explained an equal proportion of the overall shared variance (e.g., for two dimensions, each dimension explained half of the shared variance; see Methods). However, it is typically the case for recorded neuronal activity that some dimensions explain more shared variance than others; in other words, neuronal activity co-fluctuates more strongly along some patterns than others (Sadtler et al., 2014; Williamson et al., 2016; Mazzucato et al., 2016; Gallego et al., 2018; Huang et al., 2019; Stringer et al., 2019a; Ruff et al., 2019b). We sought to assess the influence of the relative strength of each dimension on

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

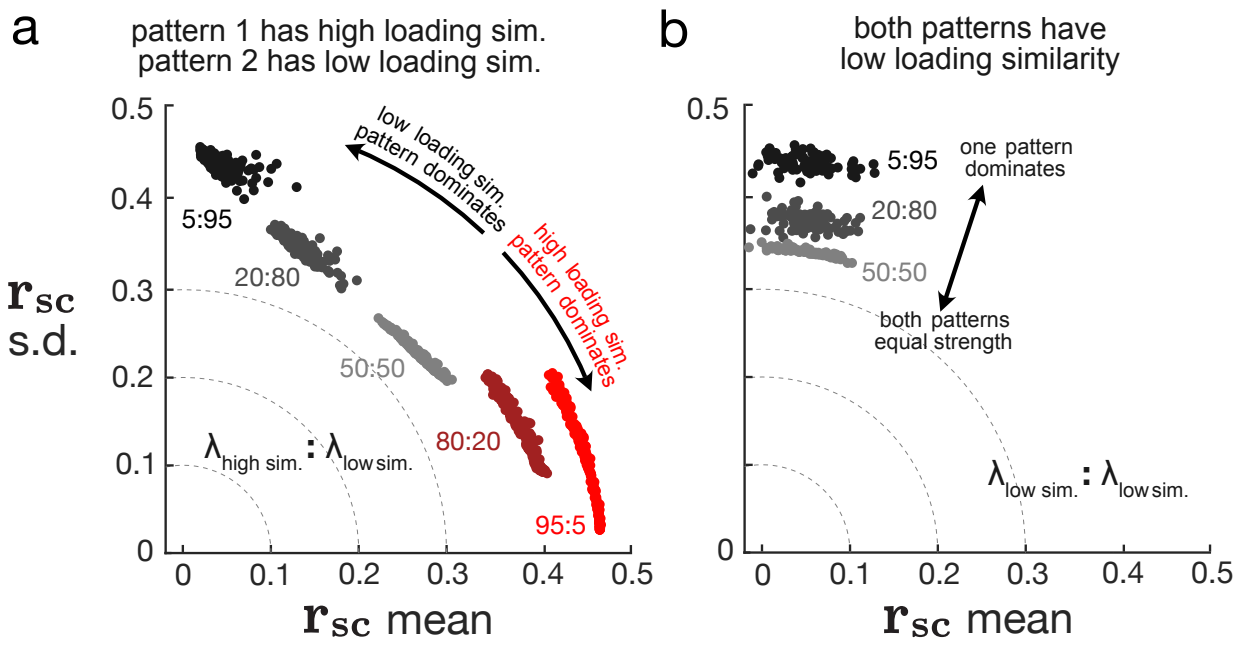


Figure 4: Relative strengths of dimensions affect r_{sc} distributions. With dimensionality of 2, we systematically varied the relative strengths of the two dimensions with a fixed total %sv of 50%. We considered two scenarios: 1) one dimension has high loading similarity and the other dimension has low loading similarity (panel **a**) and 2) both dimensions have low loading similarity (panel **b**). Each dot represents one simulated covariance matrix and r_{sc} distribution. The color of the dots indicate different relative strengths between the two dimensions, and numbers next to each cloud of dots indicate the ratio between the relative strength associated with each dimension. For example, in panel **a**, red dots correspond to the high loading similarity dimension being 19 times stronger (95:5) than the low loading similarity dimension. Black dots correspond to the low loading similarity dimension being 19 times stronger (5:95) than the high loading similarity dimension. In panel **b**, since both patterns have low loading similarity, clouds for 80:20 and 95:5 are very similar to clouds for 20:80 and 5:95 respectively and are thus omitted for clarity.

pairwise metrics.

We reasoned that stronger dimensions would play a larger role than weaker dimensions in determining the r_{sc} distribution and pairwise metrics. Extending equation (1) to multiple dimensions, we show that the r_{sc} between a pair of neurons can be expressed as the sum of a contribution from each constituent dimension (Supplementary Math Note C). The stronger a dimension, the larger the magnitude of its contribution to r_{sc}, and thus the larger its impact on r_{sc} mean and s.d.

To test this empirically, we performed a simulation with two dimensions, while systematically varying the relative strength of each dimension. We considered two scenarios: (1) one dimension has a pattern with high loading similarity and one dimension has a pattern with low loading similarity (Fig. 4*a*), and (2) both dimensions have patterns with low loading similarity (Fig. 4*b*). Note that both dimensions cannot have patterns with high loading similarity because they would not be orthogonal (Supplementary Math Note E).

In scenario (1) where one dimension's pattern has high loading similarity and the other has low loading similarity, r_{sc} mean and r_{sc} s.d. reflects the loading similarity of the dominant dimension (Fig. 4*a*). When the dimension with a high loading similarity pattern dominated, r_{sc} mean was large and r_{sc} s.d. was small (Fig. 4*a*, red dots are close to horizontal axis). When the dimension with a low loading similarity pattern dominated, r_{sc} mean was small and r_{sc} s.d. was large (Fig. 4*a*, black dots are close to vertical axis). When the two dimensions were of equal strength (i.e., neither

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dimension dominated), r_{sc} mean and r_{sc} s.d. were both intermediate values (Fig. 4*a*, light gray dots are between red and black dots). Thus, the dimensions along which neuronal activity co-fluctuates most strongly have a greater influence on pairwise metrics.

In scenario (2) where both dimensions have patterns of low loading similarity, r_{sc} mean was low and r_{sc} s.d. was high (Fig. 4*b*), similar to when there is one dimension with low loading similarity (Fig. 3*e*, blue dots). When we made one dimension stronger than the other, r_{sc} mean remained low and r_{sc} s.d. remained high (Fig. 4*b*, light gray dots and black dots are both close to vertical axis) because both patterns had low loading similarity. However, the radius of the arc increased (Fig. 4*b*, black dots farther from the origin than light gray dots), and was close to the arc that would have been produced with a single dimension (Fig. 3*g*, '1 dim.'). Thus, whereas changing the number of dimensions causes discrete jumps in the arc radius (Fig. 3*g*), changing the relative strength of each dimension allows for r_{sc} mean and r_{sc} s.d. to vary continuously between the arcs for different dimensionalities. Put another way, changing the relative strength of each dimension varies the "effective dimensionality" of population activity in a continuous manner. Neuronal activity for which one dimension dominates another (Fig. 4*b*, black dots) has a lower effective dimensionality than when both dimensions have equal strength (Fig. 4*b*, light gray dots).

Reporting only a single statistic provides an incomplete description of population covariability

Figure 5 summarizes the relationships that we have established between pairwise metrics and population metrics. Rotating a co-fluctuation pattern from a low loading similarity to a high loading similarity increases r_{sc} mean and decreases r_{sc} s.d. along an arc (Fig. 5, arrow outside pink arc). Decreasing %sv decreases both r_{sc} mean and s.d. (Fig. 5, arrow pointing toward origin), and increasing dimensionality also tends to decrease r_{sc} mean and s.d. (Fig. 5, pink to yellow shaded regions).

These results provide a cautionary tale that using a single statistic on its own provides an opaque description of population-wide covariability. For example, a change in r_{sc} mean could correspond to changes in loading similarity, %sv, dimensionality, or a combination of the three. Likewise, reporting dimensionality on its own would be incomplete because the role of a dimension in explaining population-wide covariability depends how much shared variance it explains and the loading similarity of its co-fluctuation pattern. For example, consider a decrease in dimensionality by 1. This would have little impact on population-wide covariability if the removed dimension explains only a small amount of shared variance, whereas it could have a large impact if the removed dimension explains a large amount of shared variance.

Considering multiple statistics together provides a richer description of population-wide covariability. For example, in the case where population activity co-fluctuates along a single dimension, r_{sc} mean and r_{sc} s.d. can be used together to approximate %sv (using distance from the origin) and deduce whether loading similarity is low (r_{sc} s.d. $>$ r_{sc} mean) or high (r_{sc} mean $>$ r_{sc} s.d.), whereas r_{sc} mean alone would not provide much information about %sv or loading similarity (cf. Fig. 5). In the next section, we further demonstrate using neuronal recordings how relating pairwise and population metrics using the framework we have developed (Fig. 5) provides a richer description of how neurons covary than using a single statistic (e.g., r_{sc} mean) alone.

Case study: V4 neuronal recordings during spatial attention

When spatial attention is directed to the receptive fields of neurons in area V4 of macaque visual cortex, r_{sc} mean among those neurons decreases (Cohen and Maunsell, 2009; Mitchell et al., 2009;

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

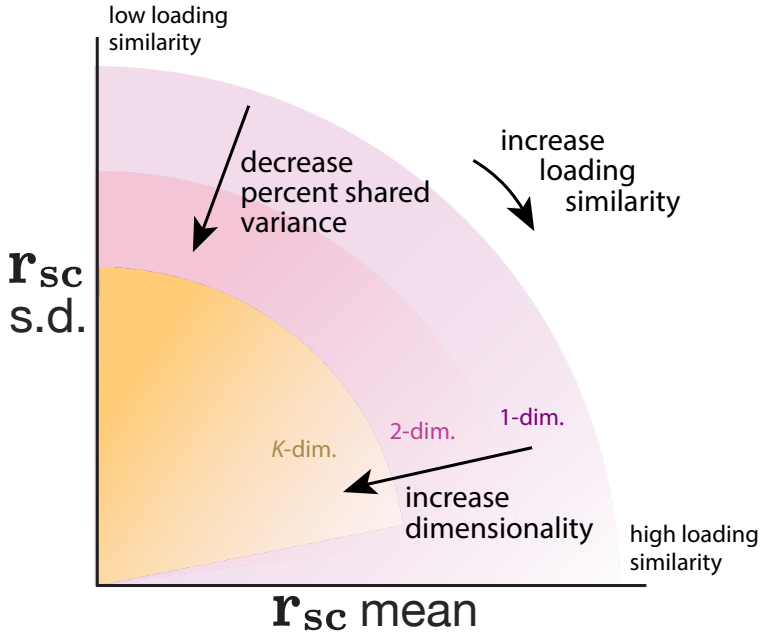


Figure 5: Summary of relationship between pairwise and population metrics. A change in r_{sc} mean and r_{sc} s.d. may correspond to changes in loading similarity, %sv, dimensionality, or a combination of the three. Shaded regions indicate the possible r_{sc} mean and r_{sc} s.d. values for different dimensionalities; increasing dimensionality tends to decrease r_{sc} mean and r_{sc} s.d. (shaded regions for larger dimensionalities become smaller). Within each shaded region, decreasing %sv decreases both r_{sc} mean and s.d. radially toward the origin. Finally, rotating co-fluctuation patterns such that the loadings are more similar (going from low to high loading similarity) results in moving clockwise along an arc such that r_{sc} mean increases and r_{sc} s.d. decreases. We also note two subtle trends. First, there are more possibilities for loading similarity to be low than high (Supplementary Math Note E), suggesting that r_{sc} s.d. will generally tend to be larger than r_{sc} mean if neuronal activity varied along a randomly chosen co-fluctuation pattern (shading within each region is darker near the vertical axis than the horizontal axis). Second, this effect becomes exaggerated for higher-dimensional neuronal activity as many dimensions can have low loading similarity but only one dimension can have high loading similarity (Supplementary Math Note E). Thus, it becomes progressively unlikely for r_{sc} s.d. to be 0 as dimensionality increases (shaded regions for larger dimensionalities lifted off the horizontal axis).

Gregoriou et al., 2014; Snyder et al., 2016, 2018). This decrease has often been attributed to a reduction in shared modulations among the neurons. However, we have shown both mathematically and in simulations that several distinct changes in population metrics (e.g., decrease in loading similarity, decrease in %sv, or an increase in dimensionality) could underlie this decrease in r_{sc} mean (Fig. 5). Here, we sought to assess which aspects of population-wide covariability underlie, and how each of them contribute to, the overall decrease in r_{sc} mean.

We analyzed activity recorded simultaneously from tens of neurons in macaque V4 while the animal performed an orientation-change detection task (Fig. 6*a*; previously reported in Snyder et al., 2018). To probe spatial attention, we cued the animal to the location of the stimulus that was more likely to change in orientation. As expected, perceptual sensitivity increased for orientation changes in the cued stimulus location (Fig. 6*a* inset, red dot above black dot). 'Attend-in' trials were those in which the cued stimulus location was inside the aggregate receptive fields (RFs) of the recorded V4 neurons, whereas 'attend-out' trials were those in which the cued stimulus location was in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

opposite visual hemifield.

For pairwise metrics, r_{sc} mean decreased when attention was directed into the RFs of the V4 neurons (Fig. 6*b*, left panel), consistent with previous studies (Cohen and Maunsell, 2009; Mitchell et al., 2009; Gregoriou et al., 2014; Snyder et al., 2016, 2018). We further found that r_{sc} s.d. was lower for attend-in trials than for attend-out trials, an effect not reported previously (Fig. 6*b*, right panel; also see Supplementary Fig. 1 for session-by-session pairwise metrics).

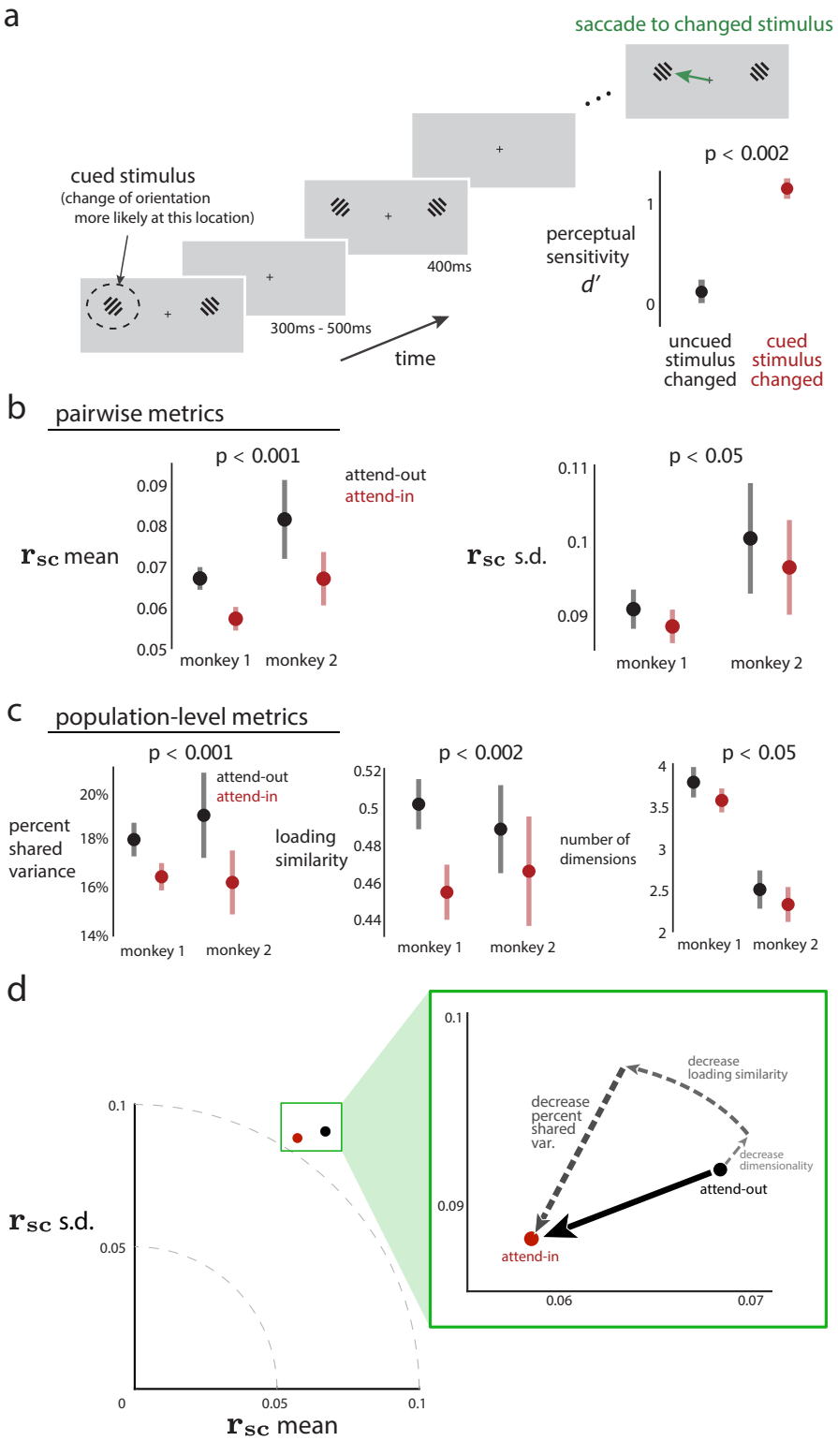
The decrease in both r_{sc} mean and r_{sc} s.d. could arise from several different types of distinct changes in population-wide covariability (Fig. 5). To compute the population metrics, we applied factor analysis (FA) separately to attend-out and attend-in trials (see Methods). FA is the most basic dimensionality reduction method that characterizes shared variance among neurons (Cunningham and Yu, 2014), and is consistent with how we created covariance matrices in Figures 3 and 4. We found three distinct changes in population metrics. First, neuronal activity during attend-in trials had lower %sv than during attend-out trials (Fig. 6*c*, left), consistent with previous interpretations that attention reduces the strength of shared modulations (Rabinowitz et al., 2015; Ecker et al., 2016; Huang et al., 2019; Ruff et al., 2019b). Second, we also found lower loading similarity for attend-in trials than attend-out trials for the dominant dimension (i.e., the dimension that explains the largest proportion of the shared variance; Fig. 6*c*, middle). This implies that, with attention, neurons in the population co-fluctuate in a more heterogeneous manner (i.e., more pairs of neurons co-fluctuate in opposition, and fewer pairs co-fluctuate together). Third, we found that dimensionality was slightly lower for attend-in than attend-out trials (Fig. 6*c*, right). Thus, on average, a smaller number of distinct shared signals were present when attention was directed into the neurons' RFs. The small change in dimensionality is consistent with the relative strength of each dimension (i.e., eigenspectrum shape) being similar for attend-in and attend-out (Supplementary Fig. 2). Taken together, this collection of observations of both pairwise and population metrics leads to a more refined view of how attention affects population-wide covariability.

The pairwise (Fig. 6*b*) and population (Fig. 6*c*) metrics are computed based on the same recorded activity and each represents a different view of population activity. The central contribution of our work is to provide a framework by which to understand these two perspectives and five different metrics in a coherent manner. Using the relationships between pairwise and population metrics we have established in the r_{sc} mean versus r_{sc} s.d. space (Fig. 5), we can decompose the decrease in r_{sc} mean and s.d. into: 1) a small decrease in dimensionality (Fig. 6*d*, small dashed arrow), 2) a decrease in loading similarity (Fig. 6*d*, medium dashed arrow), and 3) a substantial decrease in %sv (Fig. 6*d*, large dashed arrow). Overall, this analysis demonstrates the insufficiency of any one measure of correlated variability, and the value in considering pairwise and population metrics together, with a bridge that allows one to navigate between the two.

Discussion

Coordinated variability in the brain has long been linked to the neural computations underlying a diverse range of functions, including sensory encoding, decision making, attention, learning, and more. In this study, we sought to relate two major bodies of work investigating the coordinated activity among neurons: studies that measure spike count correlation between pairs of neurons (r_{sc}) and studies that use dimensionality reduction to measure population-wide covariability. We considered three population metrics and established analytically and empirically that: 1) increasing loading similarity corresponds to increasing r_{sc} mean and decreasing r_{sc} s.d., 2) decreasing percent shared variance (%sv) corresponds to decreasing both r_{sc} mean and s.d., and 3) increasing dimensionality tends to decrease r_{sc} mean and s.d. Applying this understanding to recordings in macaque

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 6 ***(previous page)*****: An observed decrease in r_{sc} mean of macaque V4 neurons during a spatial attention task corresponds to changes in multiple population metrics. a**. Experimental task design. On each trial, monkeys maintained fixation while Gabor stimuli were presented for 400 ms (with 300-500 ms in between presentations). When one of the stimuli changed orientation, animals were required to saccade to the changed stimulus to obtain a reward. At the beginning of a block of trials, we performed an attentional manipulation by cuing animals to the location of the stimulus that was more likely to change for that block (dashed circle denotes the cued stimulus and was not presented on the screen). The cued location alternated between blocks. Animals were more likely to detect a change in stimulus at cued rather than uncued locations (inset in bottom right, $p < 0.002$ for both animals; data for monkey 1 is shown). During this task, we recorded activity from V4 neurons whose receptive fields (RFs) overlapped with one of the stimulus locations. **b**. r_{sc} mean (left panel) and r_{sc} s.d. (right panel) across recording sessions for two animals. Black denotes 'attend-out' trials (i.e., the cued location was outside the recorded V4 neurons' RFs), and red denotes 'attend-in' trials (i.e., the cued location was inside the RFs). Data was pooled across both animals to compute p-values reported in titles for comparison of attend-out (black) and attend-in (red). For individual animals, r_{sc} mean was lower for attend-in than attend-out ($p < 0.001$ for each animal). r_{sc} s.d. was also lower for attend-in than attend-out ($p < 0.05$ for monkey 1, and $p = 0.148$ for monkey 2). **c**. Population metrics identified across recording sessions for two animals (same data as in **b**). Black denotes attend-in trials, red denotes attend-out trials. Data was again pooled across animals to compute p-values reported in titles for comparing attend-out and attend-in. %sv was lower for attend-in than attend-out ($p < 0.001$ for monkey 1 and $p < 0.02$ for monkey 2). Loading similarity was lower for attend-in than attend-out ($p < 0.001$ for monkey 1 and $p = 0.162$ for monkey 2). Dimensionality was lower for attend-in than attend-out ($p = 0.113$ for monkey 1 and $p = 0.174$ for monkey 2). In panels **a**-**c**, dots indicate means and error bars indicate 1 s.e.m., both computed across recording sessions. **d**. Summary of the real data results. Attention decreases both r_{sc} mean and r_{sc} s.d. (black dot to red dot). These decreases in pairwise metrics correspond to a combination of decreases in %sv, loading similarity, and dimensionality (dashed arrows).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

V4, we found that the previously-reported decrease in mean spike count correlation associated with attention stemmed from a decrease in %sv, a decrease in loading similarity, and decrease in dimensionality. This analysis revealed that attention involves multiple changes in how neurons interact that are not well captured by a single statistic alone. Overall, our work demonstrates that common ground exists between the literatures of spike count correlation and dimensionality reduction approaches, and builds the intuition and formalism to navigate between them.

Our work provides a cautionary tale for attempting to summarize population-wide covariability using one, or a small number of, statistics. For example, reporting only r_{sc} mean is incomplete because several distinct changes in population-wide covariability can correspond to the same change in r_{sc} mean. In a similar vein, reporting only dimensionality is incomplete because it does not indicate how strongly the neurons covary, nor their co-fluctuation patterns. For this reason, we recommend reporting several different pairwise and population metrics (e.g., the five used in this study along with the eigenspectrum of the shared covariance matrix), as long as they can be reliably measured from the data available. This not only allows for a deeper and more complete understanding of how neurons covary, but also it allows one to make tighter connections to previous literature that uses the same metrics. Future work may seek to revisit previous results of correlated neuronal variability that are based on a single statistic (e.g., r_{sc} mean), and reinterpret them within a framework that considers multiple perspectives and statistics of population-wide covariability, such as that presented here.

There are some situations where it is not feasible to reliably measure population statistics, such as recording from a small number of neurons in deep brain structures (Nevet et al., 2007; Liu et al., 2013). In such situations, the r_{sc} can be measured between pairs of neurons recorded in each session and then averaged across sessions to obtain the r_{sc} mean (Supplementary Fig. 3). Based on our findings, we recommend that studies which report r_{sc} mean also report r_{sc} s.d. because the latter provides additional information about population-wide covariability. For example, in the special case of one latent dimension (typically not known in advance for real data), measuring r_{sc} mean and r_{sc} s.d. allows one to estimate the loading similarity and %sv (cf. Fig. 3*e-f*). In general, even when there is more than one latent dimension in the population, r_{sc} s.d. provides value in situating the data in the r_{sc} mean versus r_{sc} s.d. plot (cf. Fig. 5). Changes in r_{sc} mean and s.d. can then inform changes in population metrics based on the relationships established in this work (cf. Fig. 6*d*).

We considered three population metrics — dimensionality, percent shared variance (%sv), and loading similarity — that summarize the structure of population-wide covariability and are rooted in well-established concepts in existing literature. First, dimensionality has been used to describe how neurons covary across conditions (i.e., an analysis of trial-averaged firing rates; Churchland et al., 2012; Rigotti et al., 2013; Mante et al., 2013; Cowley et al., 2016; Kobak et al., 2016; Sohn et al., 2019), as well as how neurons covary from trial to trial (Yu et al., 2009; Santhanam et al., 2009; Sadtler et al., 2014; Rabinowitz et al., 2015; Mazzucato et al., 2016; Williamson et al., 2016; Bittner et al., 2017; Athalye et al., 2017; Williams et al., 2018; Stringer et al., 2019a; Recanatesi et al., 2019). We focused on the latter in our study to connect with the r_{sc} literature, which also seeks to understand the shared trial-to-trial variability between neurons. To focus on the shared variability among neurons, we used factor analysis (FA) to measure dimensionality. Another commonly-used dimensionality reduction method, principal components analysis (PCA), although appropriate for studying trial-averaged activity, does not distinguish between variability that is shared among neurons and variability that is independent to each neuron. Second, investigating the loading similarity has provided insight about whether shared variability among neurons arises from a shared global factor which drives neurons to increase and decrease their activity together (Ecker et al., 2014; Okun et al., 2015; Lin et al., 2015; Rabinowitz et al., 2015; Williamson et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2016; Huang et al., 2019) or whether the co-fluctuations involve a more intricate pattern across the neuronal population (Snyder et al., 2018; Insanally et al., 2019; Cowley et al., 2020). Understanding how loading similarity and these patterns of shared variability interact with (e.g., align with or are orthogonal to) patterns of stimulus encoding and downstream readouts will be important to understand how the brain perceives and computes (Averbeck et al., 2006; Moreno-Bote et al., 2014; Kohn et al., 2016; Ni et al., 2018; Ruff and Cohen, 2019a; Cowley et al., 2020; Rumyantsev et al., 2020; Bartolo et al., 2020). Third, we have previously reported %sv for area V1 (Williamson et al., 2016), area M1 (Hennig et al., 2018), and network models (Williamson et al., 2016; Bittner et al., 2017). Conceptually, %sv and r_{sc} mean are both designed to capture the strength of shared variability in a population of neurons. Thus, we might initially think that there should be a one-to-one correspondence between the two quantities. Indeed, if the population activity is described by one co-fluctuation pattern with a high loading similarity, there is a direct relationship between %sv and r_{sc} mean (Fig. 3*f*). However, in general, %sv and r_{sc} mean do not have a one-to-one correspondence between them (Fig. 3*f*, moderate or low loading similarity).

Although pairwise correlation and dimensionality reduction have most commonly been computed based on spike counts, several studies have also computed these metrics on neuronal activity recorded using other modalities, such as calcium imaging (Harvey et al., 2012; Ahrens et al., 2012; Dechery and MacLean, 2018; Stringer et al., 2019a; Rumyantsev et al., 2020). The relationships that we established here between pairwise and population metrics are properties of covariance matrices in general and do not rely on or assume recordings of neuronal spikes. Thus, the intuition built here can be applied to other recording modalities.

Our work here focused on studying interactions within a single population of neurons. Technological advances are enabling recordings from multiple distinct populations simultaneously, including neurons in different brain areas, neurons in different cortical layers, or different neuron types (e.g., Ahrens et al., 2013; Jiang et al., 2015; Jun et al., 2017). Studies are dissecting the interactions between these distinct populations using pairwise correlation (Smith et al., 2012; Pooresmaeili et al., 2014; Oemisch et al., 2015; Zandvakili and Kohn, 2015; Ruff and Cohen, 2016a; Snyder et al., 2016) and dimensionality reduction (Semedo et al., 2014; Buesing et al., 2014; Bittner et al., 2017; Perich et al., 2018; Semedo et al., 2019; Ames and Churchland, 2019; Ruff and Cohen, 2019a; Veuthey et al., 2020; Cowley et al., 2020). As we have shown here for a single population of neurons, considering a range of metrics from both the pairwise correlation and dimensionality reduction perspectives, and understanding how they relate to one another, will provide rich descriptions of how different neuronal populations interact.

Data and code availability

Relevant data and analysis code used to generate the results are available upon request from the authors.

Acknowledgements

The authors would like to thank Samantha Schmitt for help with data collection, and João Semedo for helpful discussions. BRC was supported by CV Starr Foundation Fellowship. ACS was supported by NIH grant K99EY025768. MAS and BMY were supported by NIH CRCNS R01 MH118929, NIH R01 EB026953, and NSF NCS BCS 1954107/1734916. MAS was supported by NIH R01 EY022928, NIH R01 EY029250, NIH P30 EY008098. BMY was supported by Simons

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

453 Foundation 364994 and 543065, NIH R01 HD071686, NIH CRCNS R01 NS105318, and NSF NCS
454 BCS1533672.

455 Author contributions

456 Mathematical derivations: AU, RM; simulations: AU, RM, BRC; data collection: ACS; data
457 analysis: AU, BRC, writing: AU, RM, BRC; reviewing and editing: all authors. AU, RM, BRC
458 contributed equally to this work (first authors). MAS, BMY contributed equally to this work
459 (senior authors).

460 Competing interests

461 The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Abbott, L. F. and Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural computation*, 11(1):91–101, 1999.

Adibi, M., McDonald, J. S., Clifford, C. W., and Arabzadeh, E. Adaptation improves neural coding efficiency despite increasing correlations in variability. *Journal of Neuroscience*, 33(5):2108–2120, 2013.

Ahrens, M. B., Li, J. M., Orger, M. B., Robson, D. N., Schier, A. F., Engert, F., and Portugues, R. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature*, 485(7399):471–477, 2012.

Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M., and Keller, P. J. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods*, 10(5):413, 2013.

Ames, K. C. and Churchland, M. M. Motor cortex signals for each arm are mixed across hemispheres and neurons yet partitioned within the population response. *eLife*, 8:e46159, October 2019.

Athalye, V. R., Ganguly, K., Costa, R. M., and Carmena, J. M. Emergence of Coordinated Neural Dynamics Underlies Neuroprosthetic Learning and Skillful Control. *Neuron*, 93(4):955–970.e5, February 2017.

Averbeck, B. B., Latham, P. E., and Pouget, A. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366, 2006.

Bair, W., Zohary, E., and Newsome, W. T. Correlated firing in macaque visual area mt: time scales and relationship to behavior. *Journal of Neuroscience*, 21(5):1676–1697, 2001.

Bartolo, R., Saunders, R. C., Mitz, A. R., and Averbeck, B. B. Information-limiting correlations in large neural populations. *Journal of Neuroscience*, 40(8):1668–1678, 2020.

Bittner, S. R., Williamson, R. C., Snyder, A. C., Litwin-Kumar, A., Doiron, B., Chase, S. M., Smith, M. A., and Yu, B. M. Population activity structure of excitatory and inhibitory neurons. *PloS one*, 12(8):e0181773, 2017.

Bondy, A. G., Haefner, R. M., and Cumming, B. G. Feedback determines the structure of correlated variability in primary visual cortex. *Nature Neuroscience*, page 1, 2018.

Buesing, L., Machado, T. A., Cunningham, J. P., and Paninski, L. Clustered factor analysis of multineuronal spike data. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 3500–3508. Curran Associates, Inc., 2014.

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, 13(3):369–378, 2010.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., and Shenoy, K. V. Neural population dynamics during reaching. *Nature*, 487(7405):51–6, Jul 2012.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cohen, M. R. and Kohn, A. Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819, 2011.

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

Cohen, M. R. and Maunsell, J. H. A neuronal population measure of attention predicts behavioral performance on individual trials. *Journal of Neuroscience*, 30(45):15241–15253, 2010.

Cowley, B. R., Smith, M. A., Kohn, A., and Yu, B. M. Stimulus-driven population activity patterns in macaque primary visual cortex. *PLOS Computational Biology*, 12(12):e1005185, 2016.

Cowley, B. R., Snyder, A. C., Acar, K., Williamson, R. C., Yu, B. M., and Smith, M. A. Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. *Neuron*, 108(3):551–567.e8, November 2020.

Cunningham, J. P. and Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, 2014.

Dechery, J. B. and MacLean, J. N. Functional triplet motifs underlie accurate predictions of single-trial responses in populations of tuned and untuned V1 neurons. *PLOS Computational Biology*, 14(5):e1006153, May 2018.

Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., and Tolias, A. S. Decorrelated neuronal firing in cortical microcircuits. *science*, 327(5965):584–587, 2010.

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., Smirnakis, S. M., Bethge, M., and Tolias, A. S. State dependence of noise correlations in macaque primary visual cortex. *Neuron*, 82(1):235–248, 2014.

Ecker, A. S., Denfield, G. H., Bethge, M., and Tolias, A. S. On the structure of neuronal population activity under fluctuations in attentional state. *Journal of Neuroscience*, 36(5):1775–1789, 2016.

Erisken, S., Vaiceliunaite, A., Jurjut, O., Fiorini, M., Katzner, S., and Busse, L. Effects of locomotion extend throughout the mouse early visual system. *Current Biology*, 24(24):2899–2907, 2014.

Gallego, J. A., Perich, M. G., Miller, L. E., and Solla, S. A. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017.

Gallego, J. A., Perich, M. G., Naufel, S. N., Ethier, C., Solla, S. A., and Miller, L. E. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*, 9(1):4233, October 2018.

Gao, P. and Ganguli, S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, 32:148–155, June 2015.

Gregoriou, G. G., Rossi, A. F., Ungerleider, L. G., and Desimone, R. Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4. *Nature Neuroscience*, 17(7):1003–1011, 2014.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Gu, Y., Liu, S., Fetsch, C. R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G. C., and Angelaki, D. E. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron*, 71(4):750–761, 2011.

Harvey, C. D., Coen, P., and Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62–68, 2012.

Hennig, J. A., Golub, M. D., Lund, P. J., Sadtler, P. T., Oby, E. R., Quick, K. M., Ryu, S. I., Tyler-Kabara, E. C., Batista, A. P., Yu, B. M., and Chase, S. M. Constraints on neural redundancy. *eLife*, 7:e36774, August 2018.

Herrero, J. L., Gieselmann, M. A., Sanayei, M., and Thiele, A. Attention-induced variance and noise correlation reduction in macaque v1 is mediated by nmda receptors. *Neuron*, 78(4):729–739, 2013.

Huang, C., Ruff, D. A., Pyle, R., Rosenbaum, R., Cohen, M. R., and Doiron, B. Circuit models of low-dimensional shared variability in cortical networks. *Neuron*, 101(2):337–348, 2019.

Huang, X. and Lisberger, S. G. Noise Correlations in Cortical Area MT and Their Potential Impact on Trial-by-Trial Variation in the Direction and Speed of Smooth-Pursuit Eye Movements. *Journal of Neurophysiology*, 101(6):3012–3030, June 2009.

Insanally, M. N., Carcea, I., Field, R. E., Rodgers, C. C., DePasquale, B., Rajan, K., DeWeese, M. R., Albanna, B. F., and Froemke, R. C. Spike-timing-dependent ensemble encoding by non-classically responsive cortical neurons. *eLife*, 8, 2019.

Jeanne, J., Sharpee, T., and Gentner, T. Associative Learning Enhances Population Coding by Inverting Interneuronal Correlation Patterns. *Neuron*, 78(2):352–363, April 2013.

Jiang, X., Shen, S., Cadwell, C. R., Berens, P., Sinz, F., Ecker, A. S., Patel, S., and Tolias, A. S. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, 350(6264), November 2015.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, *c*C., et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232, 2017.

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

Kelly, R. C., Smith, M. A., Samonds, J. M., Kohn, A., Bonds, A., Movshon, J. A., and Lee, T. S. Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. *Journal of Neuroscience*, 27(2):261–264, 2007.

Kiani, R., Cueva, C. J., Reppas, J. B., and Newsome, W. T. Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials. *Current Biology*, 24(13): 1542–1547, 2014.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Kohn, A. and Smith, M. A. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience*, 25(14):3661–3673, 2005.

Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. Correlations and neuronal population information. *Annual Review of Neuroscience*, 39:237–256, 2016.

Lee, D., Port, N. L., Kruse, W., and Georgopoulos, A. P. Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *Journal of Neuroscience*, 18(3):1161–1170, 1998.

Lin, I.-C., Okun, M., Carandini, M., and Harris, K. D. The nature of shared cortical variability. *Neuron*, 87(3):644–656, 2015.

Liu, S., Gu, Y., DeAngelis, G. C., and Angelaki, D. E. Choice-related activity and correlated noise in subcortical vestibular neurons. *Nature Neuroscience*, 16(1):89, 2013.

Luo, T. Z. and Maunsell, J. H. Neuronal modulations in visual cortex are associated with only one of multiple components of attention. *Neuron*, 86(5):1182–1188, 2015.

Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, 2013.

Maynard, E. M., Hatsopoulos, N. G., Ojakangas, C. L., Acuna, B. D., Sanes, J. N., Normann, R. A., and Donoghue, J. P. Neuronal interactions improve cortical population coding of movement direction. *Journal of Neuroscience*, 19(18):8083–8093, 1999.

Mazor, O. and Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron*, 48(4):661–673, 2005.

Mazzucato, L., Fontanini, A., and La Camera, G. Stimuli reduce the dimensionality of cortical activity. *Frontiers in systems neuroscience*, 10:11, 2016.

Minces, V., Pinto, L., Dan, Y., and Chiba, A. A. Cholinergic shaping of neural correlations. *Proceedings of the National Academy of Sciences of the United States of America*, 114(22):5725–5730, 2017.

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area v4. *Neuron*, 63(6):879–888, 2009.

Miura, K., Mainen, Z. F., and Uchida, N. Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. *Neuron*, 74(6):1087–1098, 2012.

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. Information-limiting correlations. *Nature Neuroscience*, 17(10):1410, 2014.

Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., and Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nature neuroscience*, 22(10):1677–1686, 2019.

Nevet, A., Morris, G., Saban, G., Arkadir, D., and Bergman, H. Lack of spike-count and spike-time correlations in the substantia nigra reticulata despite overlap of neural responses. *Journal of neurophysiology*, 98(4):2232–2243, 2007.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J., and Cohen, M. R. Learning and attention reveal a general relationship between population activity and behavior. *Science*, 359(6374):463–465, 2018.

Nienborg, H., R. Cohen, M., and Cumming, B. G. Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annual review of neuroscience*, 35:463–483, 2012.

Oemisch, M., Westendorff, S., Everling, S., and Womelsdorf, T. Interareal Spike-Train Correlations of Anterior Cingulate and Dorsal Prefrontal Cortex during Attention Shifts. *Journal of Neuroscience*, 35(38):13076–13089, September 2015.

Okun, M., Steinmetz, N. A., Cossell, L., Iacaruso, M. F., Ko, H., Barthó, P., Moore, T., Hofer, S. B., Mrsic-Flogel, T. D., Carandini, M., et al. Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521(7553):511–515, 2015.

Pang, R., Lansdell, B. J., and Fairhall, A. L. Dimensionality reduction in neuroscience. *Current Biology*, 26(14):R656–R660, July 2016.

Perich, M. G., Gallego, J. A., and Miller, L. E. A Neural Population Mechanism for Rapid Learning. *Neuron*, 100(4):964–976.e7, November 2018.

Ponce-Alvarez, A., Thiele, A., Albright, T. D., Stoner, G. R., and Deco, G. Stimulus-dependent variability and noise correlations in cortical mt neurons. *Proceedings of the National Academy of Sciences*, 110(32):13162–13167, 2013.

Pooresmaeili, A., Poort, J., and Roelfsema, P. R. Simultaneous selection by object-based attention in visual and frontal cortex. *Proceedings of the National Academy of Sciences*, 111(17):6467–6472, April 2014.

Qi, X.-L. and Constantinidis, C. Correlated discharges in the primate prefrontal cortex before and after working memory training. *The European Journal of Neuroscience*, 36(11):3538–3548, 2012.

Rabinowitz, N. C., Goris, R. L., Cohen, M., and Simoncelli, E. P. Attention stabilizes the shared gain of v4 populations. *Elife*, 4, 2015.

Recanatesi, S., Ocker, G. K., Buice, M. A., and Shea-Brown, E. Dimensionality in recurrent spiking networks: Global trends in activity and local origins in connectivity. *PLOS Computational Biology*, 15(7):e1006446, July 2019.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., and Fusi, S. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590, 2013.

Romo, R., Hernández, A., Zainos, A., and Salinas, E. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron*, 38(4):649–657, 2003.

Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., and Doiron, B. The spatial structure of correlated neuronal variability. *Nature Neuroscience*, 20(1):107, 2017.

Ruff, D. A. and Cohen, M. R. Global cognitive factors modulate correlated response variability between v4 neurons. *Journal of Neuroscience*, 34(49):16408–16416, 2014a.

Ruff, D. A. and Cohen, M. R. Attention can either increase or decrease spike count correlations in visual cortex. *Nature Neuroscience*, 17(11):1591–1597, November 2014b.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ruff, D. A. and Cohen, M. R. Attention increases spike count correlations between visual cortical areas. *Journal of Neuroscience*, 36(28):7523–7534, 2016a.

Ruff, D. A. and Cohen, M. R. Stimulus dependence of correlated variability across cortical areas. *Journal of Neuroscience*, 36(28):7546–7556, 2016b.

Ruff, D. A. and Cohen, M. R. Simultaneous multi-area recordings suggest that attention improves performance by reshaping stimulus representations. *Nature Neuroscience*, 22(10):1669–1676, 2019a.

Ruff, D. A., Xue, C., Kramer, L. E., Baqai, F., and Cohen, M. R. Low rank mechanisms underlying flexible visual representations. *bioRxiv*, page 730978, 2019b.

Rumyantsev, O. I., Lecoq, J. A., Hernandez, O., Zhang, Y., Savall, J., Chrapkiewicz, R., Li, J., Zeng, H., Ganguli, S., and Schnitzer, M. J. Fundamental bounds on the fidelity of sensory cortical coding. *Nature*, 580(7801):100–105, 2020.

Runyan, C. A., Piasini, E., Panzeri, S., and Harvey, C. D. Distinct timescales of population coding across cortex. *Nature*, 548(7665):92–96, 2017.

Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., Yu, B. M., and Batista, A. P. Neural constraints on learning. *Nature*, 512(7515):423, 2014.

Santhanam, G., Yu, B. M., Gilja, V., Ryu, S. I., Afshar, A., Sahani, M., and Shenoy, K. V. Factor-analysis methods for higher-performance neural prostheses. *Journal of neurophysiology*, 102(2): 1315–1330, 2009.

Semedo, J., Zandvakili, A., Kohn, A., Machens, C. K., and Yu, B. M. Extracting latent structure from multiple interacting neural populations. In *Advances in Neural Information Processing Systems*, pages 2942–2950, 2014.

Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M., and Kohn, A. Cortical areas interact through a communication subspace. *Neuron*, 2019.

Shadlen, M. N. and Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience*, 18(10):3870–3896, 1998.

Sharpee, T. O. and Berkowitz, J. A. Linking neural responses to behavior with information-preserving population vectors. *Current Opinion in Behavioral Sciences*, 29:37–44, October 2019.

Smith, M. A. and Kohn, A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *The Journal of Neuroscience*, 28(48):12591–12603, 2008.

Smith, M. A. and Sommer, M. A. Spatial and temporal scales of neuronal correlation in visual area v4. *Journal of Neuroscience*, 33(12):5422–5432, 2013.

Smith, M. A., Jia, X., Zandvakili, A., and Kohn, A. Laminar dependence of neuronal correlations in visual cortex. *Journal of Neurophysiology*, 109(4):940–947, November 2012.

Snyder, A. C., Morais, M. J., and Smith, M. A. Dynamics of excitatory and inhibitory networks are differentially altered by selective attention. *Journal of neurophysiology*, 116(4):1807–1820, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Snyder, A. C., Yu, B. M., and Smith, M. A. Distinct population codes for attention in the absence and presence of visual stimulation. *Nature Communications*, 9(1):4382, 2018.

Sohn, H., Narain, D., Meirhaeghe, N., and Jazayeri, M. Bayesian Computation through Cortical Latent Dynamics. *Neuron*, 103(5):934–947.e5, September 2019.

Solomon, S. S., Chen, S. C., Morley, J. W., and Solomon, S. G. Local and global correlations between neurons in the middle temporal area of primate visual cortex. *Cerebral Cortex*, 25(9): 3182–3196, 2015.

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765):361–365, July 2019a.

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., and Harris, K. D. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437), April 2019b.

Veuthey, T. L., Derosier, K., Kondapavulur, S., and Ganguly, K. Single-trial cross-area neural population dynamics during long-term skill learning. *Nature Communications*, 11(1):4057, August 2020.

Vyas, S., Even-Chen, N., Stavisky, S. D., Ryu, S. I., Nuyujukian, P., and Shenoy, K. V. Neural Population Dynamics Underlying Motor Learning Transfer. *Neuron*, 97(5):1177–1186.e3, March 2018.

Williams, A. H., Kim, T. H., Wang, F., Vyas, S., Ryu, S. I., Shenoy, K. V., Schnitzer, M., Kolda, T. G., and Ganguli, S. Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron*, 98(6):1099–1115.e8, June 2018.

Williamson, R. C., Cowley, B. R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M. A., and Yu, B. M. Scaling properties of dimensionality reduction for neural populations and network models. *PLoS computational biology*, 12(12):e1005141, 2016.

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., and Sahani, M. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology*, 102(1):614–635, 2009.

Zandvakili, A. and Kohn, A. Coordinated neuronal activity enhances corticocortical communication. *Neuron*, 87(4):827–839, 2015.

Zohary, E., Shadlen, M. N., and Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370:140–143, 1994.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Spike count covariance matrix

Both pairwise metrics and population metrics are computed directly from the spike count covariance matrix Σ of size $n \times n$ for a population of n neurons. Each entry in Σ is the covariance between the activity of neuron i and neuron j:

$$\Sigma_{ij} = \mathrm{cov}(x_i, x_j) = \mathrm{E}[(x_i - \mu_i)(x_j - \mu_j)] \tag{2}$$

where x_i and x_j represent the activity of neurons i and j, respectively, and μ_i and μ_j represent the mean activity of neurons i and j, respectively. The variance of the ith neuron is equal to Σ_{ii}.

Pairwise metrics

We computed the spike count correlation (r_{sc}) between neurons i and j directly from the spike count covariance matrix:

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} \tag{3}$$

We then summarized the distribution of r_{sc} values across all pairs of neurons in the population with two pairwise metrics: the r_{sc} mean and r_{sc} standard deviation (s.d.).

Population metrics

The metrics we use for characterizing population-wide covariability are based on factor analysis (FA; Santhanam et al., 2009; Yu et al., 2009; Churchland et al., 2010; Harvey et al., 2012; Williamson et al., 2016; Bittner et al., 2017; Athalye et al., 2017; Huang et al., 2019), a dimensionality reduction method. We chose FA because it is the most basic dimensionality reduction method that explicitly separates variance that is shared among neurons from variance that is independent to each neuron. This allows us to relate the population metrics provided by FA to spike count correlation, which is designed to measure shared variability between pairs of neurons. One might consider using principal component analysis (PCA), but it does not distinguish shared variance from independent variance. Thus, FA is more appropriate than PCA for studying the shared variability among a population of neurons.

Decomposing the spike count covariance matrix

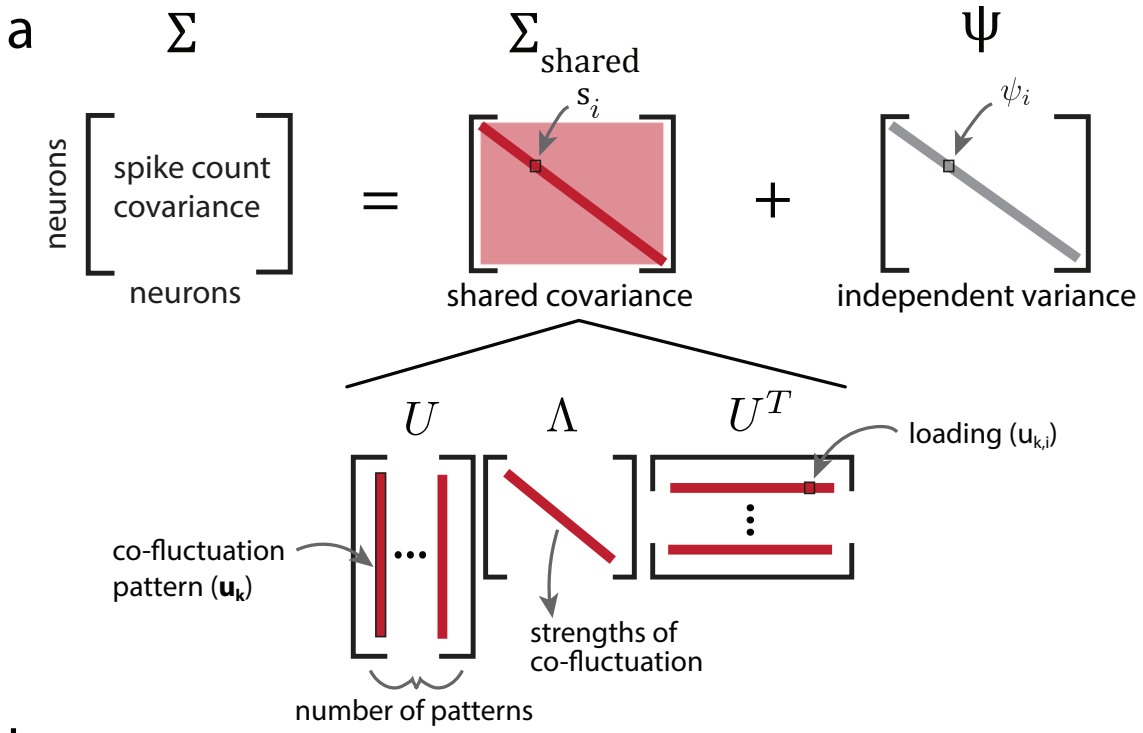
FA decomposes the spike count covariance matrix Σ into a low-rank shared covariance matrix, which captures the variability shared among neurons in the population, and an independent variance matrix, which captures the portion of variance of each neuron unexplained by the other neurons (Fig. 7*a*):

$$\Sigma = \Sigma_{\mathrm{shared}} + \Psi \tag{4}$$

where $\Sigma_{\mathrm{shared}} \in \mathbb{R}^{n \times n}$ is the shared covariance matrix for n neurons, and $\Psi \in \mathbb{R}^{n \times n}$ is a diagonal matrix containing the independent variance of each neuron. The low-rank shared covariance matrix can be expressed using the eigendecomposition as (Fig. 7*a*):

$$\Sigma_{\mathrm{shared}} = U \Lambda U^T \tag{5}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



b

population metric	mathematical definition	intuition
loading similarity	$\text{loading sim.} = 1 - \frac{\text{var}(\mathbf{u_k})}{1\ /\ \text{n}}$	similarity in weights in co-fluctuation pattern $\mathbf{u_k}$
percent shared variance	$\%\text{sv} = \frac{1}{\text{n}} \sum_{i=1}^{\text{n}} \frac{s_i}{s_i+\psi_i} \cdot 100\%$	% of neuron's variability explained by other neurons in the population
dimensionality	$\text{dim.} = \text{rank}(U)$	number of co-fluctuation patterns

Figure 7: Decomposition of the spike count covariance matrix and defining population metrics. **a**. We use factor analysis to decompose the spike count covariance matrix Σ into the sum of a low-rank shared covariance matrix Σ_{shared} and a diagonal independent variance matrix Ψ. The ith diagonal entry of Σ_{shared} (s_i) corresponds to the spike count variance that neuron i shares with other neurons in the population (i.e., shared variance), while the ith diagonal entry of Ψ_i corresponds to spike count variance of neuron i that cannot be explained by the other neurons (i.e., independent to neuron i). We can further decompose Σ_{shared} via an eigendecomposition to extract the co-fluctuation patterns (i.e., the eigenvectors) and the strength of each latent co-fluctuation (i.e., the eigenvalues). **b**. The population metrics used in this study are loading similarity, percent shared variance (%sv), and dimensionality.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where $U \in \mathbb{R}^{n \times d}$ and $\Lambda \in \mathbb{R}^{d \times d}$, with $d < n$. The rank (i.e., dimensionality) of the shared covariance matrix, d, indicates the number of latent variables. Each column of U is an eigenvector and represents a co-fluctuation pattern containing the loading weights of each neuron (i.e., how much each neuron contributes to that dimension). The matrix Λ is a diagonal matrix where each diagonal element is an eigenvalue and represents the amount of variance along the corresponding co-fluctuation pattern (e.g., in Fig. 2 panel *a* has larger eigenvalue than panel *c*).

Based on this matrix decomposition, we defined the three metrics that describe the population-wide covariability:

- **Loading similarity**: the similarity of loading weights across neurons for a given co-fluctuation pattern. Scalar value between 0 (the weights are maximally dissimilar, defined precisely below) and 1 (all weights are the same).
- **Percent shared variance (%sv)**: the percentage of each neuron's variance that is explained by other neurons in the population. Percentage between 0% and 100%.
- **Dimensionality**: the number of dimensions (i.e., co-fluctuation patterns). Integer value.

We give the precise definitions of these population metrics below and in Fig. 7*b*.

Loading similarity

We sought to define loading similarity such that, for a given co-fluctuation pattern, if the weights for all neurons are the same, we would measure a loading similarity of 1. When the weights are as different as possible, we would measure a loading similarity of 0. We define the loading similarity based on the variance across the n weights (for n neurons) in a co-fluctuation pattern $\mathbf{u_k}$. The smallest possible variance is 0; the largest possible variance, for a unit vector $\mathbf{u_k}$, is $1/n$ (Supplementary Math Note F). Thus, we define loading similarity for a co-fluctuation pattern $\mathbf{u_k} \in \mathbb{R}^n$ as:

$$\text{loading similarity}(\mathbf{u_k}) = 1 - \frac{\text{var}(\mathbf{u_k})}{\max_{\mathbf{v_k}} \text{var}(\mathbf{v_k})} = 1 - \frac{\text{var}(\mathbf{u_k})}{1/n} \tag{6}$$

where the loading similarity is computed on unit vectors (i.e., $\mathbf{u_k}$ has a norm of 1). The notation $\text{var}(\mathbf{u_k})$ denotes that the variance is being taken across the n elements of the vector $\mathbf{u_k}$. The denominator of equation (6) acts as a normalizing factor, bounding the loading similarity value between 0 and 1.

The loading similarity distinguishes between a co-fluctuation pattern along which all neurons in the population have the same weight in which case they change their activity up and down together (Fig. 2*a*; loading similarity of 1), from one in which weights are different and some neurons increase their activity when others decrease their activity (Fig. 2*b*; loading similarity of 0). The loading weights we use here are closely related to 'population coupling' (Okun et al., 2015) and 'modulator weights' (Rabinowitz et al., 2015). For some types of shared fluctuations, these weights are similar across neurons in a population (i.e., high loading similarity; Okun et al., 2015; Rabinowitz et al., 2015; Huang et al., 2019). For other types of shared fluctuations, the weights vary substantially across neurons in the population (i.e., low loading similarity; Snyder et al., 2018; Cowley et al., 2020).

We show in Supplementary Math Note E why, if one dimension has high loading similarity, the other dimensions must have low loading similarity. The reason is that co-fluctuation patterns are defined to be mutually orthogonal. If one co-fluctuation pattern has all weights close to the same

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

value (i.e., high loading similarity), then all other co-fluctuation patterns must have substantial diversity in their weights (i.e., low loading similarity) to satisfy orthogonality.

Percent shared variance

The percent shared variance (%sv) measures the percentage of each neuron's spike count variance that is explained by other neurons in the population (Williamson et al., 2016; Bittner et al., 2017; Hennig et al., 2018). Equivalently, we can think of %sv in terms of latent co-fluctuations. Because latent co-fluctuations capture the shared variability among neurons, the %sv measures how much of each neuron's variance is explained by the latent co-fluctuations. The activity of neurons may be tightly linked to the latent co-fluctuation (e.g., Fig. 2*a*), in which case a large percentage of each neuron's variance is shared with other neurons, or may only be loosely linked to the latent co-fluctuation (e.g., Fig. 2*c*), in which case a small percentage of each neuron's variance is shared with other neurons. Mathematically, we define the %sv for a neuron i:

$$\%\text{sv for neuron i} = \frac{\Sigma_{shared,ii}}{\Sigma_{ii}} \cdot 100\% = \frac{s_i}{s_i + \psi_i} \cdot 100\% \tag{7}$$

where s_i is the i^{th} entry along the diagonal of the shared covariance matrix (Fig. 7*a*, Σ_{shared}), and ψ_i is the i^{th} entry along the diagonal of the independent covariance matrix (Fig. 7*a*, Ψ). A %sv of 0% indicates that the neuron does not covary with (i.e., is independent of) other neurons in the population, whereas a %sv of 100% indicates that the neuron's activity can be entirely accounted for by the activity of other neurons in the population. To compute %sv for an entire population of neurons, we averaged the %sv of the individual neurons. All %sv values reported in this study are the %sv for the neuronal population.

Dimensionality

Dimensionality refers to the number of latent co-fluctuations needed to describe population-wide covariability. For example, the population-wide covariability can be described by one latent co-fluctuation (Fig. 2*a*) or by several latent co-fluctuations (Fig. 2*d*). In the population activity space, dimensionality corresponds to the number of axes along which the population activity varies (see Fig. 2*d*, bottom inset). Mathematically, the dimensionality is the rank of the shared covariance matrix (i.e., the number of columns in U, Fig. 7*a*).

Creating the spike count covariance matrices with specified population metrics

To relate pairwise and population metrics, we created spike count covariance matrices of the form in equation (4) with specified population metrics. Importantly, we did not simulate spike counts, nor fit a factor analysis model to simulated data. Rather, we created covariance matrices using (4) and computed pairwise correlations directly from the entries of the covariance matrix, as shown in (3). Across simulations (Figs. 3 and 4), we simulated with $n = 30$ neurons and set independent variances (i.e., diagonal elements of Ψ in equation (4)) to 1.

Specifying co-fluctuation patterns to obtain different loading similarities

Each co-fluctuation pattern $\mathbf{u_k}$ is a vector with $n = 30$ entries (one entry per neuron). We generated a single co-fluctuation pattern by randomly drawing 30 independent samples from a Gaussian distribution with a mean of 2.5. We choose a nonzero mean so that we could obtain co-fluctuation patterns with loading similarities close to 1 when drawing from the Gaussian distribution (i.e., a mean of 0 would have resulted in almost all co-fluctuation patterns having a loading similarity close to 0). To get a range of loading similarities between 0 and 1, we used different standard deviations for the Gaussian. For a small standard deviation value, all entries in the co-fluctuation

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pattern are close to 2.5, resulting in a high loading similarity. For larger standard deviations, some loading weights are positive and some negative, with large variability in their values, resulting in co-fluctuation patterns with low loading similarity. We increased the Gaussian standard deviation from 0.1 to 5.5 with increments of size 0.1. For each increment, we generated 50 patterns and normalized them to have unit norm. In total, we created a set of 2,750 random patterns.

The following procedure describes the construction of shared covariance matrices with one co-fluctuation pattern. We chose a single pattern $\mathbf{u_1} \in \mathbb{R}^{30\times 1}$ (i.e., U has only 1 column) from the set of 2,750. We constructed the shared covariance matrix by computing $U\Lambda U^T$, where Λ was chosen to achieve a desired percent shared variance (see below). The covariance matrix was then computed according to equation (4). We created a covariance matrix, yielding a spread of loading similarities between 0 and 1 (Fig. 3*e-f*). In the next section, we describe the procedure for creating a covariance matrix with more dimensions.

Specifying the percent shared variance

To achieve a given %sv, either the independent variance or the amount of shared variability (i.e., the eigenvalues) of each dimension can be adjusted. In the main text, we set the independent variance of each neuron to $\Psi_i = 1$, and changed the total amount of shared variability by multiplying each eigenvalue (each diagonal element in Λ from equation (5)) by the same constant value, a. To obtain a specified %sv, we identified a by searching through a large set of possible values (from 10^{-4} to 10^3 with step size 10^{-3}). We allowed for a tolerance of $\epsilon = 10^{-3}$ between the desired %sv and the %sv that was achieved after scaling the eigenvalues by a. In Supplementary Figure 4, we allowed the independent variances to be different across neurons, and the results were qualitatively similar to that shown in the main text.

Increasing dimensionality

To assess how changing dimensionality affects pairwise metrics, we created covariance matrices whose shared covariance matrix comprised more than 1 dimension. To create a shared covariance matrix with d dimensions, we randomly chose d patterns from the set of 2750 we had generated above (see 'Specifying co-fluctuation patterns to obtain different loading similarities'). We then orthogonalized the chosen patterns using the Gram-Schmidt process to obtain d orthonormal (i.e., orthogonal and unit length) co-fluctuation patterns $U \in \mathbb{R}^{30\times d}$. We formed the shared covariance matrix using $U\Lambda U^T$, where $\Lambda \in \mathbb{R}^{d\times d}$ is a diagonal matrix containing the eigenvalues (i.e., the strength of each dimension; see 'Specifying the relative strengths of each dimension' below). We repeated this procedure to produce 3,000 sets of d orthonormal patterns (i.e., 3,000 different U matrices), each of which was used to create a shared covariance matrix. The spike count covariance was computed according to equation (4).

Specifying the relative strengths of each dimension

In simulating shared covariance matrices with more than one dimension, we chose the relative strength of each dimension by specifying the eigenspectrum (diagonal elements of Λ in equation (5)). We worked with three sets of eigenspectra. First, a flat eigenspectrum had eigenvalues that were all equal (Fig. 3*g*). Second, for two dimensions, we varied the ratio of the two eigenvalues between 95:5, 80:20, 50:50, 20:80, and 5:95 (Fig. 4). Third, we considered an eigenspectrum in which each subsequent eigenvalue falls off according to an exponential function (Supplementary Fig. 5). Only the relative (and not the absolute) eigenvalues (i.e., the shape of the eigenspectrum) affect the results, because the eigenspectrum was subsequently scaled to achieve a desired %sv (see 'Specifying the values of percent shared variance').

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Analysis of V4 neuronal recordings from a spatial attention task

Electrophysiological recordings

We analyzed data from a visual spatial attention task reported in a previous study (Snyder et al., 2018). Briefly, we implanted a 96-electrode "Utah" array (Blackrock Microsystems; Salt Lake City, UT) into visual cortical area V4 of an adult male rhesus macaque monkey (data from two monkeys were analyzed; in our study, monkey 1 corresponds to "monkey P" and monkey 2 corresponds to "monkey W" from Snyder et al. (2018)). After recording electrode voltages (Ripple Neuro.; Salt Lake City, UT), we used custom software to perform off-line spike sorting (Kelly et al., 2007, freely available at `https://github.com/smithlabvision/spikesort`). This yielded 93.2 ± 8.9 and 61.9 ± 27.4 candidate units per session for monkey 1 and 2, respectively.

To further ensure the isolation quality of recorded units, we removed units from our analyses according to the following criteria. First, we removed units with a signal-to-noise ratio of the spike waveform less than 2.0 (Kelly et al., 2007). Second, we removed units with overall mean firing rates less than 1 Hz, as estimates of r_{sc} for these units tends to be poor (Cohen and Kohn, 2011). Third, we removed units that had large and sudden changes in activity due to unstable recording conditions. For this criterion, we divided the recording session into ten equally-sized blocks and for each unit computed the difference in average firing rate between adjacent blocks. We excluded units with a change in average firing rate greater than 60% of the maximum firing rate (where the maximum is taken across the ten equally-sized blocks). Fourth, we removed an electrode from each pair of electrodes that were likely electrically-coupled. We identified the coupled electrodes by computing the fraction of threshold crossings that occurred within 100 μs of each other for each pair of electrodes. We then removed the fewest number of electrodes to ensure this fraction was less than 0.2 (i.e., pairs with an unusually high number of coincident spikes) for all pairs of electrodes. Fifth, we removed units that did not sufficiently respond to the visual stimuli used in the experiment. Evoked spike counts (i.e., a neuron's response after stimulus presentation) were taken between 50 ms to 250 ms after stimulus onset, and spontaneous spike counts (i.e., a neuron's response during a blank screen) were taken in a 200 ms window that ended 50 ms before stimulus onset. For each unit, we computed a sensitivity measure d' between evoked and spontaneous activity:

$$d' = \frac{\mu_{\text{evoked}} - \mu_{\text{spontaneous}}}{\sqrt{\frac{1}{2}(\sigma^2_{\text{evoked}} + \sigma^2_{\text{spontaneous}})}}$$

for mean spike counts μ_{evoked} and $\mu_{\text{spontaneous}}$ and spike count variances σ^2_{evoked} and $\sigma^2_{\text{spontaneous}}$. We removed units with $d' < 0.5$ from analyses, as these units had spontaneous and evoked responses that were difficult to distinguish.

After applying these five criteria, 44.5 ± 11.3 and 18.8 ± 6.7 units per session (mean ± s.d. over sessions) remained for monkeys 1 and 2, respectively. Although these remaining units likely contained both single-unit and multi-unit activity, we refer to each unit as a neuron for simplicity.

Visual stimulus change-detection task

Animals were trained to perform a change-detection task with a spatial attention cue to the location of the visual stimulus that was more likely to change (Snyder et al., 2018). In the visual change-detection task (Fig. 6*a*), animals fixated a central dot while Gabor stimuli were presented in two locations on a computer screen. One location was chosen to be within the aggregate receptive fields (RFs) of the recorded V4 neurons (mapped prior to running the experiment), and the other location was placed at the mirror symmetric location in the opposite hemifield. Animals maintained fixation while a sequence of Gabor stimuli were presented. Each drifting Gabor stimulus (oriented at either 45° or 135°) was presented for 400 ms, followed by a blank screen presented for a random

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

interval (between 300 and 500 ms). The sequence continued, with a fixed probability for each presentation, until one of the two stimuli changed orientation when presented (i.e., the 'target'). Upon target presentation, animals were required to make a saccade to the target to earn a juice reward. We manipulated spatial attention in the experiment by cueing the more probable target location in blocks. At the beginning of each block, the cue was denoted by presenting only one Gabor stimulus at the more probable target location (90% likely), and requiring animals to detect orientation changes at this location for 5 trials. Consistent with the results of previous studies, we found that animals had greater perceptual sensitivity for orientation changes at the cued (i.e., attended) location than the uncued location (Fig. 6*a*, inset in the bottom right) and shorter reaction times (Snyder et al., 2018).

Data processing and computing spike counts

We first separated the trials into two groups: (1) "attend in" trials, for which the cued stimulus was inside the recorded neurons' RFs and (2) "attend out" trials, for which the cued stimulus was outside the RFs. Since the initial orientation of the stimulus at the cued location could be one of two values (i.e., 45° or 135°), we further divided trials, resulting in a total of 4 groups of trials per session (attend in & 45°, attend out & 45°, attend in & 135°, attend out & 135°). Each combination of cued location and stimulus orientation was treated as an independent sample. The same neurons were used for each of the 4 groups within each session, ensuring a fair comparison between the attend-in and attend-out conditions.

We analyzed all stimulus presentations for which the target stimulus did not change. For each stimulus presentation, we took spike counts in a 200 ms window starting 150 ms after stimulus onset. For each of the 4 groups, we formed a spike count matrix $X \in \mathbb{R}^{n \times t}$, containing the spike counts of the n recorded neurons for the t trials belonging to that group. These spike count matrices were then used to compute both the pairwise and population metrics (described below). For all analyses (Fig. 6), we excluded recording sessions with fewer than 10 neurons. Additionally, because population metrics depend on the number of trials (Williamson et al., 2016), for each session we equalized the number of trials across the 4 groups by randomly subsampling from groups with larger numbers of trials.

Computing pairwise metrics for V4 spike counts

We computed pairwise metrics on each combination of attention state ('attend in' and 'attend out') and stimulus orientation. We computed the correlation matrix for X as described above in 'Pairwise metrics' and then computed r_{sc} mean and r_{sc} s.d. For each attention state, we averaged the r_{sc} mean and r_{sc} s.d. over sessions and different stimulus orientations.

Computing population metrics for V4 spike counts

We fit the parameters of a factor analysis model (see Fig. 7*a*) to each spike count matrix X (as described above) using the expectation-maximization (EM) algorithm (Dempster et al., 1977). For each session, this was performed separately for each attention state and stimulus orientation. Using the FA parameters, we then computed the three population metrics (Fig. 7*b*). For dimensionality, we first found the number of dimensions d that maximized the cross-validated data likelihood. We fit an FA model with d dimensions, and then found the number of dimensions required to explain 95% of the shared variance, termed d_{shared} (Williamson et al., 2016). We report d_{shared} because it tends to be a more reliable estimate of dimensionality than the number of dimensions that maximizes the cross-validated data likelihood. We computed %sv as described by equation (7). We report the loading similarity as defined in equation (6) for the co-fluctuation pattern that explained the most shared variability (i.e., the eigenvector with the largest eigenvalue), since it contributes most to describing the population-wide covariability. For 'attend in' and 'attend out'

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conditions, we averaged the population metrics across sessions and stimulus orientations.

Statistics

We employed paired permutations tests for all statistical comparisons of pairwise metrics and population metrics between 'attend-in' and 'attend-out' conditions (Fig. 6*b*-*c*). First, for a given, metric we computed the average difference between attend-in and attend-out. Then, we computed a null distribution by randomly permuting attend-in and attend-out labels and recomputing the average difference in the permuted data. We ran 10,000 permutations to obtain a null distribution of 10,000 samples. We computed p-values as the proportion of samples in the null distribution that were more extreme than the average difference in the data, corresponding to $p < 0.0001$ as the highest attainable level of significance in our statistical analyses.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Math Notes

A Relationship between correlation, loading similarity, and %sv (one latent dimension)

We establish here the mathematical relationship between r_{sc}, loading similarity, and %sv. This will provide the formalism for understanding why decreasing %sv decreases both r_{sc} mean and s.d. (Fig. 3*f*), that a high loading similarity corresponds to large r_{sc} mean and low r_{sc} s.d. (Fig. 3*e*), and that a low loading similarity corresponds to small r_{sc} mean and large r_{sc} s.d. (Fig. 3*e*).

Let n be the number of neurons, and let $\mathbf{w}$ be the co-fluctuation pattern (i.e., loading vector $[w_1, w_2, ..., w_n]^T \in \mathbb{R}^{n\times 1}$), $\lambda \in \mathbb{R}_+$ be the strength of the co-fluctuation pattern (i.e., eigenvalue of the shared covariance matrix), and $\Psi \in \mathbb{R}^{n\times n}$ be a diagonal matrix specifying the independent variance of each neuron $(\psi_1, \psi_2, ..., \psi_n)$. Then the covariance matrix of the population activity is (see Methods and Fig. 7):

$$\Sigma = \Sigma_{shared} + \Psi = \mathbf{w}\lambda\mathbf{w}^T + \Psi$$

From this, we observe that $\Sigma_{ij} = \Sigma_{shared,ij} = \lambda w_i w_j$ on the off-diagonal entries (i.e., if $i \neq j$). Along the diagonals, $\Sigma_{shared,ii} = \lambda w_i^2$ and $\Sigma_{ii} = \lambda w_i^2 + \psi_i$. The correlation (i.e., r_{sc} if Σ is a spike count covariance matrix) between neurons i and j can be written as:

$$\begin{aligned}\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} &= \frac{\lambda w_i w_j}{\sqrt{(\lambda w_i^2 + \psi_i)(\lambda w_j^2 + \psi_j)}} \\ &= \sqrt{\frac{\lambda w_i^2}{\lambda w_i^2 + \psi_i}}\sqrt{\frac{\lambda w_j^2}{\lambda w_j^2 + \psi_j}}\,\text{sign}(w_i w_j) \\ &= \sqrt{\phi_i \phi_j}\,\text{sign}(w_i w_j)\end{aligned} \tag{8}$$

where ϕ_i and ϕ_j represent the %sv (as proportions) for neurons i and j, respectively, and $\text{sign}(w_i w_j) = +1$ if $w_i w_j > 0$ or -1 if $w_i w_j < 0$. The last line follows from the fact that %sv is defined in equation (7) as:

$$\phi_i = \frac{\Sigma_{shared,ii}}{\Sigma_{ii}} = \frac{\lambda w_i^2}{\lambda w_i^2 + \psi_i} \tag{9}$$

Equations (8) and (9) provide a basis for understanding the relationships between r_{sc}, %sv, and loading similarity. The r_{sc} mean and s.d. are computed across all pairs of neurons ρ_{ij}, for $i < j$.

For establishing a relationship between pairwise metrics and %sv, consider decreasing the overall %sv of the population, while keeping the loadings w_i fixed. This corresponds to decreasing λ in equation (9), which implies ϕ_i for each neuron decreases, and thus the product $\sqrt{\phi_i \phi_j}$ decreases for all pairs. The magnitude of each ρ_{ij} decreases (i.e., each ρ_{ij} moves closer to 0). As such, decreasing %sv of the population decreases the distance of a point from the origin in the r_{sc} mean versus r_{sc} s.d. plot, all else being equal (Fig. 3*f*).

For establishing a relationship between pairwise metrics and loading similarity, consider two extreme cases: 1) when loading similarity is 1 (as high as possible) 2) when it is 0 (as low as possible). We first assume that each neuron has the same independent variance ψ_i for simplicity, as we did in Figure 3. A loading similarity of 1 corresponds to each $w_i = +\frac{1}{\sqrt{n}}$ or each $w_i = -\frac{1}{\sqrt{n}}$. In either case, $\text{sign}(w_i w_j)$ is always $+1$. Furthermore, ϕ_i is the same for every neuron and $\sqrt{\phi_i \phi_j} = $ %sv

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(i.e., the %sv of the population, expressed as a proportion) for every pair of neurons. Thus, all ρ_{ij} = %sv for all pairs of neurons i and j. In this case, r_{sc} mean = %sv and r_{sc} s.d. = 0. If the independent variances ψ_i are different across neurons, we can still get each sign$(w_i w_j) = +1$ and each ϕ_i to be the same by setting each $w_i = +\sqrt{\psi_i}$ or each $w_i = -\sqrt{\psi_i}$. This would also result in ρ_{ij} = %sv for all pairs of neurons i and j, and thus r_{sc} mean = %sv and r_{sc} s.d. = 0. In this case, the loading similarity is still high (all w_i are the same sign; we can show that load. sim.> 0.5), but not equal to 1.

Now, consider a scenario in which half the loadings are $+\frac{1}{\sqrt{n}}$ and the other half are $-\frac{1}{\sqrt{n}}$ (and assume again that ψ_i are the same for every neuron). This is one way to obtain a loading similarity of 0. In this case, ϕ_i are still the same for every neuron, so $\sqrt{\phi_i \phi_j}$ = %sv for all pairs. However, sign$(w_i w_j) = -1$ for $\left(\frac{n}{2}\right)^2 = \frac{n^2}{4}$ pairs, and sign$(w_i w_j) = +1$ for $2 \times \binom{n/2}{2} = \frac{n^2}{4} - \frac{n}{2}$ pairs. We can show that r_{sc} mean $= -\frac{\%sv}{n-1}$ and, by using equation (10) from Supplementary Math Note B below, r_{sc} s.d. = %sv $\sqrt{1 - \frac{1}{(n-1)^2}}$. Thus, for a large number of neurons n, this case (where loading similarity=0) corresponds to small negative r_{sc} mean (close to 0), and large r_{sc} s.d. (close to the %sv). As an example, for 30 neurons and %sv=50%, this corresponds to r_{sc} mean = -0.0172 and r_{sc} s.d. = 0.4997.

With this analysis, we have established that for one latent dimension:

- Decreasing %sv decreases the magnitudes of correlations (i.e., each ρ_{ij} closer to 0). r_{sc} mean and s.d. both decrease (as seen empirically in Fig. 3*f*).
- Starting from a loading similarity near 1, a decrease in loading similarity involves flips in the signs of some correlations (i.e., some ρ_{ij} become $-\rho_{ij}$). r_{sc} mean decreases but r_{sc} s.d. increases (as seen empirically in Fig. 3*f*).
- Both r_{sc} mean and %sv measure shared variance among neurons, but they are not always equal. Equations (8) shows that the two quantities are equal if all sign$(w_i w_j)$ are the same (i.e., when loading similarity is high). However, in general r_{sc} mean and shared variance (%sv) are not the same—e.g., when loading similarity is low, or when there are multiple dimensions (Supplementary Math Note C).

In this section, we consider the extremes of loading similarity. In the next section, we analyze how gradual changes in loading similarity affect r_{sc} mean and s.d. for a fixed %sv.

B Circular arc in r_{sc} mean versus r_{sc} s.d. plot for one latent dimension and fixed %sv

We establish here mathematically that gradually varying the loading similarity for one latent dimension and fixed %sv results in an arc-like relationship between r_{sc} mean and r_{sc} s.d., and that the radius of the arc is approximately equal to the %sv (Fig. 3*e-f*).

We use the same notation as in Supplementary Math Note A. Let $E[.]$ and $Var(.)$ denote the mean and variance across all neurons or all pairs of neurons, depending on context. In particular, we are interested in $E[\rho] = r_{sc}$ mean, $\sqrt{Var(\rho)} = r_{sc}$ s.d., where the expectation and variance are computed across ρ_{ij} for all pairs of neurons in a given population (i.e., the upper triangle of the correlation matrix, ρ_{ij} for $i > j$).

Let c be the distance of a point (corresponding to one instance of the population activity covariance matrix) from the origin in the r_{sc} mean versus r_{sc} s.d. plot (i.e., $c = \sqrt{(r_{sc}\text{ mean})^2 + (r_{sc}\text{ s.d.})^2}$). We want to know whether c is the same for all population activity covariance matrices with one

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

latent dimension and fixed %sv. This would correspond to point being equidistant from the origin, and thus a circular arc. We can write c as:

$$\begin{aligned} c^2 &= (r_{\text{sc}} \text{ mean})^2 + (r_{\text{sc}} \text{ s.d.})^2 \\ &= E\left[\rho\right]^2 + Var\left(\rho\right) \\ &= E\left[\rho\right]^2 + E\left[\rho^2\right] - E\left[\rho\right]^2 \\ &= E\left[\rho^2\right] \end{aligned}$$

Thus, the squared distance (i.e., squared radius) is equal to $E\left[\rho^2\right]$, the mean of ρ_{ij}^2 across all pairs in the population. Let m be the number of pairs (i.e., $m = \binom{n}{2} = \frac{n(n-1)}{2}$). Now, using equations (8) and (9) derived in Supplementary Math Note A:

$$\begin{aligned} E\left[\rho^2\right] &= \frac{1}{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\rho_{ij}^2 \\ &= \frac{1}{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{(\lambda w_i^2)(\lambda w_j^2)}{\left(\lambda w_i^2+\psi_i\right)\left(\lambda w_j^2+\psi_j\right)} \\ &= \frac{1}{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_i\phi_j \end{aligned}$$

where ϕ_i and ϕ_j are the %sv of neurons i and j (expressed as proportions), as defined in Supplementary Math Note A. We can show that $2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_i\phi_j = \sum_{i=1}^{n}\sum_{j=1}^{n}\phi_i\phi_j - \sum_{i=1}^{n}\phi_i^2$. Intuitively, if we have a symmetric matrix Φ with entries $\Phi(i,j) = \phi_i\phi_j$, and we want to find the sum of the off-diagonal elements ($2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_i\phi_j$), then we can take the sum of all elements and subtract the diagonal elements ($\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_i\phi_j - \sum_{i=1}^{n}\phi_i^2$). Using this equivalence, it follows:

$$\begin{aligned} E\left[\rho^2\right] &= \frac{1}{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_i\phi_j \\ &= \frac{1}{2m}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_i\phi_j - \sum_{i=1}^{n}\phi_i^2\right) \\ &= \frac{1}{2m}\left(\sum_{i=1}^{n}\phi_i\sum_{j=1}^{n}\phi_j - \sum_{i=1}^{n}\phi_i^2\right) \\ &= \frac{1}{2m}\left(n^2E\left[\phi\right]^2 - \sum_{i=1}^{n}\phi_i^2\right) \\ &= \frac{1}{n-1}\left(nE\left[\phi\right]^2 - E\left[\phi^2\right]\right) \\ &= \frac{1}{n-1}\left(nE\left[\phi\right]^2 - Var\left(\phi\right) - E\left[\phi\right]^2\right) \\ &= \frac{1}{n-1}\left((n-1)E\left[\phi\right]^2 - Var\left(\phi\right)\right) \end{aligned}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$
\begin{aligned}
&= E\left[\phi\right]^2 - \frac{1}{n-1} Var\left(\phi\right) \\
&= \left(\%sv\right)^2 - \frac{1}{n-1} Var\left(\phi\right)
\end{aligned} \tag{10}
$$

This provides an equation for the squared radius (i.e., squared distance from the origin) of a point in the r_{sc} mean versus r_{sc} s.d. plot. In the above derivation, $E\left[\phi\right]$ and $Var\left(\phi\right)$ are taken across the percent shared variance of each neuron in the population ϕ_i. Thus, $E\left[\phi\right]$ is equal to our population metric %sv. Now, we will bound $Var\left(\phi\right)$, which by definition is greater than or equal to 0. Since $0 \leq \phi_i \leq 1$, one instance where the maximum variance occurs is when there are an equal number of $\phi_i = 0$ and $\phi_i = 1$ (and $E\left[\phi\right] = 0.5$). Then,

$$
\begin{aligned}
Var\left(\phi\right) &= \frac{1}{n} \sum_{i=1}^{n} \left(\phi_i - 0.5\right)^2 \\
&= \frac{1}{n} \left(\frac{n}{2}(1-0.5)^2 + \frac{n}{2}(0-0.5)^2\right) \\
&= \frac{1}{n} \left(0.25n\right) \\
&= 0.25
\end{aligned}
$$

So $0 \leq Var\left(\phi\right) \leq 0.25$. For a small number of neurons n, the second term is non-negligible. For example, for a model with 6 neurons and %sv = 50%, the radius of the data points may vary between 0.4472 and 0.5. As the number of neurons increases, the second terms becomes negligible, and data points lie approximately along an arc with radius equal to %sv. For example, for 30 neurons as in our simulations and a %sv of 50%, the radius only varies between 0.4913 and 0.5.

To summarize, equation (10) computes the distance from the origin of a point for a given population of neurons. For a fixed %sv, $Var\left(\phi\right)$ can be the same or differ across many simulation runs. If $Var\left(\phi\right) = 0$ or is the same across runs, then the points will lie perfectly along an arc, with radius specified by equation (10). However, if $Var\left(\phi\right)$ is different across runs, the distances of each point from the origin will differ slightly, so they will lie close to, but not exactly along, an arc.

With this analysis, we have shown that in the case of one latent dimensions:

- A point (i.e., corresponding to a given population of neurons, simulated or real) on the r_{sc} mean versus r_{sc} s.d. plot has distance from the origin (i.e., radius) less than or equal to %sv.
- If the %sv for individual neurons (ϕ_i) are all the same (see Supplementary Math Note A), then the radius equals %sv.
- As the number of neurons increases, the radius becomes asymptotically closer to %sv.

C Relationship between correlation, loading similarity, and %sv (multiple latent dimensions)

In Supplementary Math Note A, we established a mathematical relationship between r_{sc}, loading similarity, and %sv in the case of one latent dimension. Here, we generalize equation (8) to include multiple dimensions in order to better understand the relationship between r_{sc} and dimensionality. We demonstrate here that the general relationships between r_{sc}, %sv, and loading similarity for one latent dimension also hold true for multiple latent dimensions. For multiple latent dimensions, the relative strengths of each dimension is an important consideration—a stronger dimension plays a bigger role in determining the r_{sc} distribution. Finally, we consider the relationship between

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dimensionality itself and r_{sc}. We will discover below that increasing dimensionality tends to decrease the magnitude of r_{sc} values.

First, consider the case of two latent dimensions. Again, let n be the number of neurons, let $\mathbf{w}$ be the co-fluctuation pattern (i.e., loading vector $[w_1, w_2, ..., w_n]^T \in \mathbb{R}^{n\times 1}$) with eigenvalue λ_w, let $\mathbf{v}$ be another pattern orthogonal to $\mathbf{w}$ ($[v_1, v_2, ..., v_n]^T \in \mathbb{R}^{n\times 1}$; $\mathbf{v} \perp \mathbf{w}$), with eigenvalue λ_v, and let $\Psi \in \mathbb{R}^{n\times n}$ be a diagonal matrix specifying the independent variance of each neuron $(\psi_1, \psi_2, ..., \psi_n)$. Then the covariance is $\Sigma = \Sigma_{shared} + \Psi = \Sigma_w + \Sigma_v + \Psi = \mathbf{w}\lambda_w\mathbf{w}^T + \mathbf{v}\lambda_v\mathbf{v}^T + \Psi$. On the off-diagonals entries (i.e., if $i \neq j$), $\Sigma_{ij} = \lambda_w w_i w_j + \lambda_v v_i v_j$. Along the diagonals, $\Sigma_{shared,ii} = \Sigma_{w,ii} + \Sigma_{v,ii} = \lambda_w w_i^2 + \lambda_v v_i^2$ and $\Sigma_{ii} = \lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i$.

Because the shared covariance matrix Σ_{shared} can be expressed as a sum of two component matrices $\Sigma_w + \Sigma_v$, we can express the %sv of neuron i (ϕ_i) as

$$\begin{aligned}\phi_i = \frac{\Sigma_{shared,ii}}{\Sigma_{ii}} &= \frac{\Sigma_{w,ii}}{\Sigma_{ii}} + \frac{\Sigma_{v,ii}}{\Sigma_{ii}} \\ &= \frac{\lambda_w w_i^2}{\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i} + \frac{\lambda_v v_i^2}{\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i} \\ &= \phi_i^{(w)} + \phi_i^{(v)}\end{aligned}$$

where $\phi_i^{(w)}$ is the %sv variance of neuron i explained by dimension $\mathbf{w}$ and $\phi_i^{(w)}$ is the %sv variance of neuron i explained by dimension $\mathbf{v}$.

With this decomposition of ϕ_i, and following similar steps as in equation (8):

$$\rho_{ij} = \sqrt{\phi_i^{(w)}\phi_j^{(w)}}\,\text{sign}(w_i w_j) + \sqrt{\phi_i^{(v)}\phi_j^{(v)}}\,\text{sign}(v_i v_j) \tag{11}$$

where %sv values (ϕ) are represented as proportions. Equation (11) relates r_{sc}, %sv, and loading similarity for the case of two latent dimensions. Next, we compare these relationships for one versus two latent dimensions.

We will show that, for two latent dimensions, the relative strength of each dimension (i.e., the ratio $\lambda_w : \lambda_v$) is an important consideration. For two latent dimensions, decreasing the overall %sv by decreasing both $\phi^{(w)}$ and $\phi^{(v)}$ equally (e.g., $\lambda_w = \lambda_v$ and both decrease equally) pushes each ρ_{ij} closer to 0–r_{sc} mean and s.d. will decrease. This is similar to what happens for one latent dimension when %sv is decreased. On the other hand, even if the overall %sv is held constant, but $\phi^{(w)}$ increases relative to $\phi^{(v)}$ (i.e., increase the strength of $\mathbf{w}$ relative to $\mathbf{v}$), pairwise correlations could change. Each ρ_{ij} will largely be determined by $\phi^{(w)}$ and $\mathbf{w}$—r_{sc} mean and s.d. will be more similar to what they would be if only $\mathbf{w}$ existed (Fig. 4*a*). In other words, each ρ_{ij} for two latent dimensions is the sum of the ρ_{ij} that would have been produced by each of the two constituent dimensions on their own. The dimension with larger relative strength λ will have larger ϕ; the stronger dimension will play a larger role in determining each value of ρ_{ij} and thus the resulting r_{sc} distribution.

Using this logic, we can deduce that increasing the loading similarity of one of the dimensions would increase r_{sc} mean and decrease r_{sc} s.d. for the same reasons as for one latent dimension (Supplementary Math Note A). Doing so for a relatively stronger dimension would result in larger changes in r_{sc} than doing so for a relatively weaker dimension.

We have shown how having multiple latent dimensions can affect the relationship between r_{sc}, %sv, and loading similarity. Now, we show that dimensionality itself and r_{sc} are related—larger dimensionality tends to decrease r_{sc} mean and s.d. To see this, we can generalize equation (11) for $d < n$ orthogonal latent dimensions $\mathbf{u_1}, \ldots, \mathbf{u_d} \in \mathbb{R}^n$.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$\rho_{ij} = \sum_{k=1}^{d} \sqrt{\phi_i^{(u_k)} \phi_j^{(u_k)}} \ \text{sign}(u_{k_i} u_{k_j})$$

Considering the sign of one term, ρ_{ij} could have the same sign for $\text{sign}(u_{k_i} u_{k_j})$ across all dimensions $\mathbf{u_1}, \ldots, \mathbf{u_d}$; in this case, larger dimensionality acts to increase the correlation between neurons i and j (ρ_{ij}) above the level corresponding to a single dimension. However, because the loading vectors $\mathbf{u_1}, \ldots, \mathbf{u_d}$ are orthogonal, a pair of neurons i and j is likely to have many $\text{sign}(u_{k_i} u_{k_j})$ of opposite sign across dimensions; in this case, larger dimensionality pushes the correlation between neurons i and j (ρ_{ij}) closer to 0. Thus, we would expect the magnitude of correlations to decrease as more dimensions are added (i.e., a tendency for r_{sc} mean and s.d. to decrease; Fig. 3g). In the next section, we show this relationship mathematically.

D Increasing dimensionality decreases arc radius

We establish here that increasing dimensionality results in a decrease in the radius of the arc in the r_{sc} mean versus r_{sc} s.d. plot (Fig. 3g). We extend the math for an arc for one latent dimension (Supplementary Math Note B) to multiple latent dimensions. We will refer to the one latent dimension as the '1-d case' and multiple (k) latent dimensions as the 'k-d case'.

We use the same notation as in Supplementary Math Note C. Consider the distance c of a point (corresponding to one instance of the population activity covariance matrix) from the origin in the r_{sc} mean versus r_{sc} s.d. plot. From Supplementary Math Note B, $c^2 = E[\rho^2]$. For this 2-d case, the correlation between neurons i and j is $\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} = \frac{\lambda_w w_i w_j + \lambda_v v_i v_j}{\sqrt{\left(\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i\right)\left(\lambda_w w_j^2 + \lambda_v v_j^2 + \psi_j\right)}}$.

Thus we can write ρ_{ij}^2 as:

$$\begin{aligned}
\rho_{ij}^2 &= \frac{(\lambda_w w_i w_j + \lambda_v v_i v_j)^2}{\left(\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i\right)\left(\lambda_w w_j^2 + \lambda_v v_j^2 + \psi_j\right)} \\
&= \frac{\lambda_w^2 w_i^2 w_j^2 + \lambda_w \lambda_v 2 w_i w_j v_i v_j + \lambda_v^2 v_i^2 v_j^2}{\left(\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i\right)\left(\lambda_w w_j^2 + \lambda_v v_j^2 + \psi_j\right)} \\
&= \phi_i \phi_j - \frac{\lambda_w \lambda_v \left(w_i^2 v_j^2 - 2 w_i w_j v_i v_j + w_j^2 v_i^2\right)}{\left(\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i\right)\left(\lambda_w w_j^2 + \lambda_v v_j^2 + \psi_j\right)} \\
&= \phi_i \phi_j - \frac{\lambda_w \lambda_v (w_i v_j - w_j v_i)^2}{\left(\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i\right)\left(\lambda_w w_j^2 + \lambda_v v_j^2 + \psi_j\right)}
\end{aligned}$$

where the % shared variance of neuron i in this 2-d case is $\phi_i = \frac{\Sigma_{shared,ii}}{\Sigma_{ii}} = \frac{\lambda_w w_i^2 + \lambda_v v_i^2}{\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i}$.

Then letting m is the number of pairs in the population, and following similar steps to (10) in Supplementary Math Note B, we arrive at:

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$
\begin{aligned}
E[\rho^2] &= \frac{1}{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\rho_{ij}^2 \\
&= (\%sv)^2 - \frac{1}{n-1}Var\left(\phi\right) - \frac{1}{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\lambda_w\lambda_v(w_iv_j - w_jv_i)^2}{\left(\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i\right)\left(\lambda_w w_j^2 + \lambda_v v_j^2 + \psi_j\right)}
\end{aligned}
\tag{12}
$$

Not including the negative sign in front, note that this final term is non-negative (given that λ_w and λ_v are non-negative, as for any covariance matrix). Thus, comparing the final line in equation (12) to the final line from equation (10), we observe that the distance of the point for the 2-d case in the r_{sc} mean versus r_{sc} s.d. plot is necessarily smaller than or equal to the distance for the corresponding 1-d case.

More generally, for a k-dimensional case we can show that:

$$
\begin{aligned}
E[\rho^2] =& (\%sv)^2 - \frac{1}{n-1}Var\left(\phi\right) \\
&- \frac{1}{m}\sum_{w,v}\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\lambda_w\lambda_v(w_iv_j - w_jv_i)^2}{\left(\lambda_w w_i^2 + \lambda_v v_i^2 + \psi_i\right)\left(\lambda_w w_j^2 + \lambda_v v_j^2 + \psi_j\right)}\right]
\end{aligned}
\tag{13}
$$

where the sum $\sum_{w,v}$ is taken over all unique pairs of loading vectors (w, v). Indeed, as more latent dimensions are subsequently added, the radius of the r_{sc} mean versus r_{sc} s.d. plot decreases (Fig. 3*g*). Intuitively, this final term accounts for how population activity covaries along many different dimensions in the high-d firing rate space. As more *orthogonal* dimensions are added, population activity is further pulled in different directions in the high-d space, more interaction terms come into play, and the magnitude of correlations is further decreased. This tends to decrease both r_{sc} mean and r_{sc} s.d., explaining why the radius of the arc in the r_{sc} mean versus r_{sc} s.d. plot tends to decrease as dimensionality increases.

We note that r_{sc} mean and r_{sc} s.d. do not necessarily *both* need to decrease. For example, consider a pattern with a loading similarity of 1; loading weights for all neurons would have the same value, r_{sc} across all pairs would be the same value, and thus r_{sc} s.d. would be 0 (see Supplementary Math Note A). When a second pattern of necessarily low loading similarity (see Supplementary Math Note E) is added, r_{sc} values across pairs of neurons would differ, and r_{sc} s.d. would be larger than 0. Therefore, r_{sc} s.d. can increase when going from the 1-d case to the 2-d case. However, the corresponding decrease in r_{sc} mean would be larger in magnitude than the increase in r_{sc} s.d., resulting in an overall decrease in arc radius (Fig. 3*g*, 1 to 2 dimensions, data points closest to the horizontal axis).

The third term in equation (13) can also help explain variability of the radius ($E[\rho^2]$) across different random instantiations with the same population metrics (Figs. 3*g* and 4). Consider a fixed %sv. For the 1-d case, the radius is determined by the first two terms of the above equation, and any variability in radius will be caused by different values of $Var(\phi)$ across different instantiations. For the 2-d case, the third term also plays a factor in determining the radius, and this term varies across different random instantiations, typically to a larger degree than the second term for large numbers of neurons n (see Supplementary Math Note B). Thus, the 2-d and k-d cases have greater variability in $E[\rho^2]$ than 1-d cases (Fig. 3*g*, Fig. 4). Other subtle factors can affect the variability of $E[\rho^2]$. For example, variability in $E[\rho^2]$ can increase or decrease depending on the relative strengths of each dimension and their corresponding loading similarities (Fig. 4 and Supplementary Fig. 5).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This can be explained by the third component of equation (13), in particular by the terms involving λ_w and λ_v.

E Properties of loading similarities across different co-fluctuation patterns

We asked whether there was a relationship between the loading similarities of different co-fluctuation patterns in the same model. In our simulations and V4 data analysis, we ensured that we obtain unique co-fluctuation patterns by constraining dimensions to be orthogonal. Thus, we might conjecture that if one pattern has high loading similarity (e.g., $[1, \ldots, 1]$), then another pattern in the same model necessarily has low loading similarity (e.g., $[1, -1, 1, -1, \ldots, -1, 1]$). Indeed, this is true because the sum across the loading similarities of each pattern in a model is at most 1. We show this property of loading similarity here.

Let $\mathbf{w}$ and $\mathbf{v}$ be vectors representing two co-fluctuation patterns in the same model. We use the notation $\mathbf{w} \cdot \mathbf{v}$ to refer to the element-wise product between $\mathbf{w}$ and $\mathbf{v}$, resulting in a vector that is the same size as $\mathbf{w}$ and $\mathbf{v}$. Furthermore, we use $E[\mathbf{w}]$, $Var(\mathbf{w})$, and $Cov(\mathbf{w})$ as shorthand to refer to computations across the elements of a vector (and *not* as operations on a random variable): e.g., $E[\mathbf{w}] = \frac{1}{n}\sum_{i=1}^{n} w_i$, and $Cov[\mathbf{w}, \mathbf{v}] = E[\mathbf{w} \cdot \mathbf{v}] - E[\mathbf{w}]E[\mathbf{v}] = \frac{1}{n}\sum_{i=1}^{n} w_i v_i - \left(\frac{1}{n}\sum_{i=1}^{n} w_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} v_i\right)$. Also, in this section we refer to the loading similarity of vector $\mathbf{w}$ as $ls(\mathbf{w})$ for shorthand.

We first show a constraint on loading similarities for a model with two co-fluctuation patterns (i.e. loading vectors for each dimension). Let n be the number of neurons and let $\mathbf{w}, \mathbf{v} \in \mathbb{R}^n$ be two loading vectors. As in our simulations and data analysis (see Methods), $\mathbf{w}$ and $\mathbf{v}$ are orthogonal unit vectors: $\sum_{i=1}^{n} w_i^2 = 1$, $\sum_{i=1}^{n} v_i^2 = 1$, and $\sum_{i=1}^{n} w_i v_i = 0$. Then, using these constraints,

$$\begin{aligned} Cov(\mathbf{w}, \mathbf{v}) &= E[\mathbf{w} \cdot \mathbf{v}] - E[\mathbf{w}]E[\mathbf{v}] \\ &= \frac{1}{n}\sum_{i=1}^{n} w_i v_i - E[\mathbf{w}]E[\mathbf{v}] \\ &= -E[\mathbf{w}]E[\mathbf{v}] \\ Var(\mathbf{w}) &= E[\mathbf{w} \cdot \mathbf{w}] - E[\mathbf{w}]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} w_i^2 - E[\mathbf{w}]^2 \\ &= \frac{1}{n} - E[\mathbf{w}]^2 \end{aligned} \tag{14}$$

Because correlation is bounded between -1 and 1, we know that $|Cov(\mathbf{w}, \mathbf{v})| \leq \sqrt{Var(\mathbf{w})Var(\mathbf{v})}$. It follows that:

$$\begin{aligned} Cov^2(\mathbf{w}, \mathbf{v}) &\leq Var(\mathbf{w})Var(\mathbf{v}) \\ E[\mathbf{w}]^2 E[\mathbf{v}]^2 &\leq \left(\frac{1}{n} - E[\mathbf{w}]^2\right)\left(\frac{1}{n} - E[\mathbf{v}]^2\right) \\ 0 &\leq \frac{1}{n^2} - \frac{1}{n}\left(E[\mathbf{w}]^2 + E[\mathbf{v}]^2\right) \\ nE[\mathbf{w}]^2 + nE[\mathbf{v}]^2 &\leq 1 \\ ls(\mathbf{w}) + ls(\mathbf{v}) &\leq 1 \end{aligned} \tag{15}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The last step follows from the definition of loading similarity:

$$ls(\mathbf{w}) \equiv 1 - \frac{Var(\mathbf{w})}{1/n} = 1 - \frac{\frac{1}{n} - E[\mathbf{w}]^2}{1/n} = nE[\mathbf{w}]^2$$

The final inequality in equation (15) proves the intuition provided at the beginning of this section–if $ls(\mathbf{w})$ is large, then $ls(\mathbf{v})$ must be small (at most $1 - ls(\mathbf{w})$). More strongly, if $ls(\mathbf{w}) = 1$, then $ls(\mathbf{v}) = 0$.

Generally, for a model with d dimensions and patterns $\mathbf{u_1}, \ldots, \mathbf{u_d} \in \mathbb{R}^n$, we can show that $\sum_{i=1}^{d} ls(\mathbf{u_i}) \leq 1$. To see this, we can construct a matrix C with entries $c_{ij} = Cov(\mathbf{u_i}, \mathbf{u_j}) = -E[\mathbf{u_i}]E[\mathbf{u_j}]$ for $i \neq j$, and $c_{ii} = Var(\mathbf{u_i}) = \frac{1}{n} - E[\mathbf{u_i}]^2$ (derived from the constraints in equation (14)). Note that $C \in \mathbb{R}^{d \times d}$, with variances on the diagonal and covariances on off-diagonals, is a covariance matrix, which implies $det(C) \geq 0$. For a 3-d model, $det(C) = \frac{1}{n^2}\left(1 - nE[\mathbf{u_1}]^2 - nE[\mathbf{u_2}]^2 - nE[\mathbf{u_3}]^2\right) \geq 0$, which implies $ls(\mathbf{u_1}) + ls(\mathbf{u_2}) + ls(\mathbf{u_3}) \leq 1$. In general, for a d-dimensional model (with $d \leq n$):

$$det(C) = \frac{1}{n^{d-1}}\left(1 - \left(\sum_{i=1}^{d} nE[\mathbf{u_i}]^2\right)\right) \geq 0$$
$$\sum_{i=1}^{d} ls(\mathbf{u_i}) \leq 1 \tag{16}$$

Equation (16) has several implications:

- If one knows the loading similarities of all dimensions $\mathbf{u_1}, \ldots, \mathbf{u_d}$ in a model, then the maximum possible loading similarity of any new dimension is $1 - \sum_{i=1}^{d} ls(\mathbf{u_i})$. It follows that two dimensions with high loading similarity cannot co-exist in the same model.
- If one dimension has $ls = 1$, then all other dimensions in the model (or that would be added to the model) necessarily have $ls = 0$. Note that there is only one possibility for a pattern to have $ls = 1$ (i.e., $\mathbf{u} = [\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}]^T$, such that $Var(\mathbf{u}) = 0$). This implies that there are many possibilities for a pattern to have $ls(\mathbf{u}) = 0$. More loosely, there are relatively few ways for a pattern to have high loading similarity, but many more ways for a pattern to have low loading similarity.

F Maximum variance of a unit vector

We defined loading similarity for a co-fluctuation pattern $\mathbf{u}$ (normalized to have norm 1) of n neurons to be $1 - \frac{var(\mathbf{u})}{1/n}$, where the variance is computed along the elements of $\mathbf{u}$. This value lies between 0 and 1 because the maximum variance across the elements of $\mathbf{u}$ is $1/n$. We now show this mathematically.

Let $\mathbf{u} \in R^n$ be a unit vector. Because $\mathbf{u}$ is a unit vector, $\sum_{i=1}^{n} u_i^2 = 1$. Using these facts:

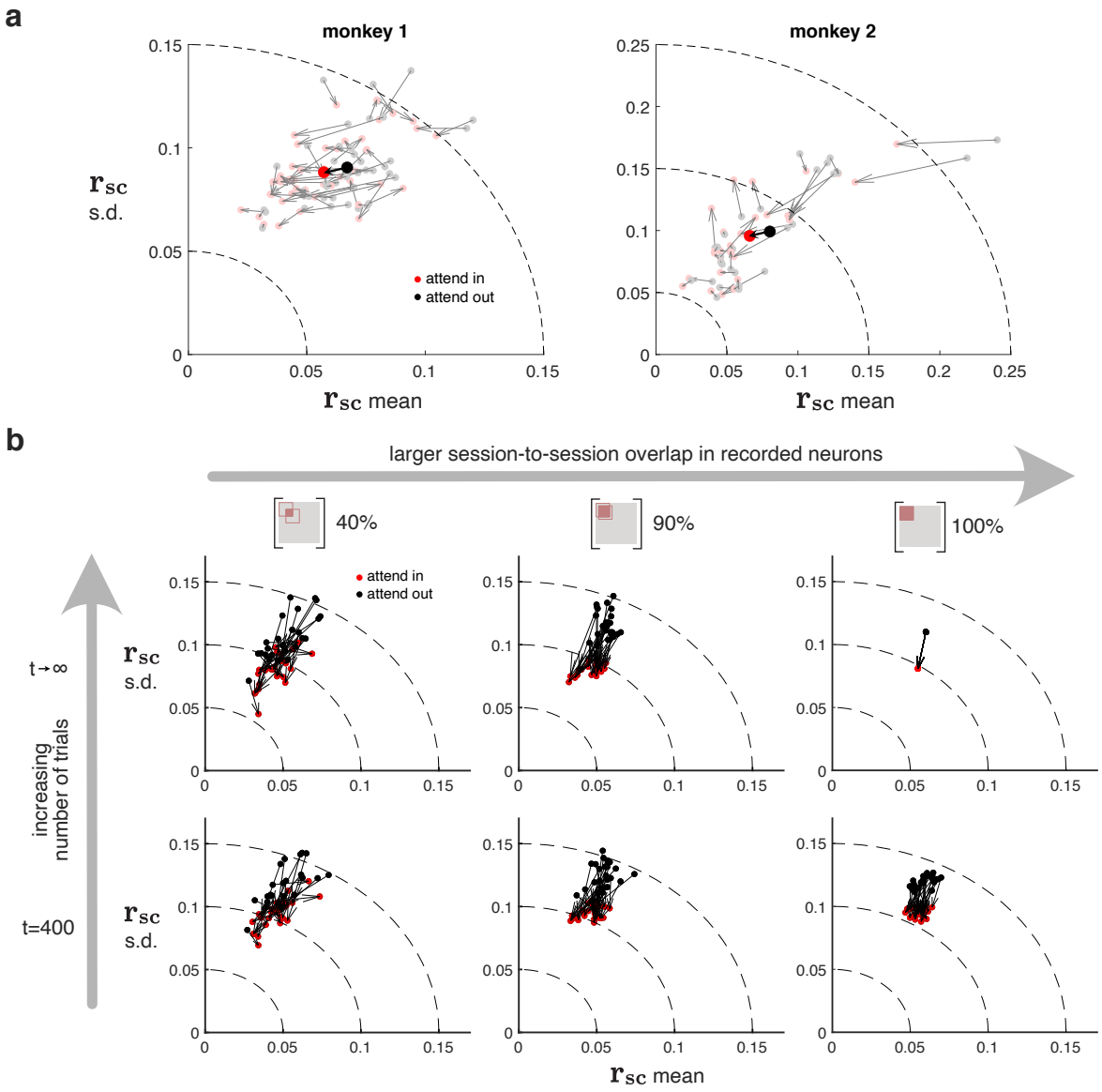
$$\begin{aligned} Var(\mathbf{u}) &= E[\mathbf{u}^2] - E[\mathbf{u}]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} u_i^2 - E[\mathbf{u}]^2 \\ &= \frac{1}{n} - E[\mathbf{u}]^2 \\ &\leq \frac{1}{n} \end{aligned}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This holds with equality when $E[\mathbf{u}] = 0$ (i.e., when the mean across the elements in a co-fluctuation pattern is 0). This implies that the smallest loading similarity is 0 (when $Var(\mathbf{u}) = 1/n$), and the largest loading similarity is 1 (when $Var(\mathbf{u}) = 0$).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figures



Supplementary Figure 1: Session-to-session variability in pairwise metrics can be accounted for by finite trial counts and neuron sampling.

We sought to understand to what extent the average pairwise correlations reported in Figure 6*b* varied from session to session, and the possible sources for this across-session variability. Here, we consider two possible sources: 1) the number of recorded trials, and 2) the overlap in recorded neurons across sessions. Although we likely recorded from many of the same neurons across sessions, the population of neurons recorded on each session was not identical due to recording instability (sometimes referred to as electrode drift), a well-known property of electrophysiological recordings.

(continued on next page...)

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 1: (continued from previous page...)

a. In V4 recordings, we observed substantial session-to-session variability in pairwise metrics. Each gray arrow, corresponding to an individual session, begins at r_{sc} mean and s.d. values for 'attend-out' (light gray dots) and ends at r_{sc} estimates for 'attend-in' (light red dots). The average change in r_{sc} mean and s.d (black arrow, black dot, solid red dot) correspond to the values we report in Figure 6*b*.

b. Using simulations, we assessed how varying the number of trials and overlap of neurons across sessions contribute to the across-session variability of r_{sc} mean and s.d.

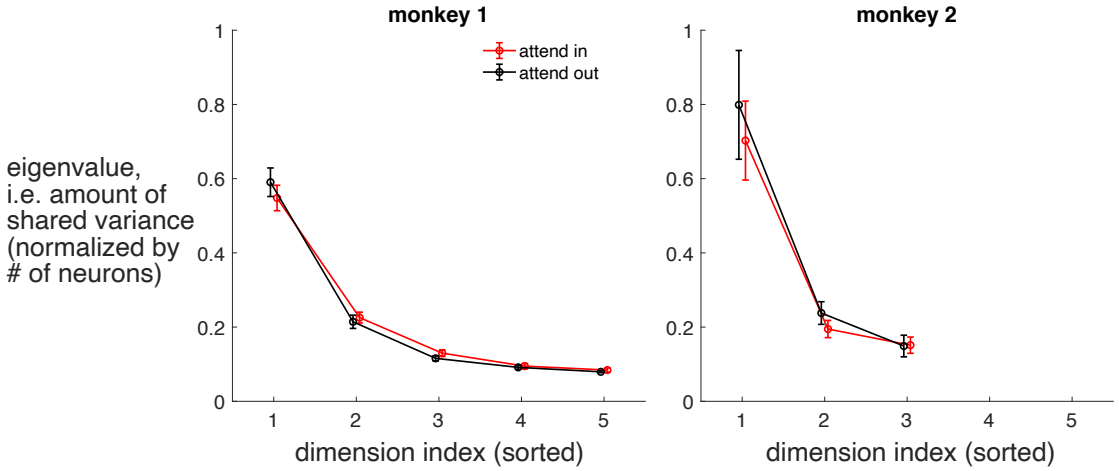
The simulation procedure is described as follows. We first considered the setting of a large number of trials with varying amounts of neuron overlap. We created covariance matrix using factor analysis (equation 4) for 2,500 neurons and computed r_{sc} mean and s.d. from this matrix. Since we had access to the true covariance matrix, we considered this setting as having an "infinite" number of trials (top row). For each simulated recording session, we sampled 50 neurons from the set of 2,500 neurons. To mimic the effects of recording different neurons on different sessions, we sampled neurons such that a percentage of neurons were the same as those in the previous session (e.g., with 50 neurons in a session, a 40% overlap corresponds to 20 neurons being common with the previous session from two sessions, whereas 100% corresponds to having the same 50 neurons on all sessions). To simulate 'attend in' and 'attend out' trials, we chose FA parameters to be similar to those estimated from the V4 data (attend in: load. sim.=0.45, %sv=15%, dim.=3; attend out: load. sim.=0.5, %sv=20%, dim.=3).

We found that as the percentage of neuron overlap between sessions increased, the across-session variability in the r_{sc} estimates decreased (moving left to right in top row). This indicates that if different neurons are recorded in different sessions, one should expect differences in estimates of pairwise metrics across sessions.

We then considered how the number of recorded trials influences the across-session variability. We used the same simulation procedure as described above, except for one key difference. Instead of using the true covariance matrix to directly compute r_{sc}, we instead simulated trials of "neuronal activity" from that covariance matrix. We simulated 400 trials because that was, on average, the number of trials we recorded per condition per session (i.e., each light-colored dot in panel *a* is computed from 400 trials). We then estimated the spike count covariance matrix based on these 400 trials. We found that the level of across-session variability did not substantially differ for 400 trials (bottom row) versus an infinite number of trials (top row), except in the case of 100% overlap, i.e. no recording instability.

Using these simulations as a reference, it appears that much of the across-session variability in our reported V4 pairwise metrics (panel *a*) can be accounted for by neural recording instability and finite trial sampling. Additional across-session variability in the V4 results may be due to other cognitive factors, such as across-session differences in the animal's motivation level.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Figure 2: Eigenspectra for V4 population activity in 'attend in' and 'attend out' conditions.

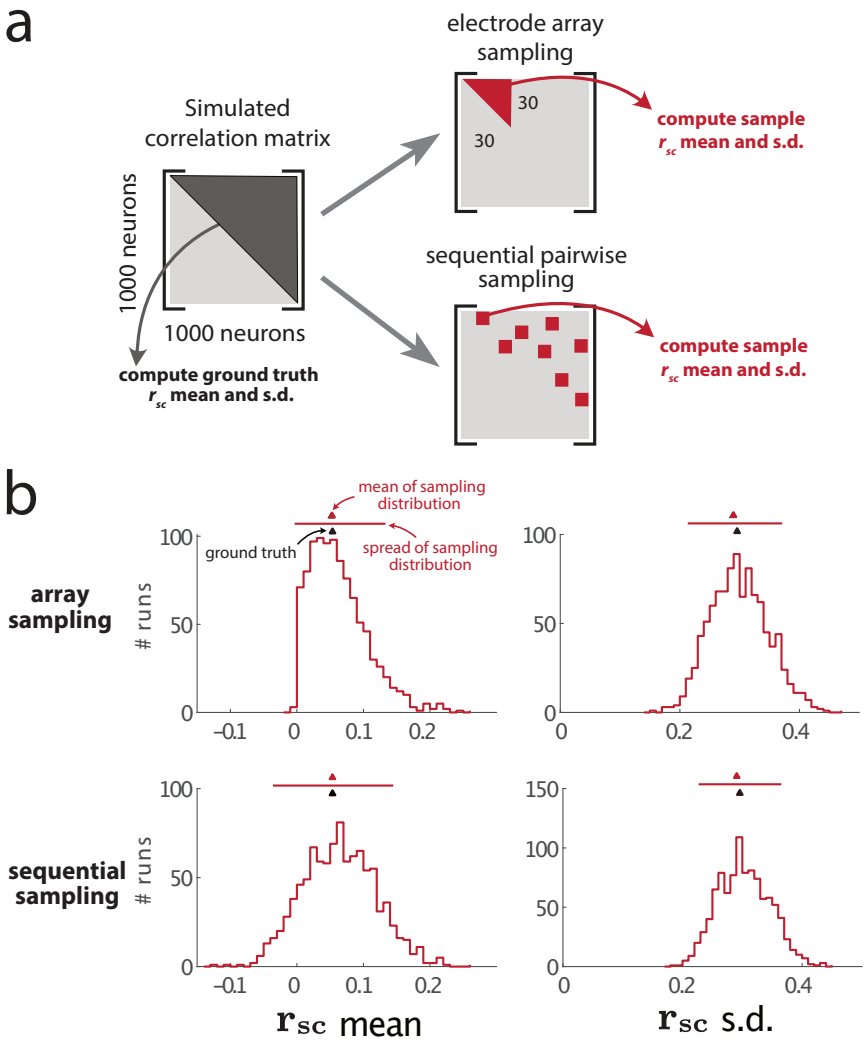
Although we observed only a modest change in dimensionality with attention (Fig. 6*c*), our simulations showed that the relative strength of each dimension (i.e., shape of the shared eigenspectrum) could alter the "effective dimensionality" of population activity and have large effects on pairwise metrics (Fig. 4*a*). Here, we asked whether the relative strengths of each dimension changed with attention.

The eigenspectra were computed in the following way. We decomposed the V4 spike count covariance matrix into shared and independent components using factor analysis (see Methods). We then computed the eigendecomposition of the shared covariance matrix (Fig. 7, $\Sigma_{\text{shared}} = U\Lambda U$). We found that eigenvalues (diagonal of Λ) tended to increase linearly with the number of neurons recorded; therefore, in order to combine across sessions, we normalized the eigenvalues by dividing by the number of neurons recorded in each session. After normalizing, we computed the eigenspectrum averaged across sessions and stimulus orientations. Because the dimensionality identified by cross-validation differed across sessions, there were a different number of sessions that contributed to each average. We did not plot mean eigenvalues when there were fewer than 5 sessions to average (i.e., dimensions ≥ 6 for monkey 1; dimensions ≥ 4 for monkey 2). Error bars indicate standard error. Data points have been jittered horizontally for visual clarity.

We found that the shape of the eigenspectra was qualitatively similar for 'attend in' and 'attend out' conditions (red and black curves have similar shape). In both conditions, the eigenvalues of the shared covariance matrix decayed (dot for each subsequent dimension was below dot for the previous dimension), indicating that a small number of dimensions were needed to explain the population-wide covariability.

When comparing eigenspectra (i.e., the amount of shared variance explained by each dimension), one also needs to consider the firing rates under each condition. Mean firing rates tend to be higher for attend-in than attend-out trials. Higher firing rates typically correspond to higher spike count variance due to the Poisson-like firing of neurons. All else being equal, the higher mean firing rates imply higher levels of both shared variance and independent variance (Churchland et al., 2010). Thus, a direct comparison of the eigenspectra should be done with caution. Nonetheless, we plotted attend-in and attend-out together to relate our results to previous reports (Huang et al., 2019; Ruff et al., 2019b). Consistent with these studies, we found that attention decreased the strength of the strongest dimension (red below black dot for dimension index 1), though the magnitude of the decrease we observed was more consistent with Ruff et al. (2019b) than Huang et al. (2019). Had we been able to equalize the mean firing rate across the two conditions, we likely would have observed an even greater difference between attend-in and attend-out. We note that the caveat described here for comparison of eigenspectra (i.e., the amount of shared variance) does not apply to comparisons of %sv (Fig. 6*c*) because %sv is normalized by the overall spike count variance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Figure 3: Pairwise metrics can be estimated equally well by either sampling pairs of neurons over many sessions or simultaneously sampling many neurons.

Many studies that report r_{sc} mean measured the activity of one, or a few, pairs of neurons in each recording session, and combined r_{sc} across separate recording sessions. We wondered if our estimates of pairwise metrics may change if we could only sample random pairs of neurons, instead of simultaneously sampling a population of neurons using a Utah array (as in our V4 recordings). This "sequential pairwise sampling" is often done out of necessity due to physiological restrictions (e.g., recordings in deep brain structures) or technological limitations (e.g., recording with pairs of single recording electrodes, or a single tetrode). More recent studies often use multi-electrode array recording technologies, such as Utah arrays, that simultaneously record the activity of tens of neurons. We term this "electrode array sampling"; it is the type of sampling we considered in our main text simulations and V4 recordings. Here, we asked whether 1) sequential pairwise sampling and electrode array sampling provide similar estimates when computing pairwise statistics and 2) whether each type of sampling reflects the pairwise statistics of a larger neuronal population.

(continued on next page...)

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 3: (...continued from previous page)

a. To address these questions in simulation, we created, we created a large 'ground truth' covariance matrix of population activity, and then simulated recording sessions by sampling entries from this matrix using either a sequential pairwise or electrode array approach. We created a covariance matrix for a population of 1000 neurons using one co-fluctuation pattern (i.e., one latent dimension), a loading similarity of 0.2, and a %sv of 30% (Fig. 7). We also tried other values for the population metrics and obtained similar results to that shown below. From the covariance matrix, from which we computed a $1,000 \times 1,000$ correlation matrix (left, black upper triangle). The r_{sc} mean and s.d. of this correlation matrix were deemed to be the ground truth values that we aimed to estimate using sequential pairwise or electrode array sampling. To simulate an electrode array recording, we randomly sampled 30 neurons to obtain a 30×30 covariance matrix (top, red triangle). To simulate sequential recordings of pairs of neurons, we randomly sampled 30 entries of the population correlation matrix (bottom, random red squares). We repeated each type of sampling 1,000 times (i.e., 1,000 bootstrap runs), and for each run we computed r_{sc} mean and r_{sc} s.d. This procedure provided us with sampling distributions of the r_{sc} mean and s.d.

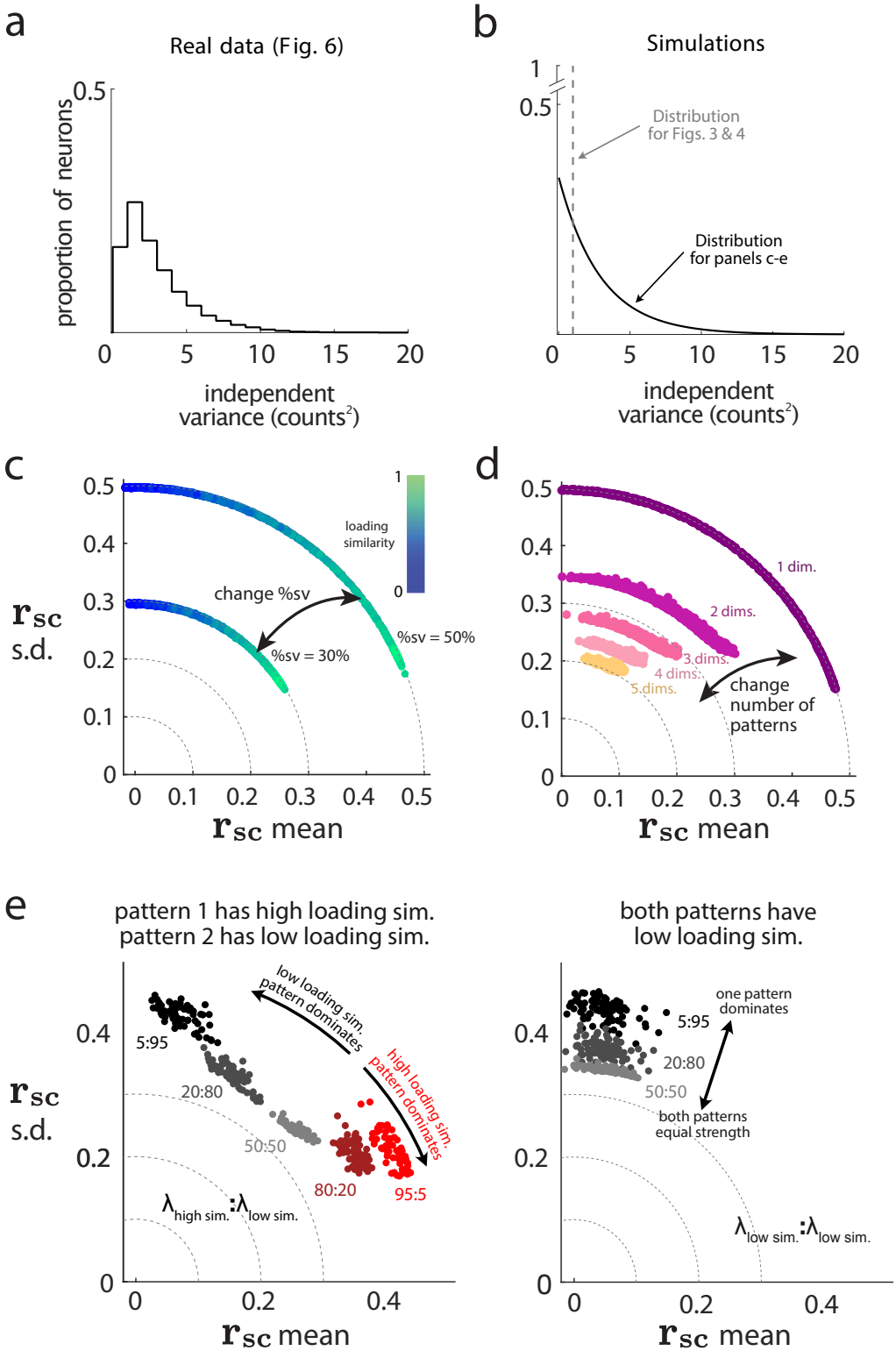
b. For both array sampling (top row) and sequential pairwise sampling (bottom row), we compared the estimates of r_{sc} mean and s.d. on each run with the ground truth values. To assess whether either metric was consistently under or overestimated, we took the difference between the mean of the sampling distribution (red triangles) and ground truth values (black triangles). We also considered how consistent/replicable the estimates were, quantified by the spread of the sampling distribution (red bars, 90% confidence interval).

For r_{sc} mean (left column), both array sampling (top row) and sequential sampling (bottom row) resulted in little to no statistical bias (red triangles and black triangles are closely aligned), and both types of sampling showed similar replicability (red bars are of similar lengths). We note that array sampling resulted in few negative r_{sc} mean estimates because the 30×30 correlation matrix obtained with array sampling must be a positive semi-definite matrix, limiting the range of possible r_{sc} mean values and making negative r_{sc} means unlikely.

For r_{sc} s.d. (right column), both array sampling (top row) and sequential sampling (bottom row) slightly underestimated the ground truth value (red triangle lies to the left of black triangle). This underestimation bias was small relative to the larger variance in estimates of r_{sc} s.d. (red bars).

Overall, we found that both sequential pairwise sampling and electrode array sampling provide similar answers for r_{sc} mean and r_{sc} s.d. Also, both types of sampling provide reasonably good estimates of r_{sc} mean and r_{sc} s.d. of a larger population, given a reasonable sampling of neurons (30 neurons for electrode, or 30 pairs for pairwise sampling here) relative to the larger population (1000 neurons here). This implies that it is still possible to gain insight into a larger population of neurons and use the intuitions provided in the main text of this work even in cases where one can only record sequentially from pairs of neurons over multiple recording sessions. Note that the results here assume the covariance matrix is accurately estimated using a large number of trials.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Figure 4: Relationship between pairwise and population metrics shows little dependence on distribution of independent variances.

(continued on next page...)

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 4: (...continued from previous page)

In our simulations (Figs. 3 and 4), we fixed the independent variance for each neuron to be the same value. However, we found that for real data, the independent variance differed across neurons. To account for this, we re-performed our simulations using a distribution of independent variances that was similar to that of the real data. We found the same trends, suggesting that different distributions of independent variances lead to similar conclusions about the relationship between pairwise and population metrics.

a. We applied factor analysis to V4 population activity recorded during a spatial attention task (see Fig. 6 and Methods) and estimated each neuron's independent variance (Fig. 7*a*, gray diagonal entries). Independent variances between neurons differed (dashed line denotes histogram).

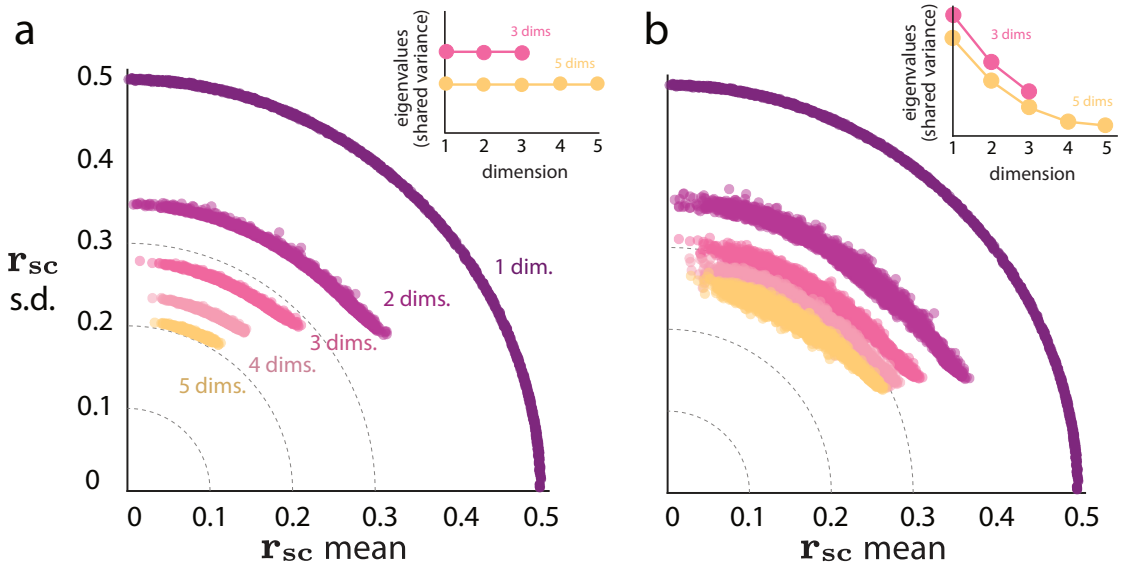
b. In the main text (Figs. 3 and 4), our simulations used a fixed independent variance of 1 for all neurons (dahsed gray line). Here (**c**-**e**), we re-performed the same simulations but drawing independent variances from an exponential distribution (solid black line) which approximates the distribution of the real data (compare solid black line here to that in in **a**). We used an exponential distribution with a mean of 3.

c. Varying %sv and loading similarity while keeping the dimensionality fixed at 1. Same conventions as in Fig. 3*f*. We set %sv by scaling the eigenvalue of the one dimension (see Methods). With changing %sv, exponentially-distributed independent variances led to similar changes in radius of the arcs as when independent variances were all 1 (compare to Fig. 3*f*). However, here the arc lengths were smaller than in Figure 3*f* because small values of r_{sc} s.d. were unlikely with different independent variances. However, an r_{sc} s.d. of 0 is still possible under special conditions that are not likely to be encountered in practice (i.e., if the loading for each neuron is set to be the square root of its independent variance; see Supplementary Math Note A).

d. Varying dimensionality and loading similarity while keeping %sv fixed at 50%. Same conventions as in Fig. 3*g*. We found that as dimensionality increases, r_{sc} mean and r_{sc} s.d. decrease, consistent with our simulations in Fig. 3*g*.

e. Varying the relative strengths of two dimensions. Same conventions as in Fig. 4. We observed similar trends whether the independent variances were all fixed to 1 (Fig. 4) or drawn from an exponential distribution. However, here with exponentially-distributed independent variances, covariance matrices for which the pattern of the dominant dimension (i.e., the dimension with the largest eigenvalue) had a high loading similarity (left panel, 95:5, red dots) yielded a higher r_{sc} s.d. than when independent variances were the same for each neuron. This is for the same reasons as noted in panel **c**.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Figure 5: Relationship between pairwise and population metrics does not depend on relative strengths of dimensions.

In Fig. 3*g*, we varied dimensionality and assessed the changes to pairwise metrics. For that simulation, we assumed that each dimension had equal strength (i.e., a flat eigenspectrum for the shared covariance matrix). Here, we asked if our simulation results held when we changed the shape of the shared covariance eigenspectrum such that it was no longer flat (i.e., some dimensions explained more shared variance than others). We found similar trends, suggesting that a different eigenspectrum shape leads to similar relationships between pairwise and population metrics as those found in the main text.

a. Results for flat eigenspectra (reproduction of Fig. 3*g*, here with inset). Each dimension of the shared covariance matrix had the same eigenvalue (top right inset), and thus each dimension explained the same amount of shared variance. Note that models with different dimensionalities had different eigenvalues (inset, 3-dim. has larger eigenvalues than those for 5-dim.), because we scaled the eigenvalues (while keeping fixed each neuron's independent variance) to enforce the %sv to be the same across models of different dimensionalities (inset, the sum of eigenvalues is equal for 3-dim. and 5-dim.).

b. In neuronal recordings, the shape of eigenspectra curves estimated from recorded neuronal activity is usually not flat (see Supplementary Fig. 2). We sought to understand how the shape of the eigenspectrum affects the relationship between pairwise and population metrics. We re-performed the simulations in Figure 3*g* except here with the eigenspectra curves were exponentially decaying according to $\lambda_i = e^{-\frac{2}{3}i}$ (inset). The overall trend held true: an increase in dimensionality led to an overall decrease in the radius of the arc (yellow dots closer to origin than purple dots).

(continued on next page...)

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.04.383604; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 5: (...continued from previous page)

We also observed two subtle differences from Figure 3*g*. First, an exponentially-decaying eigenspectrum tended to have a higher r_{sc} mean and r_{sc} s.d. compared to its corresponding flat eigenspectrum (e.g., yellow dots for 5 dims are further from origin here than in panel *a*). This occurs because, for an exponentially-decaying eigenspectrum, an added dimension explains relatively little shared variance (right side of the curves in inset). Thus, the added dimension results in only small changes in r_{sc} mean and r_{sc} s.d. On the other hand, adding a dimension to the flat eigenspectrum affects r_{sc} mean and r_{sc} s.d. as much as any other dimension, leading to larger changes in r_{sc} mean and r_{sc} s.d. than in the case of an exponentially-decaying eigenspectrum.

The second difference was, for a given dimensionality, a greater radial and angular spread for exponentially-decaying eigenspectra compared to flat eigenspectra (e.g., yellow dots for 5 dims are more spread out in the radial direction here than in panel *a*; yellow dots are also more spread out parallel to the dashed lines). This occurs because, when the eigenspectra are not flat, there is greater diversity in how the co-fluctuation patterns of different dimensions can contribute to r_{sc}. In other words, switching the eigenvalues of two dimensions with equal eigenvalues (i.e., both dimensions explain the same amount of shared variance) results in the same model and same covariance matrix—yielding the same values for r_{sc} mean and r_{sc} s.d. However, switching the strengths of two dimensions with different eigenvalues is likely to result in a different covariance matrix and thus different values of r_{sc} mean and r_{sc} s.d. Thus, for non-flat eigenspectra, the greater diversity by which co-fluctuation patterns can contribute to the shared covariance matrix leads to greater spread in the r_{sc} s.d. vs r_{sc} mean plots. The mathematical details about the spread of dots in the radial direction are provided in Supplementary Math Note D.

An implication of this analysis is that it is important to report the eigenspectrum shape whenever one reports dimensionality. For example, consider population activity described by 5 dimensions with an exponentially-decaying eigenspectrum (panel *b*, yellow dots). Because of 1) low shared variance in the last 2 dimensions, and 2) greater diversity in how dimensions contribute to the shared covariance matrix, the activity might closely resemble that described by 3 dimensions with a flat eigenspectrum (*a*, pink dots). Thus, considering both dimensionality and the eigenspectrum curve, instead of dimensionality alone, will lead to a more complete picture of the structure of population activity.