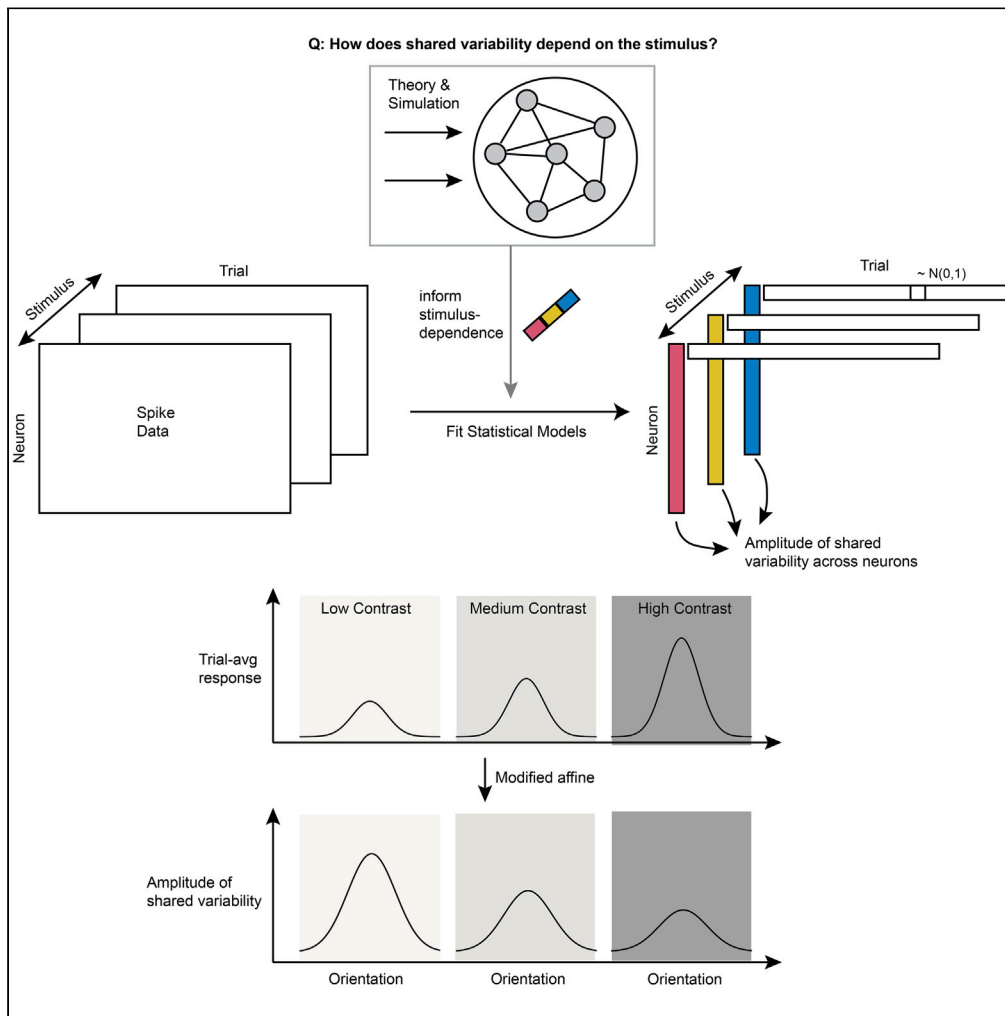**Article**

Circuit-motivated generalized affine models characterize stimulus-dependent visual cortical shared variability

Ji Xia, Anna Jasper, Adam Kohn, Kenneth D. Miller

jx2484@columbia.edu

Highlights

Compare statistical models of stimulus-dependent correlated variability in V1, V2

Recurrent neural circuit mechanism informs statistical model designs

Orientation-dependent shared variability between and within V1 and V2 is affine

A modified affine model explains contrast-dependent shared variability in V2

Article

Circuit-motivated generalized affine models characterize stimulus-dependent visual cortical shared variability

Ji Xia,[1,6,*] Anna Jasper,[2] Adam Kohn,[2,4,5] and Kenneth D. Miller[1,3]

SUMMARY

Correlated variability in the visual cortex is modulated by stimulus properties. The stimulus dependence of correlated variability impacts stimulus coding and is indicative of circuit structure. An affine model combining a multiplicative factor and an additive offset has been proposed to explain how correlated variability in primary visual cortex (V1) depends on stimulus orientations. However, whether the affine model could be extended to explain modulations by other stimulus variables or variability shared between two brain areas is unknown. Motivated by a simple neural circuit mechanism, we modified the affine model to better explain the contrast dependence of neural variability shared within either primary or secondary visual cortex (V1 or V2) as well as the orientation dependence of neural variability shared between V1 and V2. Our results bridge neural circuit mechanisms and statistical models and provide a parsimonious explanation for the stimulus dependence of correlated variability within and between visual areas.

INTRODUCTION

Neural responses in the visual cortex to repeated presentations of a fixed visual stimulus exhibit trial-to-trial variability. These trial-to-trial fluctuations are correlated or shared among neurons in a local population. There are three primary motivations for investigating and characterizing this shared variability. First, it plays a pivotal role in shaping stimulus coding.[1–4] For example, shared variability patterns that steer the population response from one stimulus toward the response to a nearby stimulus can hinder accurate stimulus coding.[3] Second, accumulating evidence suggests that the standard deviation of shared variability at each neuron is modulated by stimuli,[5–8] as well as by behavioral[9,10] and task-related[11–13] variables. Consequently, a prevailing argument posits that shared variability does not just reflect stochasticity; instead, it serves to encode additional, complementary information that is absent from the trial-averaged responses and may have computational utility.[14–16] Third, the pattern of shared variability reflects the underlying connectivity among neurons.[17,18] Therefore, characterizing shared variability serves as a valuable approach for inferring circuit structure, especially because performing experimental measurements of single neuron input-output function or connectivity is often challenging.[19–21]

In this work, we focus on investigating and characterizing how shared variability depends on stimuli. A comprehensive understanding of the stimulus-dependent nature of shared variability lays the groundwork for quantifying its impact on stimulus coding and inferring circuit structure.

Three models of stimulus-modulated shared variability have been considered in most previous work[7,22–24]: the amplitude (standard deviation) of a given neuron's shared variability might be multiplicative (proportional to the neuron's stimulus response), additive (stimulus independent), or affine (combination of a multiplicative and an additive component). For neurons in the visual cortex, previous work found that multiplicative[22] or affine[7,24] models outperform the additive model in explaining the neural data. However, further work is needed to thoroughly characterize stimulus-dependent shared variability. It is possible that the underlying modulation takes other more complicated forms than the three proposed models. Furthermore, these models cannot account for the fact that shared variability is suppressed with increasing stimulus contrast.[25,26] Moreover, these models are statistical in nature and lack a foundation in neural circuitry mechanism. In this study, we will demonstrate that consideration of such mechanisms suggests alternative forms for the modulation of shared variability.

Trial-to-trial variability is not only correlated across neurons within a brain area but also across neurons from two connected brain areas.[27–33] This correlation of neural activity between brain areas is often interpreted as area-to-area communication. Characterizing the shared variability between brain areas is vital for understanding its impact on joint stimulus coding by multiple brain areas and for inferring

[1]Center for Theoretical Neuroscience and Mortimer B Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA
[2]Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA
[3]Department of Neuroscience, Swartz Program in Theoretical Neuroscience, Kavli Institute for Brain Science, College of Physicians and Surgeons and Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York City, NY 10027, USA
[4]Department of Ophthalmology and Visual Sciences, Albert Einstein College of Medicine, Bronx, NY, USA
[5]Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY, USA
[6]Lead contact
*Correspondence: jx2484@columbia.edu
https://doi.org/10.1016/j.isci.2024.110512

Box 1. How does neural variability depend on the stimulus in a recurrent neural network?

In this box, we show that in a recurrent neural network, one should expect neural variability (fluctuation around the fixed point) to depend on the stimulus (external input), if we have a nonlinear activation function for the neurons.

Let us consider neural dynamics governed by a standard rate network equation:

$$\tau \frac{dr(t)}{dt} = -r(t) + f(Wr(t) + h(t)) \qquad \text{(Equation 1)}$$

Here, τ is the time constant of neurons. $r(t)$ is a vector that denotes the firing rates of all the neurons at time t. W is the matrix of recurrent connectivity between neurons. $h(t)$ is a vector that denotes the external input received by all the neurons at time t. $f(\cdot)$ is the activation function, which is applied element-by-element to its vector argument.

Consider the case where $h(t) = h^*$ is a constant over time. Then, $r(t)$ will evolve to a fixed point r^* that satisfies

$$r^* = f(Wr^* + h^*) \qquad \text{(Equation 2)}$$

provided that this fixed point is stable. We are interested in how $r(t)$ fluctuates around r^*, if $h(t) = h^* + \delta h(t)$, where $\delta h(t)$ denotes some external input noise. In other words, we want to know δr that satisfies the following equation:

$$\tau \frac{d(r^* + \delta r(t))}{dt} = -(r^* + \delta r(t)) + f(W(r^* + \delta r(t)) + (h^* + \delta h(t))) \qquad \text{(Equation 3)}$$

As in the study by Trousdale et al., Ocker et al., and Hennequin et al.,[17,18,26] we assume that both δr and δh are sufficiently small that we can Taylor expand the second term on the right-hand side of Equation 3 around the fixed point and only keep the first-order term:

$$f(W(r^* + \delta r(t)) + (h^* + \delta h(t))) \approx f(Wr^* + h^*) \\ + FW\delta r(t) + F\delta h(t)$$

F is a diagonal matrix with diagonal entries $f'(Wr^* + h^*)$, which is the vector of neuronal gains—the slope of the activation function at each neuron's fixed-point firing rate.

With Equation 2, we can simplify Equation 3 to get:

$$\tau \frac{d(\delta r(t))}{dt} = -\delta r(t) + FW\delta r(t) + F\delta h(t) \qquad \text{(Equation 4)}$$

If we further assume that the external input noise $\delta h(t)$ varies slowly over time (much slower than τ), the left-hand side of Equation 4 will tend to zero; thus, we can get an expression for $\delta r(t)$:

$$\delta r(t) = (I - FW)^{-1} F\delta h(t) \qquad \text{(Equation 5)}$$

Empirically, the slow input noise approximation gives a decent description of the simulation as long as autocorrelation time of the external input noise is slower than ~ 50 ms (see Supplementary section "analysis in simplified scenarios" in the study by Hennequin et al.[26]).

If the activation function $f(\cdot)$ is nonlinear, then F changes with neuronal activation level and thus depends on r^*, which, in turn, depends on the stimulus h^*. Consequently, $\delta r(t)$ also depends on the stimulus through F, as discussed before in the study by Doiron et al.[35]

To gain more intuition on what $\delta r(t)$ looks like, consider a special case where $f(x) = kx^n$ (with $n > 1$, $x > 0$), then we have $F = \text{diag} f'(f^{-1}(r^*)) = \text{diag}(nk^{1/n} r^{*(1-1/n)})$. In this case, entries of F are monotonically increasing with r^*. As a motivation for the choice, it has been shown that this power-law nonlinearity well describes the responses of visual cortical neurons,[36–38] with evidence from *in vivo* whole-cell recordings in V1.[19]

Equation 5 can be simplified in two limiting regimes, depending on the absolute values of the eigenvalues of FW. To explain this, we first consider the relationship between mean firing rate r^* and the absolute eigenvalues of FW. As we have just seen, increasing r^* leads to higher positive diagonal entries of F. Since $\det(FW) = \det F \det W$, increasing r^* leads to higher $|\det(FW)|$ through increasing $\det F$, which typically results in increased absolute eigenvalues of FW (though exceptions exist).

Our first limiting case is if the entries of FW are sufficiently small (low firing rates and/or weak weights) that all its eigenvalues have absolute values smaller than 1 (i.e., FW has a spectral radius less than 1). Then, we can expand $(I - FW)^{-1}$ to obtain

$$\delta r(t) \approx \left(I + FW + (FW)^2 + \ldots\right) F\delta h(t)$$

The largest (first) term is $F\delta h(t) \propto \text{diag}(r^{*(1-1/n)})\delta h(t)$. In this case, $\delta r(t)$ tends to be higher for those neurons with higher mean firing rate r^* under a given stimulus h^*.

In our second limiting case, the entries of FW are sufficiently large (high firing rates and/or strong weights) that all of the eigenvalues of FW have absolute values larger than 1. Then, we can expand $(I - FW)^{-1} = (I - (FW)^{-1})^{-1}(-FW)^{-1}$ to obtain

$$\delta r(t) \approx \left(I + (FW)^{-1} + (FW)^{-2} + \cdots\right)(-W)^{-1}\delta h(t)$$

> **Box 1. Continued**
>
> The largest (first) term is $-W^{-1}\delta h(t)$, which is independent of the mean firing rate r^* under a given stimulus. The higher-order terms will be stimulus dependent through F; however, the amplitude of these terms will decrease with increasing r^*. Interpolating between the two limiting cases, we expect the stimulus dependence of δr to become weaker as population mean firing rates increase.
>
> In summary, when the mean firing rates r^* are small enough so that FW has a spectral radius less than 1, the amplitude of neural variability δr is stimulus dependent and increases with increasing r^*; when the mean firing rates r^* are large enough so that all the eigenvalues of FW have absolute values larger than 1, the stimulus dependence of neural variability δr becomes weaker with increasing r^*.
>
> Here, neural variability $\delta r(t)$ could be equivalent to trial-to-trial variability if we define each time step as one trial. Otherwise, one may define that one trial consists of multiple time steps ($\{t\}$), where the trial-to-trial variability is $\delta r(t)$ summed over these time steps ($\sum_{\{t\}}\delta r(t) = (I - FW)^{-1}F\sum_{\{t\}}\delta h(t)$). Note that, in our neural circuit set up, the stimulus dependence of trial-to-trial variability will not change qualitatively according to the specific definition of trials.

inter-area connectivity. Recent studies found that variability shared between areas has lower dimensionality compared with that shared within each area,[32] which suggests flexible inter-areal communication occurs through a low-dimensional subspace. However, how this inter-areal shared variability is modulated by the external stimuli, and in particular, whether this modulation can also be captured by the previously proposed multiplicative or affine statistical models, remains unknown.[34]

Here, we investigated how shared variability within and between visual areas depends on the stimulus. We analyzed electrophysiology data simultaneously recorded in monkeys from areas V1 (primary visual cortex) and V2 (secondary visual cortex) in response to visual stimulation of drifting gratings with different orientations and contrasts. We studied a simple circuit mechanism for explaining the previously observed affine-like shared variability in V1.[7,24] We assumed that neural variability is the local fluctuation around a steady state of the recurrent circuit, driven by stimulus-independent external noise. Based on this mechanism, we made three predictions: (1) when firing rates are low, affine-like shared variability should be commonly observed in different brain areas with rectified power law activation functions. (2) In V1, if we consider varying both contrasts and orientations, affine models must be modified to allow contrast-specific coefficients, due to suppression of shared variability with increasing contrast by recurrent dynamics.[26] (3) With minimal assumptions, variability shared between V1 and V2 should also be affine-like across stimulus orientations. Consistent with (1), we found that affine models also explain V2 data well when we varied the stimulus orientations. Consistent with (2), when we varied both contrasts and orientations, we found that a generalized affine model with contrast-specific coefficients best explained shared variability in V2. Consistent with (3), affine models parsimoniously explained how the inter-areal variability shared between V1 and V2 depends on stimulus orientations. We also considered a generalized model that can have an arbitrary form of shared variability for each stimulus. This model slightly outperformed affine models when we had sufficient data to fit it, but differences were small.

Our work extends previous work in characterizing the nature of shared variability within a brain area to shared variability between brain areas. Moreover, we provide a mechanistic explanation for the commonly observed affine-like shared variability. Importantly, we propose an alternative form of stimulus-dependent shared variability that well explains the data when varying both orientation and contrast.

RESULTS

Circuit mechanism for stimulus-dependent shared variability

Numerous works have shown that in the visual cortex, shared variability is low dimensional, such that a large portion of the variance can be explained with a 1-dimensional component or a common fluctuation across neurons.[7,22,24,39,40] To date, there are two hypothesized circuit mechanisms for explaining low-dimensional shared variability: first, it may be inherited from low-dimensional external input noise and modified by recurrent processing[26,41]; second, it may be intrinsically generated by chaotic dynamics in the recurrent circuit.[40,42,43] In our work, we focused on the first possibility, because this framework is more theoretically tractable and equally biologically plausible, compared to the second possibility.

We assumed that neural variability represents fluctuations around a steady state of the recurrent circuit. Splitting neural activity into a trial-averaged mean plus variability ignores changes over trials in what is regarded as the "mean" response (for example due to adaptation), and so is a simplification, but is the standard framework used to analyze trial-to-trial variability.[35,44] Therefore, we modeled the variability as the fluctuations around a static steady state.

To investigate the stimulus dependence of shared variability, we started with deriving an analytical expression for neural variability in the recurrent neural circuit, assuming that external input noise is slow and small (Box 1). For the sake of simplicity, we assume for now that external input noise over time is perfectly correlated across all recipient neurons in our circuit model (i.e., $\delta h(t) = ab(t)$, a is a vector that denotes the amplitude of external input noise at all the neurons, and $b(t)$ is a scalar with unit variance over time). Consequently, in this model, neural variability is the same as shared variability, and private variability is not considered. As shown by Equation 5, even if the external input noise was stimulus independent, the shared variability (δr) would depend on the stimulus, because it depends on the neuronal gains (F, the slopes of the activation function at each neuron's firing rate), which in turn depend on the stimulus-driven neuronal firing rates (r^*), given that the activation function of each neuron ($f(\cdot)$) is nonlinear.

Previous work proposed that the stimulus dependence of shared variability can be well described by an affine model,[7,24] meaning that the amplitude of shared variability at each neuron can be approximated by an affine function of the neuron's trial-averaged responses. According to the circuit mechanism, when would the stimulus dependence of shared variability be affine? Based on our previous assumptions, the amplitude of shared variability φ can be expressed as $\varphi = (I - FW)^{-1}Fa$ (see Box 1). If we further assume that the activation function follows a
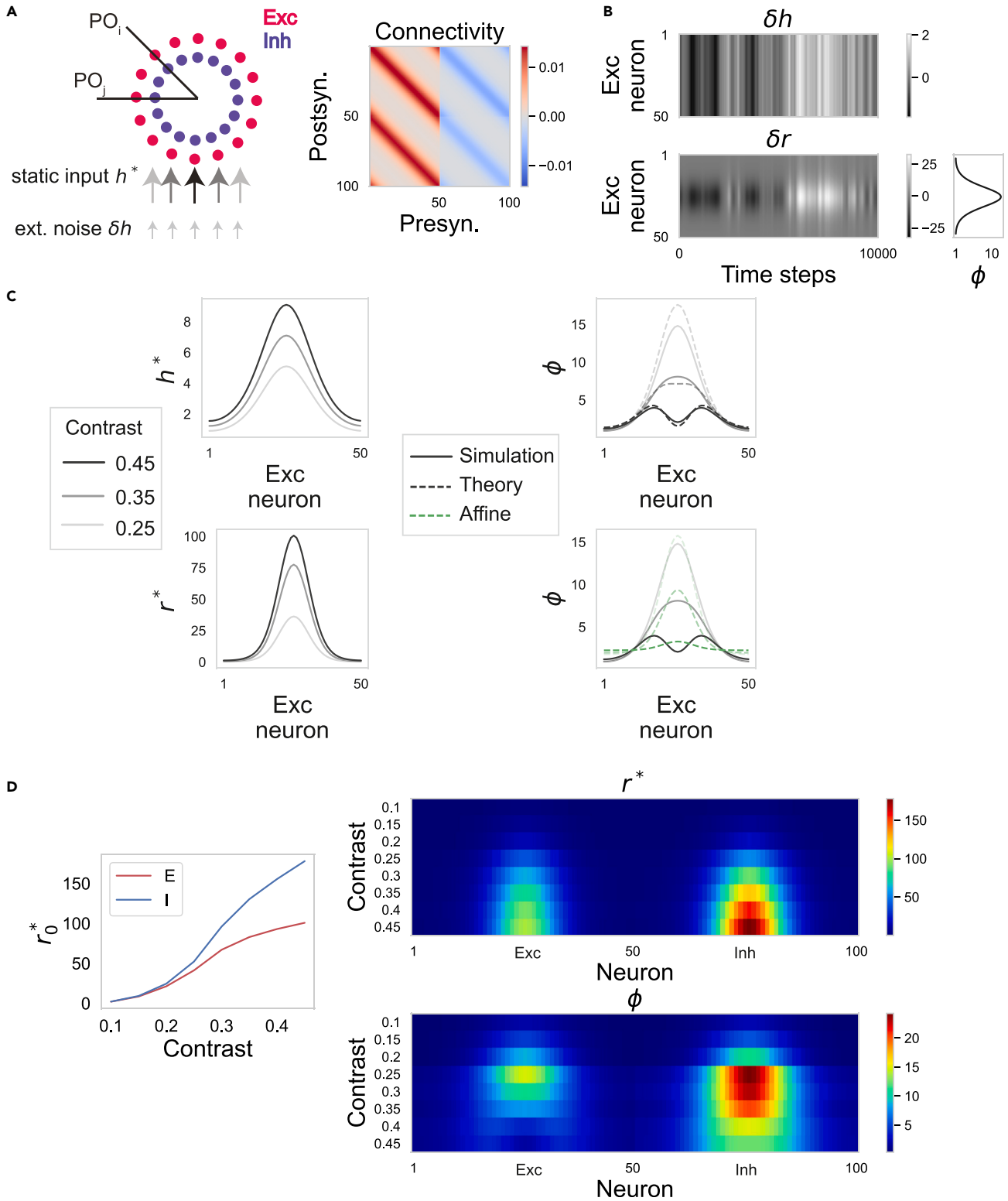
Figure 1. Circuit mechanism for stimulus-dependent shared variability

(A) Left: a schematic of the E-I neural circuit with a ring architecture. Excitatory neurons (red circles) and inhibitory neurons (blue circles) have preferred orientations (POs) based on their angular positions on the ring. Two external inputs are denoted by the arrows in the bottom. Right: the connectivity strengths between model

Figure 1. Continued

neurons. The first 50 neurons are excitatory, and the last 50 neurons are inhibitory. For simplicity, we ignore differences in numbers of E vs. I neurons. See Table 1 for a list of parameters used in the circuit model.

(B) Top: perfectly correlated external noise. Bottom: the corresponding residual responses (responses with time-averaged activity subtracted) given a visual stimulus centered between neurons 25 and 26. On the right, we plot the amplitude of shared variability φ across neurons. Here, we only show results for excitatory neurons.

(C) Left top: three external static inputs h^* representing visual inputs with different contrasts received by the model neurons (gray to black denotes low to high contrast). Left bottom: the corresponding time-averaged responses r^* of the model neurons. Right top: φ at each neuron from numerical simulation (solid line) and analytical estimation (dashed line; see Equation 5). Right bottom: the prediction of φ from the affine models (green dashed line) plotted against the simulated φ.

(D) Left: the simulated time-averaged firing rates of E and I neurons with PO matching the stimulus orientation, plotted against stimulus contrast. Right: Top: the simulated time-averaged firing rates of all neurons under stimuli with different contrasts. Bottom: the shared variability amplitudes of all neurons under stimuli with different contrast levels.

rectified power law and the firing rates are low enough so that effective recurrent connections (the product of synaptic strengths and neuronal gains) are negligible (see Box 1), then the amplitude of shared variability would be proportional to $r^{*(1-1/n)}$ (here r^* is the trial-averaged activity, and n is the exponent of the power law in the activation function). The tuning curve of δr in practice can be well approximated by an affine transformation of the tuning curve of r^*. This approximation holds since the exponent $1 - 1/n$ primarily acts to flatten the tuning curve.

To investigate the case where effective recurrent connections are strong, we directly simulated a recurrent neural network with bump-shaped orientation tuning curves, using a classical ring architecture[26,45] (Figure 1A). Neurons in the network receive two external inputs: (1) a tuned static input representing the visual stimulation and (2) perfectly correlated external noise that is slow and small (Figure 1B). We calculated the amplitude of shared variability numerically from the network simulation (see STAR Methods, Figure 1B) and analytically according to Equation 5 (Figure 1C). Note that the analytical solution (Equation 5) only approximates the circuit dynamics described by Equation 3, resulting in small differences between simulation results and the analytical predictions (upper right panel in Figure 1C). For simplicity, we set the amplitudes of external noise to be the same across all recipient neurons, so that due to the ring symmetry, the dependence of shared variability on stimulus orientation is the same as its dependence on neuronal preferred orientation (see Figure S1 for similar results when we have external noise with heterogeneous amplitudes across recipient neurons). Interestingly, even when effective recurrent connections are not negligibly small as assumed in Box 1, shared variability showed affine-like modulation across orientations for a wide range of contrast levels (Figures 1C and 1D). However, the circuit model results also showed two signatures that cannot be captured by affine models, due to recurrent dynamics. First, at high contrast levels, the shared variability amplitude exhibits an "M"-shaped dependence on stimulus orientation (Figure 1C); experimentally, "M"-shaped noise correlation has been observed in monkey area MT.[6] Second, with increasing contrast, shared variability amplitude first increases and then decreases (Figure 1D; Hennequin et al.[26]); such a decrease in shared variability amplitude with increasing stimulus strength has been observed in many cortical systems.[25] In the supplemental information (Figures S1–S3), we further analyze the behavior of shared variability in the ring circuit model to obtain an intuition for the origin of these two discrepancies from the affine models (building on Hennequin et al.[26]).

In summary, through an analysis of the dynamics in a recurrent network subject to common input noise, we suggest an origin for affine-like shared variability across orientations when the activation function adheres to a rectified power law. Additionally, we establish the need for the development of a novel statistical model to more effectively capture the stimulus-dependent nature of shared variability when both stimulus orientation and contrast are varied.

Affine models parsimoniously explain shared variability within V1 and V2 across stimulus orientations

To characterize the nature of shared variability in V1 and V2, we first analyzed electrophysiological recordings from macaque V1 and V2 under stimulation by drifting gratings of varying orientations and a single (100%) contrast (see descriptions of dataset 1 in STAR Methods). A given experimental session had 400 trials of each stimulus. Our goal is to identify a statistical model that explains the shared variability across stimuli parsimoniously. We modeled single-trial neural responses as a sum of three elements (Figure 2A): trial-averaged responses; neural variability shared across neurons, which itself may have multiple components (each component is a rank-1 matrix); and neural variability private to each neuron. For neuron c on trial i with stimulus s, its response is modeled as:

$$r_{c,i(s)} = d_{c,s} + \sum_r \varphi_{c,r,s} a_{r,i} + \epsilon_{c,i(s)}$$

$d_{c,s}$ is the trial-averaged response of neuron c to stimulus s; $a_{r,i}$ is a standard Gaussian variable, $a_{r,i} \sim \mathcal{N}(0,1)$, that governs how the rth component of the shared variability varies across trials; $\varphi_{c,r,s}$ denotes the amplitude of the rth component of the shared variability for neuron c with stimulus s (for a fixed s, vectors $\varphi_{r,s}$ are linearly independent across the component index r); and $\epsilon_{c,i(s)}$ represents the private variability, which on each trial is a sample from a Gaussian distribution with neuron- and stimulus-specific variance $\epsilon_{c,i(s)} \sim \mathcal{N}(0, \sigma_{c,s}^2)$.

We considered four types of models with different constraints on the stimulus dependence of φ (Figure 2C). For the additive model, the amplitude of shared variability does not vary across stimuli ($\varphi_{c,r,s} = \varphi_{c,r}$). For the multiplicative model, the amplitude of shared variability is proportional to the trial-averaged activity ($\varphi_{c,r,s} = \alpha_{c,r} d_{c,s}$). For the affine model, the amplitude of shared variability is constrained to be an affine transformation of the trial-averaged response ($\varphi_{c,r,s} = \alpha_{c,r} d_{c,s} + \beta_{c,r}$). For the generalized model, there are no constraints on the

Table 1. Parameters used in the circuit model with a ring architecture

N_E	50	–
N_I	50	–
τ_E	20	ms
τ_I	10	ms
k	0.3	$mV^{-n}\cdot s^{-1}$
n	2.5	–
W_{EE}	0.25	$mV\cdot s$
W_{IE}	0.24	$mV\cdot s$
W_{EI}	0.13	$mV\cdot s$
W_{II}	0.1	$mV\cdot s$
τ_{noise}	500	ms
σ^2_{noise}	0.5	mV^2
l_{syn}	45	$deg.$
l_{stim}	60	$deg.$
b	0.1	mV
A_{max}	20	mV

stimulus dependence of $\varphi_{c,r,s}$, as it can vary across s freely (i.e., it is non-parametric). In terms of number of parameters in the model, additive model ($2NS + NR$ parameters) = multiplicative model ($2NS + NR$ parameters) < affine model ($2NS + 2NR$ parameters) < generalized model ($2NS + NRS$ parameters), where N is the number of neurons, S is the number of stimuli, and R is the number of components of shared variability. If we have enough data so that none of the statistical models overfit, then we should always expect to see that the generalized model performs the best. Importantly, it is possible that, even without overfitting, a simpler model (e.g., additive or multiplicative model) can perform as well as the generalized model. In that case, we conclude that the simpler model is the best model for capturing the data, as it contains fewer parameters but achieves the same performance (see Figure S3 for fitting performances of statistical models on the surrogate data with known ground truth). Note that our proposed statistical models have a close relation with factor analysis: the generalized model is equivalent to applying factor analysis separately to the residual responses to each stimulus; the additive model is similar to applying factor analysis to the set of residual responses to all stimuli, except that it assumes stimulus-dependent private variability instead of stimulus-independent private variability.

We fit the four models to V1 and V2 neuronal responses separately, by maximum likelihood estimation with 5-fold cross-validation. We evaluated the fitting performance of each model by two cross-validated measures: (1) the log likelihood of the experimental data given the model (Figure 3A) and (2) the R^2 of the model noise covariance in accounting for the experimental noise covariance (see STAR Methods, Figure 3B). The log likelihood quantifies overall how well the variability (both shared and private) is captured by the statistical models, whereas the R^2 evaluates specifically how well the shared variability is captured by the statistical models.

All 5 experimental sessions showed the following results consistently (see Figure 3 for results from one session; Figure S5 for results from other sessions). First, quantitatively, the generalized model performs the best among all the statistical models, especially in terms of R^2 (Figures 3A and 3B). Second, qualitatively, the affine model performs almost as well as the generalized model in terms of capturing how shared variability amplitude changes across orientations (Figure 3C, inferred φ exhibits similar patterns for the two models). Third, with less than 5 components in the shared variability, we captured most of the explainable variance in the neural data (Figure 3; Figure S5). Additionally, inferred private variability has an amplitude proportional to the trial-averaged responses, with proportionality constant > 1 (more variable than Poisson noise; Figure S4).

Interestingly, we observed that experimental sessions with high firing rates tend to be better fit by additive models than multiplicative models (Figure S5), suggesting that they exhibit weaker stimulus dependence of the shared variability. This observation aligns with the theoretical predictions (see Box 1) and is supported by the circuit model simulations (Figure S5).

Generalized affine models explain how shared variability depends on stimulus contrasts and orientations jointly

We predict that affine models will fail to explain how shared variability is modulated by contrasts and orientations jointly. This is because the circuit mechanism (Figures 1B and 1D), as well as previous experimental and theoretical evidence,[25,26] suggests that the amplitude of shared variability should decrease with increasing stimulus contrasts, which cannot be captured by affine. To test this prediction, we recorded from V2 neurons using neuropixel probes while presenting drifting gratings with 8 orientations and 3 contrast levels (see descriptions of dataset 2 in STAR Methods and Figure S2).

Motivated by results of the circuit model simulation (Figures 1B and 1D) and fits to experiments varying orientation only (Figure 3), for a given contrast level, we expect the dependence of shared variability on orientations to be well approximated by an affine model. Therefore, we proposed a generalized affine model. This model incorporates contrast-specific affine coefficients α and β, where $\varphi_{c,r,s} = \alpha_{c,r,contrast} d_{c,s} + \beta_{c,r,contrast}$.
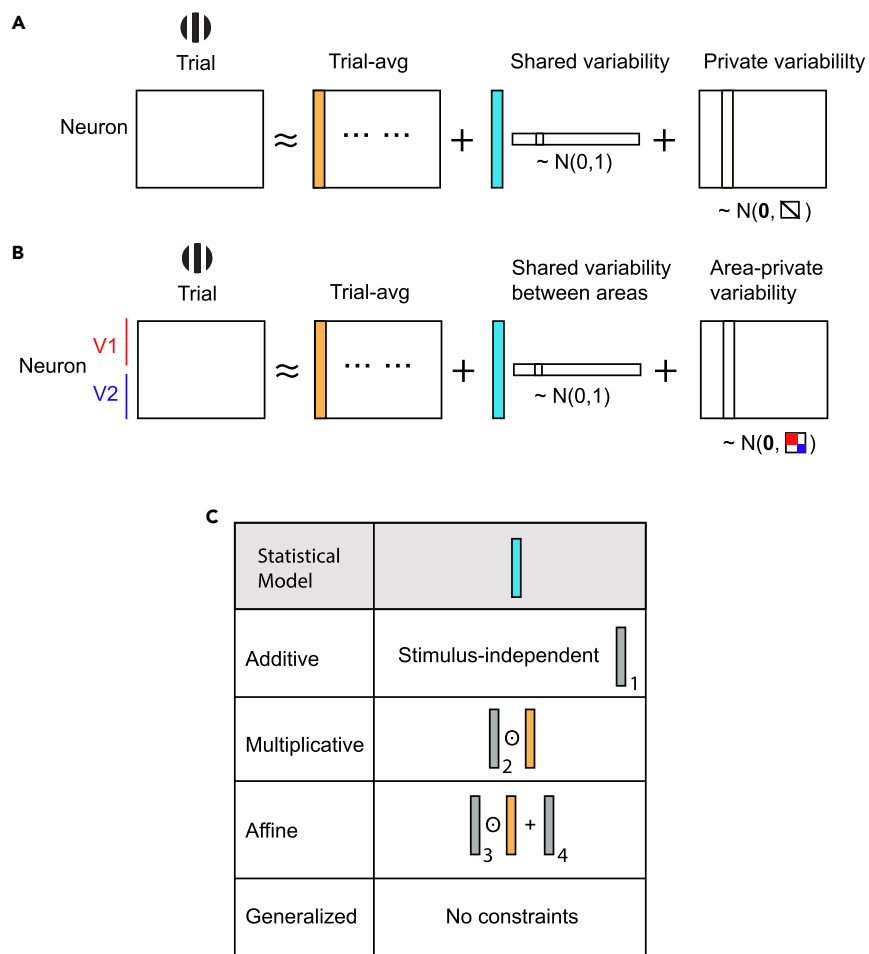
Figure 2. A schematic of the structures of statistical models

(A) The statistical model for capturing trial-to-trial variability shared within each brain area. For a given stimulus, we model the single-trial population neuronal responses within a brain area as a sum of 3 terms: the trial-averaged activity (orange column, representing pattern of trial-averaged activity across neurons; repeated identically for each trial), a 1-dimensional Gaussian variable shared across neurons (blue column representing pattern across neurons of shared variability, multiplied by a scalar Gaussian variable for each trial, represented by row; we also studied the case with multi-dimensional shared variability), and private variability that is an independent Gaussian variable for each neuron and trial.

(B) The statistical model for capturing trial-to-trial variability shared between two brain areas, V1 and V2. Similarly to A, we model the single-trial population neuronal responses as a sum of 3 terms, except that the private variability is not independent across neurons, but is independent between the two brain areas (variability within each area represented by red or blue blocks in illustrated covariance matrix, white indicates values of 0).

(C) We set up 4 types of statistical models with different constraints on the stimulus dependence of the shared variability amplitudes (i.e., of the blue column). For additive models, the shared variability amplitude is constant across stimuli (but varies across cells, indicated by gray vector). For multiplicative models, the shared variability amplitude is the element-wise product ($\odot$) of a stimulus-independent vector and the stimulus-dependent trial-averaged population vector. For affine models, the shared variability is the sum of multiplicative and additive components.

We introduced this model as a model complex enough to capture the stimulus dependence predicted by the circuit mechanism yet simple enough to avoid overfitting. Due to session duration limits and an increasing number of stimuli, we had to reduce the number of trials for each stimulus. Consequently, we only have 100 trials per stimulus. With this decreased amount of data, the generalized model overfits: as seen in our results (Figures 4A and 4B), even though the generalized models fit the data best in the training set, they perform poorly in the test set. In contrast, the generalized affine model does not overfit.

Judging by the log likelihood and R^2 of the noise covariance (Figures 4A and 4B), the generalized affine model performs the best when both contrasts and orientations are varied. To show that an unmodified affine model indeed fails to capture the modulation of shared variability in this case, we visualized the contrast dependence of the noise covariance, averaged over neuron pairs and stimulus orientations. As predicted by the circuit model, the amplitude of shared variability and the averaged noise covariance decreases with increasing contrast, which is captured by the generalized affine model and the generalized model, and importantly, not by the affine model (Figure 4C; Figure S7).
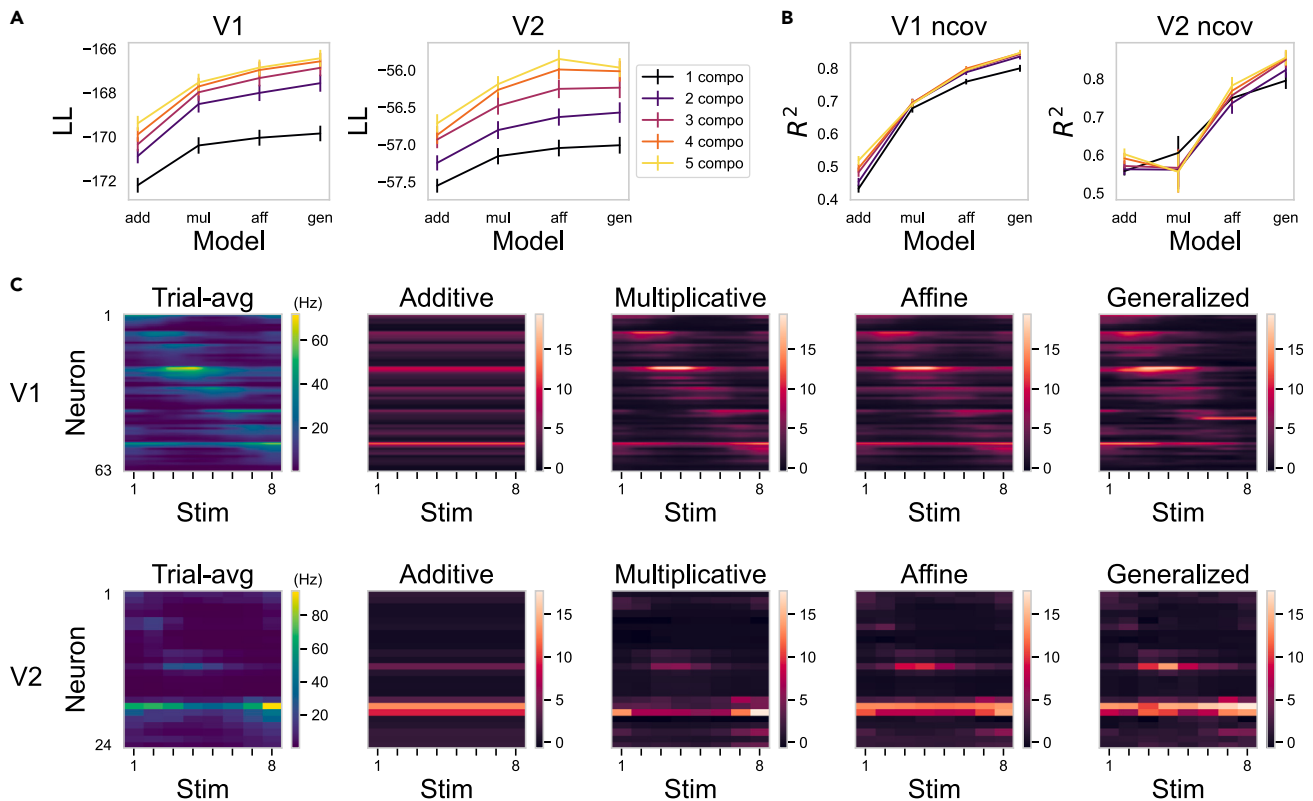
Figure 3. Affine models parsimoniously explain shared variability within V1 and V2 across stimulus orientations

Fitting results of statistical models of experimental session 1 with 8 orientations and a single contrast, 100%.

(A) The log likelihood of the test set. Error bars show the standard error of 5-fold cross-validation. Color denotes the number of components of the shared variability. Note that we ordered the statistical models based on their complexity in ascending order from left to right.

(B) The R^2 of the noise covariance ("ncov") across stimuli of the test set. Error bars show the standard error of 5-fold cross-validation. Color denotes the number of components of the shared variability.

(C) Fitted φ of the 4 statistical models with 1 component in the shared variability is shown against the trial-averaged responses (units Hz) across 8 stimulus orientations. Neurons are ordered according to their preferred orientations. The top row shows the result for V1 neurons. The bottom row shows the result for simultaneously recorded V2 neurons.

Circuit models predict affine-like variability shared between V1 and V2

So far, our analysis has focused on shared variability within each visual area. However, a significantly underexplored topic in the existing literature is the structure of variability shared between two interconnected visual areas. To gain some expectation on how variability shared between V1 and V2 should depend on the stimulus, we simulated a circuit model with two connected ring structures (Figure 5A), where each ring represents one visual area. For simplicity, we assume both V1 and V2 show stronger connectivity between neurons with similar preference for the only stimulus parameter used in our experiment, i.e., orientation. We assumed that connectivity from area 1 to area 2 is similar to connectivity within each area, except without inhibitory projections (Figure 5A).

We simulated the response of neurons in the connected ring network to changing stimulus orientation (Figure 5B). Area 1 received external inputs that consist of an orientation-tuned static external input and 1-dimensional external input noise. To calculate the amplitude of 1-dimensional variability shared within each area, we looked at the dominant component found by applying singular value decomposition to the residual activity of neurons within each ring, which is equivalent to the generalized model in the absence of private noise. To calculate the amplitude of 1-dimensional variability shared between areas, we looked at the dominant component found by applying probabilistic canonical correlation analysis[46] to the set of residual responses of the two areas. This is equivalent to a generalized joint model.

The variability shared between areas 1 and 2 consists of a pattern of neural firing across the two areas (Figure 2B). The portion of this pattern within area 1 may be similar to (Figure 5) or quite different from (Figure S8) area 1's pattern of purely within-area shared variability, depending on the chosen connectivity between area 1 and area 2. (Note that, because area 2 does not receive any external noise, its within-area and between-area noise always resemble one another.) Essentially, if the connectivity from area 1 to area 2 has a reasonably strong component along the pattern of within-area-1 shared variability, this pattern will propagate to area 2, resulting in a between-area shared variability pattern that resembles the within-area-1 shared variability (Figures 5B and 5C). Conversely, if the between-area connectivity and the
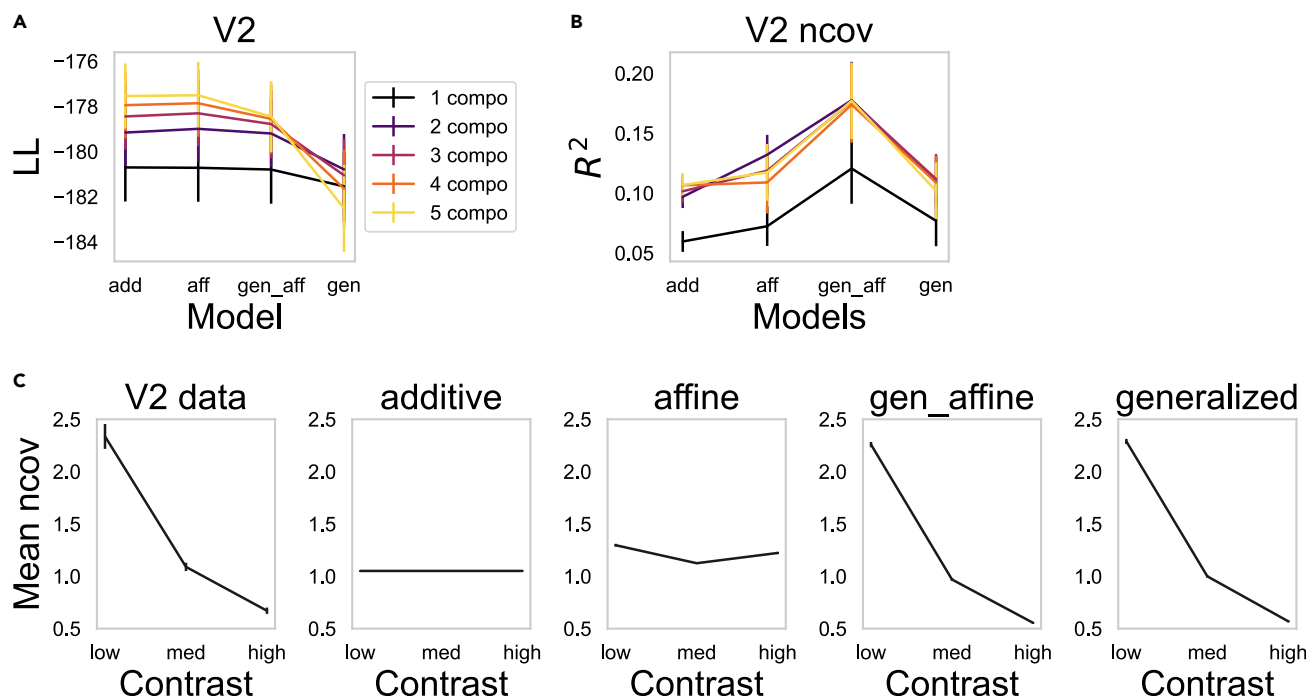
Figure 4. Generalized affine models explain how shared variability depends on stimulus contrasts and orientations jointly

Fitting performances of statistical models on the simultaneously recorded V2 data under stimulation by drifting gratings with 8 orientations and 3 contrast levels.

(A) The log likelihood of 4 different statistical models on the test set. Conventions as in Figure 3A. ("gen_aff": generalized affine).

(B) The R^2 of the noise covariance across stimuli of the test set. Conventions as in Figure 3B.

(C) The noise covariance averaged over neuron pairs and stimulus orientations is plotted against stimulus contrast levels. The first column shows the noise covariance calculated from V2 data in the test set. Error bars show the standard error of 5-fold cross-validation. The second, third, and fourth column show noise covariance calculated from different fitted statistical models with 2 components.

pattern of within-area-1 shared variability are close to orthogonal, area 1's pattern of between- and within-area shared variability may differ (see an example simulation in Figure S8). It is reasonable to assume that the trial-averaged activity in V1 has a strong component along the feedforward connectivity; otherwise, V2 would not be strongly activated. Building upon our previous analysis, which indicates that the dominant pattern of shared variability within V1 aligns closely with trial-averaged activity and is affine (Figure 3C), we predict that shared variability between visual areas will exhibit a similar orientation-dependent pattern to shared variability within each visual area and can be effectively captured by affine models.

Affine joint models parsimoniously explain variability shared between V1 and V2 across stimulus orientations

We now characterize the stimulus dependence of variability shared between V1 and V2, in responses to stimuli of varying orientation at a fixed (100%) contrast. We use four statistical models like the four used previously, modified for the two-area case (Figure 2B).

$$r_{c,i(s)} = d_{c,s} + \sum_r \varphi_{c,r,s} a_{r,i} + \epsilon_{c,i(s)}$$

The key differences from the within-area models are the following: first, we fit the statistical models to V1 and V2 data jointly as opposed to separately; thus, we refer to these statistical models as "joint statistical models." Second, here, $\varphi_{c,r,s}$ denotes the amplitude of the rth component of shared variability between areas for neuron c and stimulus s, instead of shared variability within one area. Third, we assume the private variability $\epsilon_{c,i(s)}$ is private to each brain area instead of private to each neuron (and thus can be shared within an area). Specifically, we assume that on each trial i for a given stimulus s, $\epsilon_{i(s)}$, which is a vector with length of the total number of neurons in V1 and V2, is a sample taken from a multivariate Gaussian distribution $\mathcal{N}(0, \Sigma_s)$. We assume that Σ_s is block diagonal, having non-zero entries only for within-V1 or within-V2 covariances. Note that this area-private variability ($\epsilon_{c,i(s)}$) includes both private variability at each neuron and variability that is shared only among neurons within a single brain area.

As before, we fit joint statistical models to V1 and V2 data by maximum likelihood estimation with 5-fold cross-validation, and we evaluated the performances of the joint statistical models with the previously mentioned two measures. Without overfitting, we should expect to see that the generalized model performs the best (see Figure S9 for fitting performances of joint statistical models on surrogate data with known ground truth).
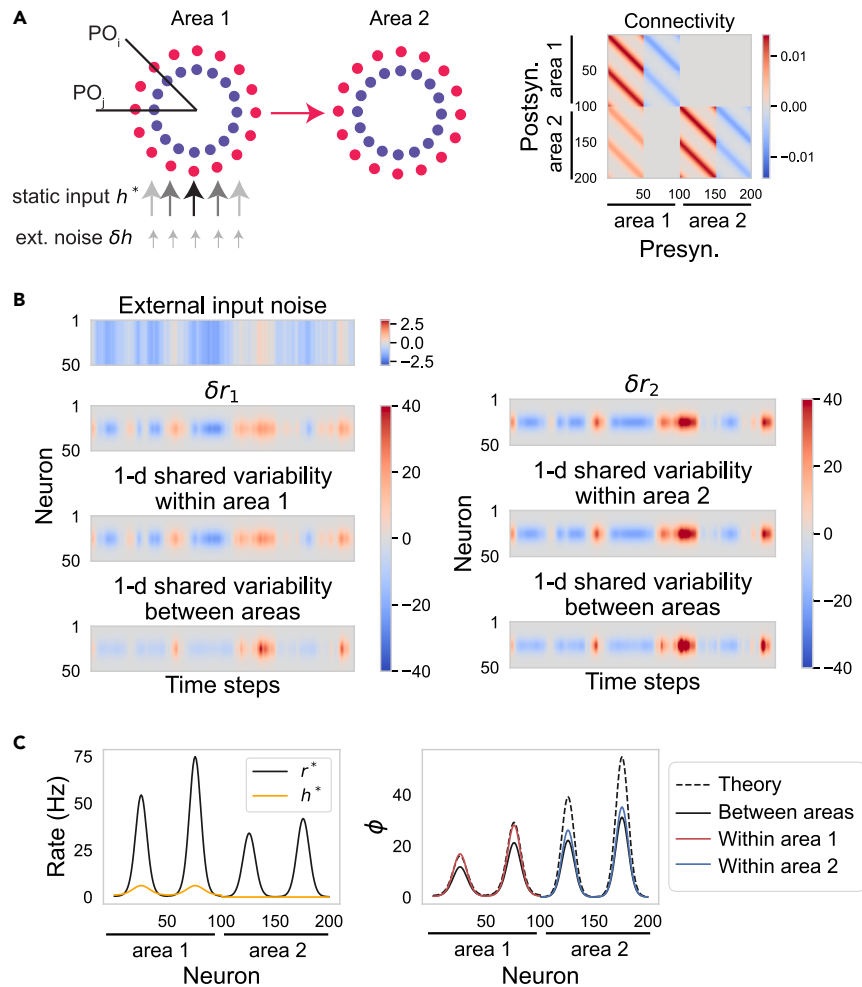
Figure 5. Circuit models predict affine-like variability shared between V1 and V2

(A) Left: schematic of two-ring models, with each ring representing one brain area. Between-area connections are excitatory and unidirectional, from area 1 to area 2. Area 1 receives an excitatory external input. Right: the connectivity strengths between the model neurons. The first 100 neurons are 50 excitatory and 50 inhibitory neurons in area 1, and the last 100 neurons are in area 2.

(B) Residual activity of excitatory neurons in area 1 (left) and area 2 (right). Left: (from top to bottom panels) external input noise δh received by the excitatory neurons in area 1; residual activity δr of the excitatory neurons in area 1; the dominant 1-dimensional variability shared within area 1; the dominant 1-dimensional variability shared between two areas. Right: similar to figures on the left but for area 2. Note that area 2 does not receive any external input noise directly.

(C) Left: the external static input h^* and the time-averaged rate responses r^* of all the simulated neurons in the connected ring model. Right: the 1-dimensional shared variability amplitudes of all the simulated neurons. Dashed: analytical prediction from Equation 5. Red: amplitude of variability shared within area 1. Blue: amplitude of variability shared within area 2. Black: amplitude of variability shared between two areas. The gap between analytical prediction and numerical simulation in area 2 is due to a violation of the small input noise assumption.

All 5 experimental sessions showed the following results consistently (Figure 6; Figure S10). Quantitatively, the generalized joint model performs the best in terms of capturing the shared variability between V1 and V2, slightly outperforming the affine joint model in terms of R^2. Qualitatively, the affine joint model inferred a similar stimulus dependence pattern for the shared variability amplitude as the generalized joint model (Figure 6C). As in Semedo et al.,[32,47] we found that 1 or 2 components for variability shared between areas are enough for capturing all the explainable variance in the experimental data. Importantly, as predicted by our circuit model simulation, variability shared between V1 and V2 exhibited similar orientation dependence as variability shared within V1 or V2 (see Figures 6C and 3C). We conclude that affine joint models parsimoniously explain how variability shared between V1 and V2 varies across orientations.

Additionally, we found that the amplitude of variability shared between areas is smaller than that within each area, as expected from the circuit model simulation (Figure 5C; Figures S11 and S12).
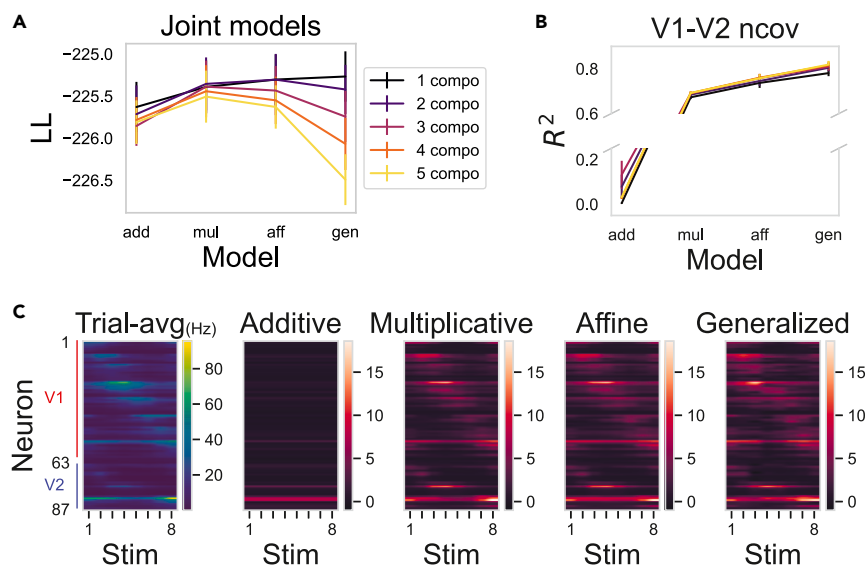
Figure 6. Affine joint models parsimoniously explain variability shared between V1 and V2 across stimulus orientations

Fitting results of joint statistical models for both V1 and V2 of experimental session 1 with 8 orientations and a single contrast, 100%.

(A) The log likelihood of 4 different joint statistical models on the test set. Conventions as in previous figures.

(B) The R^2 of the noise covariance between V1 and V2 across stimuli of the test set. Conventions as in previous figures. For visualization purposes, we omitted part of the y axis.

(C) Fitted φ of the 4 joint statistical models with 1 component is shown against the trial-averaged rates (units Hz) across stimuli with 8 orientations. Neurons are ordered according to their preferred orientations within each brain area.

DISCUSSION

We studied the stimulus dependence of shared variability within and between V1 and V2. When only stimulus orientation is varied, affine models effectively describe this stimulus dependence. However, when both orientation and contrast are varied, affine models must be modified to include contrast-specific coefficients to explain the suppression of shared variability with increasing contrast. Study of a recurrent neural circuit with power-law activation function, subjected to correlated stimulus-independent external input noise, suggests a mechanistic rationale for these statistical models.

Novelty of our work

In the previous literature, statistical models partition neural variability in one of two ways. First, motivated by the fact that spike count variance grows faster than its mean across stimuli, neural variability is partitioned into firing rate variability and spiking variability (usually modeled as Poisson noise). Second, as assumed by our work, neural variability is partitioned into shared and private variability. In general, the first method is used for explaining the variability of individual neurons,[22,48,49] whereas the second method is used for explaining variability in simultaneously recorded neuronal populations.[7,23,24] In principle, spiking variability (or firing rate variability) may contain both shared and private variability. However, if one assumes that spiking variability is private to each neuron and firing variability is shared across all the neurons, then the two partition methods would be approximately equivalent. In our study, we extend the findings using the second method in four key directions: (1) by comparing three previously proposed forms of modulation (additive, multiplicative, and affine) to an unrestricted form (generalized), we provide evidence that affine models offer a parsimonious explanation for how shared variability is modulated by stimulus orientations; (2) we establish a direct link between the statistical models and a neural circuit model, offering a straightforward mechanism for the observed affine shared variability; (3) we identified an alternative form of stimulus dependence of shared variability (generalized affine), which arises when stimulus strength (contrast) is varied; (4) we broaden the framework to explain variability shared between two connected brain areas, demonstrating that variability shared between V1 and V2 also exhibits an affine pattern across stimulus orientations.

When we only varied orientations of the grating stimuli, affine models performed quantitatively worse but qualitatively similarly to generalized models in capturing the stimulus-dependent shared variability (Figure 3). However, when we varied both orientations and contrast levels of the grating stimuli, affine models performed not only quantitatively but also qualitatively worse than the generalized affine model, because they could not capture the decrease in variability despite increasing firing rates induced by increasing contrasts (Figure 4C). Importantly, the seemingly diverse statistical behavior of shared variability across different stimulus features (Figures 3 and 4) as well as across experimental sessions (Figure S6) can be attributed to the same circuit mechanism (Box 1 and Figure 1). According to the circuit mechanism, the amplitude of shared variability depends on the stimulus through its dependence on the trial-averaged firing rates, and this dependence is modulated by recurrent connectivity and neural activation functions (see Equation 5).

Implications for stimulus coding

We found that shared variability in V1 and V2 in response to drifting gratings can be well described by affine (additive and multiplicative) modulations. How will affine shared variability impact stimulus coding?

Stimulus coding is often evaluated by how well we can discriminate two nearby (or distant) stimuli from population neuronal responses. Previous work[50] found that the linear discriminability of two nearby stimuli can be conveniently quantified by linear Fisher information: $I_F(s) = d_s'^T Q_s^{-1} d_s'$. Here, s is the stimulus variable (a scalar), d_s is trial-averaged population response at stimulus s (an $N \times 1$ vector function of s, N is the number of neurons), and d_s' is the derivative with respect to s of d_s. Q_s is the noise covariance at stimulus s. As long as Q_s has a component along $d_s' d_s'^T$, the linear Fisher information will saturate as N goes to infinity so that shared variability is detrimental to stimulus coding; otherwise, the linear Fisher information will keep increasing with increasing N.[3]

According to this theory, we can predict the impact of shared variability on coding stimulus variables, based on how much the amplitude of shared variability φ_s overlaps with d_s'. First, consider coding of stimulus orientation. If the preferred orientations of neurons are uniformly distributed, then for any s, d_s' will comprise roughly equal positive and negative entries. In addition, because the inferred affine coefficients are mostly positive (see Figures S11 and S12), φ_s will comprise mostly positive entries. Consequently, φ_s is almost orthogonal to d_s', and the affine shared variability will have limited impact on coding of orientation. Second, consider coding of stimulus contrast. The tuning curves of contrast are monotonically increasing so that all entries in d_s' are positive. In this case, φ_s is well aligned with d_s', so that the affine shared variability is detrimental to coding contrast. In addition, because the amplitude of shared variability decreases with increasing contrast, the value at which the linear discriminability of contrast saturates (with increasing N) will be higher for higher contrasts.

Note that being affine (or additive or multiplicative) does not necessarily constrain the direction of φ_s. In other words, identifying the stimulus dependence of shared variability is not very informative about how shared variability impacts coding. We can still draw conclusions about how the specific φ_s (or affine coefficients) we inferred may limit coding. However, while φ_s represents the dominant component of shared variability, even small components, if they are aligned with d_s', will cause the information coded to saturate with increasing N. Thus, conclusions about coding drawn from our model are limited to the effects of those variability components the model captures.

An alternative theory of stimulus coding posits that neurons encode a probability distribution, and neural variability represents sampling over that distribution.[51–53] In that case, the variability components identified by our model would help characterize that probability distribution.

Implications for circuit structure

Investigations of circuit mechanisms give insight into how external inputs can modulate neural variability. As shown in Box 1 and discussed in the study by Doiron et al.,[35] the alteration of patterns of trial-averaged activity induced by changes in stimuli will alter the patterns of variability, even when external input noise remains stimulus independent. Since behavioral variables and task conditions also alter trial-averaged activity levels, it is not surprising to observe that they too induce modulations in neural variability.

The circuit model also shows that, when firing rates become sufficiently high, and if external input noise is stimulus independent, then neural variability should become stimulus independent ($\delta r(t) \approx W^{-1} \delta h(t)$, see Box 1) and therefore well fit by an additive model. Indeed, we have found that in some sessions with particularly high firing rates, an additive model gives the best fit (Figure S6). In the other extreme, when the firing rates are sufficiently low to render effective connectivity very weak, the variability becomes independent of the recurrent connectivity, being determined by effective gains and external input noise ($\delta r(t) \approx F \delta h(t)$, see Box 1).

In most scenarios, recurrent connectivity affects the pattern of neural variability through Equation 5, which suggests that the variability is not necessarily always well described by affine models. Our simulation demonstrated how affine models can fall short, particularly when we vary the contrast level of the stimulus, which is supported by experimental data analysis (Figure 4). Furthermore, the circuit model predicts that the amplitude of shared variability can exhibit an "M"-shaped pattern across orientations in response to sufficiently strong tuned input (Figure 1), as has been observed in area MT.[6] However, our datasets are limited to only 8 orientations spanning a full 180° range (see STAR Methods). Consequently, our statistical models may not discern subtle variations, such as distinguishing between an "M"-shaped and a bump-shaped modulation. Alternatively, it is possible that even at the highest contrast level, V1 may not receive sufficiently strong tuned input to elicit the "M" shape. Future experiments with finer resolution within the stimulus space could resolve this issue.

In summary, our investigations of circuit mechanisms shed light on how external input can modulate neural variability. Statistical models effective in V1 and V2 may not be optimal for other brain areas, due to variations in recurrent connectivity. Nonetheless, insights into the two extreme cases of very low or very high firing rates are likely to be more general.

Implication for communication between brain areas

Previous work has shown that the communication subspace between V1 and V2, when stimulated by oriented drifting gratings, is limited to just 1 or 2 dimensions of perhaps 5 dimensions of variability within each area.[32,47] Such low dimensionality is unexpected for early processing stages like V1 and V2 in the visual hierarchy. However, the shared variability between V1 and V2 varies across stimuli, as shown by our work and a previous study,[32] so the combined communication subspace under different stimuli can be higher dimensional than only considering a single stimulus.

Limitations of the study

Our study has several limitations. Firstly, the linearized dynamics we employed as the basis for the circuit mechanism may not be accurate when trial-to-trial variability is large. A theoretical framework that incorporates nonlinear dynamics is needed for a better mechanistic model

of neural variability (e.g., see the study by Hennequin et al.[54]). Secondly, while the assumption of a Gaussian distribution for neural variability in the statistical models offers computational advantages, it is not accurate for describing discrete spike data. A discrete probability distribution such as a negative binomial distribution might be a better choice for modeling trial-to-trial variability of spike data. Thirdly, affine models (or modified affine models) may not generalize to other stimulus properties (besides orientation and contrast) or brain areas (besides V1 and V2). An approximately affine or additive model arises in the two limiting regimes of Box 1, but more complex dependencies of the amplitude of shared variability on trial-averaged responses can arise between these two regimes (see Equation 5). Therefore, we expect to find cases where the generalized model qualitatively outperforms affine models.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Neural recording and visual stimulation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data preprocessing
  - Fitting statistical models
  - Fitting joint statistical models
  - Evaluating model performance
  - Neural circuit model

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110512.

AUTHOR CONTRIBUTIONS

J.X. and K.D.M. designed the study. J.X. analyzed the data, performed the model simulation and prepared the manuscript. A.J. and A.K. designed and performed the experiments. All authors edited and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

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

2. Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nat. Rev. Neurosci. 7, 358–366.

3. Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nat. Neurosci. 17, 1410–1417.

4. Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and neuronal population information. Annu. Rev. Neurosci. 39, 237–256.

5. Kohn, A., and Smith, M.A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. J. Neurosci. 25, 3661–3673.

6. Ponce-Alvarez, A., Thiele, A., Albright, T.D., Stoner, G.R., and Deco, G. (2013). Stimulus-dependent variability and noise correlations in cortical MT neurons. Proc. Natl. Acad. Sci. USA 110, 13162–13167.

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

8. Zylberberg, J., Cafaro, J., Turner, M.H., Shea-Brown, E., and Rieke, F. (2016). Direction-selective circuits shape noise to ensure a

precise population code. Neuron *89*, 369–383.

9. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. Science *364*, 255.

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

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

12. Ruff, D.A., and Cohen, M.R. (2014). Attention can either increase or decrease spike count correlations in visual cortex. Nat. Neurosci. *17*, 1591–1597.

13. Bondy, A.G., Haefner, R.M., and Cumming, B.G. (2018). Feedback determines the structure of correlated variability in primary visual cortex. Nat. Neurosci. *21*, 598–606.

14. Haimerl, C., Savin, C., and Simoncelli, E. (2019). Flexible information routing in neural populations through stochastic comodulation. Adv. Neural Inf. Process. Syst. *32*.

15. Echeveste, R., Aitchison, L., Hennequin, G., and Lengyel, M. (2020). Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. Nat. Neurosci. *23*, 1138–1149.

16. Lange, R.D., and Haefner, R.M. (2022). Task-induced neural covariability as a signature of approximate Bayesian learning and inference. PLoS Comput. Biol. *18*, e1009557.

17. Trousdale, J., Hu, Y., Shea-Brown, E., and Josić, K. (2012). Impact of network structure and cellular response on spike time correlations. PLoS Comput. Biol. *8*, e1002408.

18. Ocker, G.K., Josić, K., Shea-Brown, E., and Buice, M.A. (2017). Linking structure and activity in nonlinear spiking networks. PLoS Comput. Biol. *13*, e1005583.

19. Priebe, N.J., Mechler, F., Carandini, M., and Ferster, D. (2004). The contribution of spike threshold to the dichotomy of cortical simple and complex cells. Nat. Neurosci. *7*, 1113–1122.

20. Hofer, S.B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., Lein, E., Lesica, N.A., and Mrsic-Flogel, T.D. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. Nat. Neurosci. *14*, 1045–1052.

21. Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjöström, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. Nature *473*, 87–91.

22. Goris, R.L.T., Movshon, J.A., and Simoncelli, E.P. (2014). Partitioning neuronal variability. Nat. Neurosci. *17*, 858–865.

23. Rabinowitz, N.C., Goris, R.L., Cohen, M., and Simoncelli, E.P. (2015). Attention stabilizes the shared gain of V4 populations. Elife *4*, e08998.

24. Arandia-Romero, I., Tanabe, S., Drugowitsch, J., Kohn, A., and Moreno-Bote, R. (2016). Multiplicative and additive modulation of neuronal tuning with population activity

affects encoded information. Neuron *89*, 1305–1316.

25. 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. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat. Neurosci. *13*, 369–378.

26. Hennequin, G., Ahmadian, Y., Rubin, D.B., Lengyel, M., and Miller, K.D. (2018). The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. Neuron *98*, 846–860.e5.

27. Nowak, L.G., Munk, M.H., James, A.C., Girard, P., and Bullier, J. (1999). Cross-correlation study of the temporal interactions between areas V1 and V2 of the macaque monkey. J. Neurophysiol. *81*, 1057–1074.

28. Jia, X., Tanabe, S., and Kohn, A. (2013). γ and the coordination of spiking activity in early visual cortex. Neuron *77*, 762–774.

29. Pooresmaeili, A., Poort, J., and Roelfsema, P.R. (2014). Simultaneous selection by object-based attention in visual and frontal cortex. Proc. Natl. Acad. Sci. USA *111*, 6467–6472.

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

31. Ruff, D.A., and Cohen, M.R. (2016). Attention increases spike count correlations between visual cortical areas. J. Neurosci. *36*, 7523–7534.

32. Semedo, J.D., Zandvakili, A., Machens, C.K., Yu, B.M., and Kohn, A. (2019). Cortical areas interact through a communication subspace. Neuron *102*, 249–259.e4.

33. Ruff, D.A., and Cohen, M.R. (2019). Simultaneous multi-area recordings suggest that attention improves performance by reshaping stimulus representations. Nat. Neurosci. *22*, 1669–1676.

34. Ruff, D.A., and Cohen, M.R. (2016). Stimulus dependence of correlated variability across cortical areas. J. Neurosci. *36*, 7546–7556.

35. Doiron, B., Litwin-Kumar, A., Rosenbaum, R., Ocker, G.K., and Josić, K. (2016). The mechanics of state-dependent neural correlations. Nat. Neurosci. *19*, 383–393.

36. Miller, K.D., and Troyer, T.W. (2002). Neural noise can explain expansive, power-law nonlinearities in neural response functions. J. Neurophysiol. *87*, 653–659.

37. Hansel, D., and van Vreeswijk, C. (2002). How noise contributes to contrast invariance of orientation tuning in cat visual cortex. J. Neurosci. *22*, 5118–5128.

38. Fourcaud-Trocmé, N., Hansel, D., van Vreeswijk, C., and Brunel, N. (2003). How spike generation mechanisms determine the neuronal response to fluctuating inputs. J. Neurosci. *23*, 11628–11640.

39. 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. (2014). State dependence of noise correlations in macaque primary visual cortex. Neuron *82*, 235–248.

40. Huang, C., Ruff, D.A., Pyle, R., Rosenbaum, R., Cohen, M.R., and Doiron, B. (2019). Circuit

models of low-dimensional shared variability in cortical networks. Neuron *101*, 337–348.e4.

41. De La Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. Nature *448*, 802–806.

42. Sompolinsky, H., Crisanti, A., and Sommers, H.-J. (1988). Chaos in random neural networks. Phys. Rev. Lett. *61*, 259–262.

43. Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat. Neurosci. *15*, 1498–1505.

44. Cohen, M.R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. Nat. Neurosci. *14*, 811–819.

45. Rubin, D.B., Van Hooser, S.D., and Miller, K.D. (2015). The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. Neuron *85*, 402–417.

46. Bach, F.R., and Jordan, M.I. (2005). A Probabilistic Interpretation of Canonical Correlation Analysis.

47. Semedo, J.D., Jasper, A.I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C.K., Kohn, A., and Yu, B.M. (2022). Feedforward and feedback interactions between visual cortical areas use different population activity patterns. Nat. Commun. *13*, 1099.

48. Churchland, A.K., Kiani, R., Chaudhuri, R., Wang, X.J., Pouget, A., and Shadlen, M.N. (2011). Variance as a signature of neural computations during decision making. Neuron *69*, 818–831.

49. Zhu, R.J.B., and Wei, X.-X. (2023). Unsupervised approach to decomposing neural tuning variability. Nat. Commun. *14*, 2298.

50. Beck, J., Bejjanki, V.R., and Pouget, A. (2011). Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. Neural Comput. *23*, 1484–1502.

51. Hoyer, P., and Hyvärinen, A. (2002). Interpreting neural response variability as Monte Carlo sampling of the posterior. Adv. Neural Inf. Process. Syst. *15*.

52. Lee, T.S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *20*, 1434–1448.

53. Orbán, G., Berkes, P., Fiser, J., and Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. Neuron *92*, 530–543.

54. Hennequin, G., and Lengyel, M. (2016). Characterizing Variability in Nonlinear Recurrent Neuronal Networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.1610.03110.

55. Smith, M.A., and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. J. Neurosci. *28*, 12591–12603.

56. Steinmetz, N.A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., et al. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. Science *372*, eabf4588.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
dataset 1	https://crcns.org/data-sets/vc/v1v2-1/about_v1v2-1 A mildly preprocessed copy is offered here: https://doi.org/10.5061/dryad.h9w0vt4s0	https://doi.org/10.6080/K0B27SHN
dataset 2	This study	https://doi.org/10.5061/dryad.h9w0vt4s0
Software and algorithms		
Deposited code	This study	https://github.com/tinaxia2016/ NeuronalVariabilityStatsModels

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ji Xia (jx2484@ columbia.edu).

Materials availability

The study did not generate new reagents.

Data and code availability

- Dataset 1 is available at (https://crcns.org/data-sets/vc/v1v2-1/), dataset 2 has been deposited at Dryad and is available at (https:// datadryad.org/stash/dataset/doi:10.5061/dryad.h9w0vt4s0).
- All original code has been deposited at https://github.com/tinaxia2016/NeuronalVariabilityStatsModels and is publicly available.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The study included data from four male macaques (Macaca fascicularis), 2-5 years old. Animals were housed in a 12-hour light-dark cycle and provided access to fresh water, food, including fresh fruit and vegetables and environmental enrichment. All procedures and animal care were approved by the IACUC of the Albert Einstein College of Medicine (protocol no: 00001407) and in accordance with NIH Guide for the Care and Use of Laboratory Animals.

METHOD DETAILS

Neural recording and visual stimulation

Animal procedures have been reported in previous work.[30,55] Briefly, animals (macaca fascicularis, 2-5 years old) were anesthetized with ketamine (10 mg/kg) and maintained on isoflurane during surgery. All recordings were performed under sufentanil (6-18 $\mu g/kg/hr$) anesthesia. Vecuronium bromide (150 $\mu g/kg/hr$) was used to prevent eye movements. All procedures were approved by the IACUC of the Albert Einstein College of Medicine.

Dataset 1 has been reported previously[30,32] and is publicly available under https://crcns.org/data-sets/vc/v1v2-1/. Recordings in V1 were performed using a Utah Array (96 channels, Blackrock Neurotech, USA) and recordings in V2 were performed using a set of tetrodes (Thomas Recording, Germany). Stimuli (full contrast drifting gratings, 8 orientations in steps of 22.5 deg, 1 cyc/d, drift rate of 3-6.25 Hz; 2.6-4.9 deg in diameter) were presented on a gamma corrected CRT screen (1024-796 pixel, 100 Hz refresh). Each stimulus was presented 400 times for 1.28 s preceded by an interstimulus interval of 1.5 s. Voltage snippets that exceeded a user defined threshold were digitized and sorted offline. The recorded V1-V2 populations had overlapping receptive fields.

Dataset 2 was recorded using four Neuropixel 1.0 (IMEC, Belgium) probes inserted in a rhomboid arrangement (see Figure S1) with the most anterior sites being roughly 3 mm posterior of the lunate sulcus. Data was recorded with a 'linear configuration' resulting in 384 sites distributed over 7.6 mm of the probe shaft. Area boundaries for V1 and V2 were determined by a combination of change in receptive field position and size and histological reconstruction of the electrode track. Stimuli (drifting gratings, 8 orientations in steps of 22.5 deg, 2 cyc/d, drift rate of 4 Hz; 8 deg in diameter; 15 %, 50 % and 100 % Michelson contrast with $L_{min} = 0$ cd/m^2 and $L_{max} = 80$ cd/m^2) were presented on a gamma corrected CRT screen (1024-796 pixel, 100 Hz refresh)." Each stimulus was presented 200 times for 0.5 s preceded by a 0.5 s inter

stimulus interval. Voltage traces were digitized using spikeGLX (https://billkarsh.github.io/SpikeGLX/) and sorted offline using Kilosort 2.5[56] and Phy2 (https://github.com/cortex-lab/phy). The V1 and V2 populations had partially overlapping receptive fields.

For dataset 2, we didn't show the fitting results of statistical models for V1, due to lack of sufficient data.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data preprocessing

For dataset 1, for each neuron and trial, we counted the spikes during the stimulus excluding the first 100 ms after stimulus onset. For dataset 2, for each neuron and trial, we counted the spikes during the stimulus. Additionally, we concatenated 2 trials as 1 trial to increase the number of spikes per trial, ensuring that the Gaussian assumption in our statistical models remains valid. We ended up with 100 trials per stimulus.

We quantified how well-tuned each neuron is by calculating the orientation tuning index (OTI). The neuron's tuning curve is just the trial average of its response at each orientation. OTI is computed as

$$\text{OTI} = \frac{d_{\max} - d_{\min}}{d_{\max} + d_{\min}}$$

where $d_{\max}$ and $d_{\min}$ are the maximum and minimum of the tuning curve, respectively. When fitting the statistical models, we excluded neurons with 0 spikes over trials for any given stimulus and excluded neurons with a fano factor > 4 in dataset 2 to avoid unstable units. Furthermore, we excluded neurons with OTI < 0.35 for both datasets. In dataset 1 (5 recording sessions), our study included 359 out of 564 units in V1 and 119 out of 147 units in V2. In dataset 2 (1 recording session), our study included 46 out of 132 units in V1 and 74 out of 191 units in V2.

Fitting statistical models

We have 5 statistical models for describing neural variability within each brain area: additive, multiplicative, affine, generalized affine and generalized models. We fit the models separately to V1 and V2 data with 5-fold cross-validation. At each fold, we split 400 trials into a training set of 320 trials and a test set of 80 trials, chosen at random. We fit the model to the training set with maximum likelihood estimation, and evaluated the fitting performance by the log likelihood of the test set. Our model assumes that the number of spikes r of N recorded neurons at trial i with stimulus s follows a multivariate Gaussian distribution $r_{i(s)} \sim \mathcal{N}(d_s, C_s)$, with the covariance $C_s = \sum_r \varphi_{r,s} \varphi_{r,s}^T + \Sigma_s$. Σ_s is a diagonal matrix with diagonal entries $\sigma_{c,s}^2$. The log likelihood is obtained as follows:

$$\sum_i \left[-\frac{N}{2} \log 2\pi - \frac{1}{2} \log \det C_s - \frac{1}{2}(r_{i(s)} - d_s)^T C_s^{-1} (r_{i(s)} - d_s)^T \right]$$

As noted in the Results, the generalized model is equivalent to factor analysis applied separately to the residual responses to each stimulus. To fit generalized models, we used the python class "sklearn.decom position.FactorAnalysis" with the default initialization parameters.

For the other 4 models, careful initialization was required for a fair comparison between models, because the log likelihood of our statistical models can have many local maxima. To initialize the parameters in the other 4 statistical models, we fit a naive version of the additive model with stimulus-independent private variability by applying factor analysis to population residual responses to all stimuli. Then the models were initialized as follows:

Initialization of the additive model:

- Initialize $\varphi_{c,r}$ with the fitted $\varphi_{c,r}$ of the naive additive model.
- Initialize $\sigma_{c,s}^2$ with the fitted $\sigma_{c,s}^2$ of the generalized model.

Initialization of the affine model:

- Initialize $\alpha_{c,r}$ as all zeros.
- Initialize $\beta_{c,r}$ as the fitted $\varphi_{c,r}$ from the naive additive model.
- Initialize $\sigma_{c,s}^2$ with the fitted $\sigma_{c,s}^2$ of the generalized model.

Initialization of the multiplicative model:

- Initialize $\alpha_{c,r}$ as fitted $\alpha_{c,r}$ from the affine model.
- Initialize $\sigma_{c,s}^2$ with the fitted $\sigma_{c,s}^2$ of the generalized model.

Initialization of the generalized affine model:

- Initialize $\alpha_{c,r,\text{contrast}}$ as all zeros.
- Initialize $\beta_{c,r,\text{contrast}}$ across all contrast levels as the fitted $\varphi_{c,r}$ from the naive additive model.
- Initialize $\sigma_{c,s}^2$ with the fitted $\sigma_{c,s}^2$ of the generalized model.

For additive, multiplicative, affine, and generalized affine models, we optimized the parameters by gradient descent with the Adam optimizer. We chose the learning rate to be 5e-3 or 1e-2 with a 0.96 decay rate. We stopped the optimization after the log likelihood of the training set converges or at the 20,000[th] iteration.

Fitting joint statistical models

We have 4 joint statistical models for describing neural variability between two brain areas: additive, multiplicative, affine and generalized joint models (we did not use generalized affine because contrast was not varied in these studies). We fit the models jointly to V1 and V2 data with 5-fold cross-validation.

We wanted to characterize the shared variability between V1 and V2, separate from shared variability within each area. This means that, in fitting a joint model to V1 and V2, the private variability is private to each area (can be correlated within each area), rather than being private to each neuron. For this reason, we cannot use the methods we applied to each single area. Instead, to fit generalized joint models, we applied probabilistic canonical correlation analysis (pCCA) to the population residual responses from V1 and V2 to each stimulus separately (as in Semedo et al.[47]). Similarly to the fitting process of within-area statistical models, we initialized the parameters in the additive, multiplicative and affine joint models with the fitted parameters from the generalized joint models or naive additive joint models with stimulus-independent area-private variability. Naive additive joint models were fitted by applying pCCA to the population residual response from V1 and V2 to all stimuli.

Evaluating model performance

We quantified how well the statistical models explain all the variance in the data by cross-validated log likelihood, as previously described (see STAR Methods section: fitting statistical models). We also quantified how well the statistical models explain the shared variability in the data by evaluating the cross-validated R^2 of the noise covariance matrix across all stimuli. Specifically, we computed the noise covariance matrix from the test set for each stimulus and concatenated the upper triangular elements of these matrices across all stimuli. We then computed the R^2 describing how well this concatenated vector was predicted by the statistical models fit to the training set. The noise covariance between neuron c_1 and neuron c_2 of a given stimulus s from the test set is calculated using the formula: $\text{Cov}_{c_1,c_2,s} = \frac{1}{k-1} \sum_{i(s)} (r_{c_1,i(s)} - d_{c_1,s})(r_{c_2,i(s)} - d_{c_2,s})$,

where k is the number of trials per stimulus in the test set, $i(s)$ is the trial index during stimulus s, r is the single-trial response, d is the trial-averaged response. The noise covariance of a given stimulus s from the fitted statistical model is calculated as $\text{Cov}_{c_1,c_2,s} = \sum_r \varphi_{c_1,r,s}\varphi_{c_2,r,s}$. Note

that we chose to look at noise covariance instead of noise correlation here, because noise correlation is also influenced by the stimulus-dependent private variability, not just the shared variability.

Similarly, we quantified how well the joint statistical models capture variability shared between V1 and V2 by cross-validated R^2 of the noise covariance between areas. Here, noise covariance is defined for all possible neuron pairs, where each neuron pair consists of one V1 neuron and one V2 neuron. The noise covariance is calculated using the formula described earlier, except that c_1 is the index of neurons in V1, and c_2 is the index of neurons in V2. We evaluated the R^2 for all elements in the noise covariance matrix between V1 and V2 across all stimuli.

Neural circuit model

We simulated a rate-based network with a ring architecture. The network contained 50 excitatory and 50 inhibitory units. The circuit dynamics was governed by:

$$\tau_i \frac{dr_i(t)}{dt} = -r_i(t) + f\left(\sum_j W_{ij} r_j(t) + h_i + \eta_i(t)\right)$$

r_i denotes the rate for neuron i, τ_i is the time constant for neuron i, W_{ij} is the synaptic strength from neuron j to neuron i, h_i is the external static input received by neuron i, η_i is the external input noise received by neuron i. The activation function f is defined as: $f(x) = k[x]_+^n$.

The connectivity matrix W is set to be translational invariant, and the strength of connections between two neurons follows a circular Gaussian of the difference in their preferred orientations:

$$W_{ij} = J_{\alpha\beta} \exp\left(\frac{\cos(2(\theta_i - \theta_j)) - 1}{l_{syn}^2}\right) \tag{Equation 6}$$

where $\alpha, \beta \in \{E, I\}$, $J_{\alpha\beta}$ is a scaling constant for synaptic strength between E and I neurons, set so that the sum of incoming E and I weights onto each E and I neuron matched the values of $W_{\alpha\beta}$. θ_i denotes the preferred orientation of neuron i. l_{syn} sets the length scale over which synaptic weights decay.

The external static input h_i representing stimulus-related input. It is constant and follows a circular Gaussian function with an added baseline:

$$h_i(\theta_{stim}) = b + cA_{max} \exp\left(\frac{\cos(2(\theta_i - \theta_{stim})) - 1}{l_{stim}^2}\right)$$

θ_{stim} denotes the orientation of the visual stimulus. The neuron with preferred orientation matching the stimulus orientation receives the strongest external static input, and l_{stim} sets the length scale over which the external static input decay. b denotes the constant baseline. c denotes the contrast level. A_{max} denotes the maximum amplitude of external static input.

The external input noise $\eta_i(t)$ is modeled as a multivariate Ornstein-Uhlenbeck process: $\tau_{\text{noise}}d\eta = -\eta dt + \sqrt{2\tau_{\text{noise}}\Sigma_{\text{noise}}}d\zeta$ Here we chose the external input noise to be perfectly correlated across neurons so that Σ_{noise} is a matrix with all the elements equals to a chosen variance σ_{noise}^2. In other words, we had 1-dimensional external input noise that is identical across neurons.

To numerically calculate the amplitude of shared variability φ, we applied SVD to the population residual responses. This is equivalent to factor analysis given that there is no private variability in the rate model. φ is calculated as the standard deviation of the dominant SVD component.

To study variability shared between two connected brain areas, we simulated two ring circuits with a feedforward connection from one to the other (Figure 5). To numerically calculate the amplitude of variability shared between the ring circuits, we applied pCCA to the population residual responses from the two ring circuits. Here, φ is calculated as the standard deviation of the dominant pCCA component.