Systems/Circuits

On the Structure of Neuronal Population Activity under Fluctuations in Attentional State

Alexander S. Ecker,[1,2,3,4] **George H. Denfield,**[2] **Matthias Bethge,**[1,3,4]* **and Andreas S. Tolias**[2,3,5]*

[1]Centre for Integrative Neuroscience and Institute for Theoretical Physics, University of Tübingen, 72076 Tübingen, Germany, [2]Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, [3]Bernstein Center for Computational Neuroscience, 72076 Tübingen, Germany, [4]Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany, and [5]Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005

Attention is commonly thought to improve behavioral performance by increasing response gain and suppressing shared variability in neuronal populations. However, both the focus and the strength of attention are likely to vary from one experimental trial to the next, thereby inducing response variability unknown to the experimenter. Here we study analytically how fluctuations in attentional state affect the structure of population responses in a simple model of spatial and feature attention. In our model, attention acts on the neural response exclusively by modulating each neuron's gain. Neurons are conditionally independent given the stimulus and the attentional gain, and correlated activity arises only from trial-to-trial fluctuations of the attentional state, which are unknown to the experimenter. We find that this simple model can readily explain many aspects of neural response modulation under attention, such as increased response gain, reduced individual and shared variability, increased correlations with firing rates, limited range correlations, and differential correlations. We therefore suggest that attention may act primarily by increasing response gain of individual neurons without affecting their correlation structure. The experimentally observed reduction in correlations may instead result from reduced variability of the attentional gain when a stimulus is attended. Moreover, we show that attentional gain fluctuations, even if unknown to a downstream readout, do not impair the readout accuracy despite inducing limited-range correlations, whereas fluctuations of the attended feature can in principle limit behavioral performance.

Key words: attention; gain modulation; noise correlation; population coding; variability

Significance Statement

Covert attention is one of the most widely studied examples of top-down modulation of neural activity in the visual system. Recent studies argue that attention improves behavioral performance by shaping of the noise distribution to suppress shared variability rather than by increasing response gain. Our work shows, however, that latent, trial-to-trial fluctuations of the focus and strength of attention lead to shared variability that is highly consistent with known experimental observations. Interestingly, fluctuations in the strength of attention do not affect coding performance. As a consequence, the experimentally observed changes in response variability may not be a mechanism of attention, but rather a side effect of attentional allocation strategies in different behavioral contexts.

Introduction

Attention was traditionally thought of as acting by increasing the response gain of a relevant population of neurons (Reynolds and Chelazzi, 2004; Maunsell and Treue, 2006). More recent studies found that attention also reduces pairwise correlations between neurons (Cohen and Maunsell, 2009; Mitchell et al., 2009; Herrero et al., 2013). Based on a simple pooling model (Zohary et al., 1994), these authors argued that the benefits of increased gain

Received May 27, 2015; revised Nov. 23, 2015; accepted Dec. 11, 2015.

Author contributions: A.S.E., G.H.D., M.B., and A.S.T. designed research; A.S.E. performed research; A.S.E. analyzed data; A.S.E., G.H.D., M.B., and A.S.T. wrote the paper.

This work was supported by the Bernstein Center for Computational Neuroscience, Tübingen, Germany (FKZ 01GQ1002), the German Excellency Initiative through the Centre for Integrative Neuroscience Tübingen (EXC307), National Eye Institute–NIH Grants R01-EY018847 and P30- EY002520-33 to A.S.T., and the National Institutes of Health Pioneer Award DP1-OD008301 to A.S.T. We thank Ralf Haefner and Philipp Berens for helpful discussions and comments.

The authors declare no competing financial interests.

This article is freely available online through the *J Neurosci* Author Open Choice option.

*M.B. and A.S.T. contributed equally to this study.

Correspondence should be addressed to Alexander S. Ecker, Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Strasse 25, 72076 Tübingen, Germany. E-mail: alexander.ecker@uni-tuebingen.de.

DOI:10.1523/JNEUROSCI.2044-15.2016

are dwarfed by the effects of reduced correlations; therefore, attention is more appropriately viewed as shaping the noise distribution.

However, in an experiment, the subject's state of attention can be controlled only indirectly and is bound to vary from one trial to the next. As a consequence, measuring neuronal variability or correlations under attention has a fundamental caveat: it is unclear to what extent the observed neuronal covariability reflects interesting aspects of information processing in the neuronal population or simply trial-to-trial fluctuations in the subject's state of attention, which is unknown to the experimenter. Despite ample evidence that attention fluctuates from trial to trial (Cohen and Maunsell, 2010, 2011), the effects of such fluctuations on neuronal population activity have so far not been investigated.

Here we analyze a simple neural population model, where neurons with overlapping receptive fields encode the direction of motion of a stimulus (see Fig. 1*A*). We assume that neurons produce spikes independently according to a Poisson process with rate λ_i and treat attention as a process that modulates the neurons' gain (see Fig. 1*B*). The firing rate of neuron i is given by

$$\lambda_i = g_i f_i(\theta), \tag{1}$$

where g_i is the attentional gain (a combination of spatial and feature attention) and $f_i(\theta)$ is the direction tuning curve. We assume that there is always a stimulus in the neurons' receptive field, but this stimulus is not necessarily attended. Crucially, in our model, the subject's attentional state is not constant across trials, even within the same attentional condition. Thus, g_i is a random variable that varies from trial to trial (see Fig. 1*C*), and its precise value is unknown to the experimenter. As a consequence, the correlations in g_i across neurons will induce correlations between the observed neural responses.

In the following, we analyze this correlation structure in detail. We find that the correlations induced by attentional fluctuations resemble many experimentally observed aspects of correlated variability, such as correlations that increase with firing rates, limited range correlations, and differential correlations. In addition, we investigate the consequences of correlations induced by fluctuating attentional gain for reading out the direction of motion of the stimulus from the population response. We show that such correlations do not impair readout, even if the decoder does not have access to the attentional state. Finally, we show that our model can account for a number of nontrivial experimental findings on correlated variability in attention paradigms.

A preliminary account of these findings has been presented previously at the Cosyne Meeting 2012 (Ecker et al., 2012). Related ideas have been developed independently by another group, whose results have been published recently (Rabinowitz et al., 2015).

Materials and Methods

This section contains a detailed description of the model and the derivations of the main results. In an effort to make the paper as accessible as possible, the Results section is self-contained. Readers not interested in the detailed derivations can skip ahead directly to Results.

Notation

We use uppercase italic letters to denote matrices, lowercase italic letters for scalar values, and lowercase boldface letters for vectors. Thus, M is a matrix and v_i is the i^{th} element of vector $\mathbf{v}$. We write the expectation of a random variable x as $\langle x \rangle$ and the conditional expectation of x given y as $\langle x \,|\, y \rangle$. By defining $\delta x = x - \langle x \rangle$, we can write the variance of x compactly as $\langle \delta x^2 \rangle$. All probabilities, expectations, etc. used herein are conditioned on the stimulus θ, which we sometimes omit to simplify the notation.

Model setup

We model a population of direction-selective neurons with identical receptive field locations and a diverse range of preferred directions ϕ_i. We use a simple model of spatial and feature attention, where λ_i, the firing rate of neuron i, is the product of a gain $g_i(\psi)$ and a tuning function $f_i(\theta)$:

$$\lambda_i(\theta, \psi) = g_i(\psi) f_i(\theta) \tag{2}$$

Here, ψ is the attended direction of motion and θ the direction of the stimulus that is shown. We assume bell-shaped tuning curves of the form

$$f_i(\theta) = \exp(\kappa \cos(\theta - \phi_i) + \gamma_i), \tag{3}$$

where κ controls the tuning width, ϕ_i is the preferred direction of neuron i, and γ_i controls its mean firing rate. Although this choice of tuning curve simplifies the mathematical treatment considerably, the results do not depend on it qualitatively. Indeed, all results on fluctuations of the attentional gain hold for arbitrary tuning curves.

Neurons are assumed to produce spikes independently according to a Poisson process with rate λ_i. Thus, the only source for noise correlations in our model is the fluctuating attentional state, which comodulates the firing rates through the gain g_i.

The gain depends on whether attention is directed to the neurons' receptive field and on the attended direction of motion. For spatial attention, we use $g = \exp(\alpha)$, which is the same for all neurons because they all have identical receptive field locations; we refer to α as the spatial gain (see Fig. 1). For feature attention, we use $g_i(\psi) = \exp(\beta h_i(\psi))$, where β is the feature gain and $h_i(\psi)$ the gain profile (see Fig. 3). We follow the feature similarity gain model (Treue and Martínez Trujillo, 1999), where a neuron's gain is enhanced if the attended feature matches the neuron's preference and suppressed otherwise. We use a cosine gain profile: $h_i(\psi) = \cos(\psi - \phi_i)$.

From the perspective of the model, there is no fundamental difference between spatial and feature attention. However, because we consider a local population with identical receptive field locations, spatial attention is a special case with a constant gain profile $h_i = 1$ and, consequently, a single common gain $g = \exp(\alpha)$. Thus, whenever we refer to spatial attention, our results apply to a situation where all neurons in the population under consideration share the same preferred feature (i.e., receptive field location in our case). Likewise, when we refer to feature attention, our results apply to any situation where the neurons in the population span a large range of preferred features (i.e., preferred direction in our case). We chose this somewhat arbitrary distinction because it reflects the typical situation encountered in experiments in areas such as V1 or MT, where neurons with similar retinotopic locations are recorded, which typically span a large range of preferred orientations or directions of motion.

Effect of fluctuating gains on spike count statistics

To study the effect of a fluctuating attentional gain on the spike count statistics, we treat the more general case of feature attention (see Fig. 4); the results for spatial attention (see Fig. 2) follow as a special case with $\beta = \alpha$ and $h_i = 1$. We assume that the attended feature is fixed and that the experimenter does not have access to the attentional gain β on individual trials but can control only its average $\langle \beta \rangle$ over many trials (e.g., by cuing the subject). We denote the variance of the trial-to-trial fluctuations of β by $\langle \delta\beta^2 \rangle$.

To obtain the mean, variance, and covariance of the spike counts y_i, we need mean, variance, and covariance of the gain g. However, due to the exponential nonlinearity in g, the exact values depend on the distribution of β. We therefore simply assume that mean and variance of β are sufficiently small that we can linearize λ_i around $\langle \beta \rangle$:

$$\lambda_i \approx (1 + \delta\beta h_i(\psi)) \exp(\langle \beta \rangle h_i(\psi)) f_i(\theta) \tag{4}$$

We note that this approximation is not strictly necessary. One could in principle obtain exact analytical results by, for example, assuming β to be Gaussian. However, because attentional modulations are usually relatively small ($\beta \approx 0.1$), an exact treatment would add only complicated correction terms that would blur the key results without making any practical difference. We therefore favored the approximate framework due to the simplicity of its results.

We obtain for the average spike count:

$$\mu_i \equiv \langle y_i \rangle \approx \exp(\langle \beta \rangle h_i(\psi))\, f_i(\theta). \tag{5}$$

Variances and covariances are obtained by application of the Law of Total Covariance:

$$\mathrm{Cov}[y_i, y_j] = \langle \mathrm{Cov}[y_i, y_j \mid \beta] \rangle + \mathrm{Cov}[\langle y_i \mid \beta \rangle, \langle y_j \mid \beta \rangle] \tag{6}$$

$$= \delta_{ij}\mu_i + \langle \delta\beta^2 \rangle h_i h_j \mu_i \mu_j \tag{7}$$

where the outer expectation (covariance) is taken over β and the inner covariance (expectation) over y_i and y_j, we plugged in the definitions of $\lambda_i = \langle y_i \mid \beta \rangle$, and used the assumption of conditionally independent Poisson spiking, $\mathrm{Cov}[y_i, y_j \mid \beta] = \delta_{ij}\lambda_i$.

By taking the ratio of the variance divided by the mean, we obtain the Fano factor as follows:

$$F = 1 + \langle \delta\beta^2 \rangle h_i^2 \mu_i. \tag{8}$$

Fluctuations in attended feature induce differential correlations

Calculating the means and covariances under fluctuations in the attended direction ψ follows the same approach as above. We start with the case where the variance $\langle \delta\psi^2 \rangle$ is small (see Fig. 5). Assuming that the subject attends to the direction that is shown (i.e., $\langle \psi \rangle = 0$), we linearize λ_i around $\psi = \theta$:

$$\lambda_i \approx (1 + \delta\psi\beta h_i'(\theta))\exp(\beta h_i(\theta))\, f_i(\theta) \tag{9}$$

where h_i' is the derivative with respect to ψ. Using this approximation, we obtain for the average spike count:

$$\mu_i = \exp(\beta h_i(\theta))\, f_i(\theta). \tag{10}$$

Again, by applying the Law of Total Covariance, we obtain the spike count covariance:

$$\mathrm{Cov}[y_i, y_j] = \delta_{ij}\mu_i + \langle \delta\psi^2 \rangle \frac{\beta^2}{\kappa^2}\, \mu_i' \mu_i', \tag{11}$$

where $\mu_i' = d\mu_i(\theta)/d\theta$ and we used the definitions $h_i(\theta) = \cos(\theta - \phi_i)$ and $f_i(\theta) = \exp(\kappa\cos(\theta - \phi_i) + \gamma_i)$ to make the substitution $h_i'\mu_i = \mu_i'/\kappa$. Thus, fluctuations of the attended direction create differential correlations (Moreno-Bote et al., 2014), that is, response variability that is identical to variability induced by changes in the stimulus (sometimes also referred to as input noise).

Next, we treat the case where the attended direction fluctuates between two discrete alternatives ψ_1 and ψ_2, as would be expected for a two-alternative forced-choice discrimination task (see Fig. 6). We define $\Delta\boldsymbol{\mu}$ as the difference between the expected spike counts for the two attention targets:

$$\Delta\mu_i = \frac{1}{2}(\langle y_i \mid \psi_1 \rangle - \langle y_i \mid \psi_2 \rangle) = [\exp(\beta h_i(\psi_1)) - \exp(\beta h_i(\psi_2))]f_0, \tag{12}$$

where we have assumed that there is no net motion in the stimulus and f_0 is the neurons' firing rate for this zero-coherence condition. Again, applying the Law of Total Covariance, we obtain the covariance:

$$\mathrm{Cov}[y_i, y_j] = \delta_{ii}\mu_i + \Delta\mu_i \Delta\mu_j \tag{13}$$

Derivation of Fisher information for modulated Poisson distribution

The joint distribution of spike counts $P(\mathbf{y} \mid \theta)$ is a compound Poisson distribution, obtained by marginalizing over the latent gain β:

$$P(\mathbf{y} \mid \theta) = \int P(\beta) \prod_i P(y_i \mid \beta, \theta) d\beta \tag{14}$$

We assume that β is drawn from a normal distribution with mean $\langle \beta \rangle$ and variance $\langle \delta\beta^2 \rangle$. Approximating as above $\lambda_i \approx (1 + \delta\beta h_i)\mu_i$, we obtain:

$$P(\mathbf{y} \mid \theta) = \int \frac{1}{\sqrt{2\pi\langle \delta\beta^2 \rangle}} \exp\left(-\frac{\delta\beta^2}{2\langle \delta\beta^2 \rangle} \right) \prod_i \frac{\lambda_i^{y_i}}{y_i!} \exp(-\lambda_i) d\delta\beta \tag{15}$$

$$\approx \frac{1}{2\pi\langle \delta\beta^2 \rangle \prod y_i!} \int \exp\left(-\frac{\delta\beta^2}{2\langle \delta\beta^2 \rangle} + \sum_i (\log\mu_i + \delta\beta h_i) y_i - (1 + \delta\beta h_i)\mu_i \right) d\delta\beta \tag{16}$$

We can solve the integral by collecting the terms related to $\delta\beta$ and completing the squares:

$$P(\mathbf{y} \mid \theta) \propto \frac{1}{\prod y_i!} \exp\left(\frac{\langle \delta\beta^2 \rangle}{2} \left(\sum h_i y_i \right)^2 \right) \times \exp\left(\sum_i y_i \left(\log\mu_i - \langle \delta\beta^2 \rangle h_i \sum_j h_j \mu_j \right) \right) \tag{17}$$

$$\equiv \phi(\mathbf{y}) \exp(\boldsymbol{\eta}(\theta)^{\mathrm{T}}\mathbf{y}). \tag{18}$$

Thus, $P(\mathbf{y} \mid \theta)$ is in the exponential family with sufficient statistics $T(\mathbf{y}) = \mathbf{y}$. Therefore, the Fisher information with respect to the stimulus θ is given by (Beck et al., 2011)

$$J = \boldsymbol{\mu}'^{\mathrm{T}} C^{-1} \boldsymbol{\mu}'. \tag{19}$$

This expression is sometimes also referred to as the linear Fisher information or J_{mean} because of its close relationship to both the variance of a locally optimal linear estimator and the linear discriminability of two nearby stimuli ($J \propto d'^2$).

Despite the fact that the covariance matrix C depends on the stimulus, J is the full Fisher information of the population. This is unlike the Gaussian case, where a stimulus-dependent covariance matrix introduces a second term into the Fisher information (Kay, 1993). This term, sometimes referred to as J_{cov}, is absent in the modulated Poisson distribution, which means that (1) fine discrimination can be performed optimally using linear methods and (2) the linear Fisher information defines the Cramér-Rao bound (i.e., the minimum variance of any unbiased estimator).

Coding accuracy under fluctuations of attentional gain

Here we show that fluctuations of attentional gains do not impair the coding accuracy of a population of neurons. We start by considering a population of conditionally independent neurons. The first ingredient to calculating the Fisher information is the inverse of the covariance matrix, which we obtain by applying the Sherman-Morrison formula to Equation 7:

$$C^{-1} = M^{-1} - \frac{M^{-1}\mathbf{u}\mathbf{u}^{\mathrm{T}}M^{-1}}{\langle \delta\beta^2 \rangle^{-1} + \mathbf{u}^{\mathrm{T}}M^{-1}\mathbf{u}}, \tag{20}$$

where $M = \mathrm{Diag}(\boldsymbol{\mu}(\theta))$ and $u_i = h_i(\psi)\mu_i(\theta)$. Plugging into the formula for Fisher information, we obtain:

$$J = J_{\text{ind}} - \frac{\left(\sum h_i \mu_i'\right)^2}{\langle \delta\beta^2\rangle^{-1} + \sum h_i^2 \mu_i} \tag{21}$$

$$= J_{\text{ind}} - O(1). \tag{22}$$

The first term J_{ind} in the above equation is the Fisher information of an independent population of neurons:

$$J_{\text{ind}} = \sum_i \frac{(\mu_i')^2}{\mu_i}. \tag{23}$$

It is therefore $O(N)$, whereas the second term is zero for homogeneous populations of neurons, where $f_i(\theta) = f(\theta - \phi_i)$, and nonzero but $O(1)$ for heterogeneous populations. To show that the second term above is $O(1)$, we assume that the amplitudes of the neurons' tuning curves are independent random variables (Shamir and Sompolinsky, 2006; Ecker et al., 2011). In this case, the quantity of interest is the expected value with respect to different realizations of the heterogeneity:

$$\left\langle \frac{\left(\sum h_i \mu_i'\right)^2}{\langle \delta\beta^2\rangle^{-1} + \sum h_i^2 \mu_i} \right\rangle \approx \frac{\left\langle \left(\sum h_i \mu_i'\right)^2 \right\rangle}{\langle \delta\beta^2\rangle^{-1} + \sum h_i^2 \langle \mu_i\rangle} = O(1). \tag{24}$$

Here the approximation holds because for large N the denominator is $O(N)$ and its SD becomes narrower relative to its mean. Therefore, the expected value of the ratio converges to the ratio of the expected values of numerator and denominator. For the numerator, we simplify:

$$\left\langle \left(\sum h_i \mu_i'\right)^2 \right\rangle = \text{Var}\left[\sum h_i \mu_i'\right] = \sum h_i^2 \text{Var}[\mu_i'] = O(N), \tag{25}$$

which holds because $\langle \sum h_i \mu_i' \rangle = 0$, both for spatial attention (where $h_i = 1$) and feature attention when the correct feature is attended (where $\mathbf{h}$ is even and $\langle \boldsymbol{\mu} \rangle$ is odd). Thus, fluctuations in attentional gains do not impair the coding accuracy of the population with respect to direction of motion, as they result in only an $O(1)$ reduction of the Fisher information, which becomes irrelevant for large populations.

To study the physiologically more realistic situation where the amount of information entering the brain is finite, we model input noise by treating the stimulus direction θ itself as a random variable with variance $\langle \delta\theta^2 \rangle$. In this case, the covariance of the spike counts is given by

$$C = C_0 + \langle \delta\theta^2 \rangle \boldsymbol{\mu}' \boldsymbol{\mu}'^{\mathrm{T}}, \tag{26}$$

where C_0 is the covariance matrix in the absence of input noise. The Fisher information is then (Moreno-Bote et al., 2014)

$$J = \frac{J_0}{1 + \langle \delta\theta^2 \rangle J_0}, \tag{27}$$

where J_0 is the Fisher information in the absence of input noise. Because J_0 is $O(N)$, $J \to 1/\langle \delta\theta^2 \rangle$. Thus, in the presence of input noise, the $O(1)$ correction term from above vanishes for large N and the Fisher information converges to the limit imposed by the input noise.

Coding accuracy under fluctuations of attended feature

We have shown above (Eq. 11) that fluctuations of the attended feature have the same effect as input noise. Including input noise with variance $\langle \delta\theta^2 \rangle$ as in the previous section, the covariance of the spike counts is therefore given by

$$C = M + \left(\langle \delta\theta^2 \rangle + \frac{\beta^2}{\kappa^2} \langle \delta\psi^2 \rangle \right) \boldsymbol{\mu}' \boldsymbol{\mu}'^{\mathrm{T}}, \tag{28}$$

where $M = \text{Diag}(\boldsymbol{\mu})$ as before. Analogous to above, the Fisher information is

$$J = \frac{J_{\text{ind}}}{1 + \left(\langle \delta\theta^2 \rangle + \frac{\beta^2}{\kappa^2} \langle \delta\psi^2 \rangle \right) J_{\text{ind}}} \tag{29}$$

and for large N, it converges to

$$J \to \frac{1}{\langle \delta\theta^2 \rangle + \frac{\beta^2}{\kappa^2} \langle \delta\psi^2 \rangle}. \tag{30}$$

Simulation of experimental results

To simulate the results of Cohen and Maunsell (2011) (see Fig. 8), we used populations of N^2 neurons, tuned to both orientation and spatial frequency, with the preferred stimuli spaced regularly on an $N \times N$ grid. For the illustration of the correlation matrices, we used $N = 50$ and homogeneous tuning curves as in all other figures showing covariance or correlation matrices. For computing the relationship between firing rate changes and correlation changes, we used $N = 64$ and heterogeneous tuning curves. For convenience and symmetry, we modeled both variables as periodic (although not strictly correct, this simplification does not affect the results qualitatively). We defined the overall attentional gain $\tilde{g}_i = \alpha_1 h_i(\psi_1) + \alpha_2 h_i(\psi_2)$. The neuronal gain $g_i = \exp(\tilde{g}_i)$. Assuming attention gains α_k and features ψ_k are independent random variables, mean and covariance of $\tilde{g}$ are given by:

$$\langle \tilde{g}_i \rangle = \langle \alpha_1 \rangle \langle h_i(\psi_1) \rangle + \langle \alpha_2 \rangle \langle h_i(\psi_2) \rangle \tag{31}$$

$$\text{Cov}[\tilde{g}_i, \tilde{g}_j] = \sum_{k=1,2} (\langle \delta\alpha_k^2 \rangle + \langle \alpha_k \rangle^2) \langle h_i(\psi_k) h_j(\psi_k) \rangle - \langle \alpha_k \rangle^2 \langle h_i(\psi_k) h_j(\psi_k) \rangle, \tag{32}$$

where we estimated the expectations over ψ_k via numerical integration. Without loss of generality, we assumed the first feature to be attended and set $\langle \alpha_1 \rangle = 0.1$, $\langle \delta\alpha_1^2 \rangle = 0.05$ and ψ Gaussian with $\langle \delta\psi_1^2 \rangle = (10°)^2$. For model I (see Fig. 8*C*,*D*), we used $\langle \alpha_2 \rangle = 0$, $\langle \delta\alpha_2^2 \rangle = 0.1$ and ψ_2, as for the attended feature, Gaussian with $\langle \delta\psi_2^2 \rangle = (10°)^2$. For model II (see Fig. 8*E*,*F*), we used $\langle \alpha_2 \rangle = 0.1$, $\langle \delta\alpha_2^2 \rangle = 0.05$ as for the attended feature but chose ψ_2 as uniformly distributed. To obtain the spike count covariances, we linearized λ_i around $\langle g_i \rangle$ as above and obtained

$$\boldsymbol{\mu}_i = \exp(\langle \tilde{g}_i \rangle) f_i(\theta) \tag{33}$$

$$\text{Cov}[y_i, y_j] = \delta_{ij} \mu_i + \text{Cov}[\tilde{g}_i, g_j] \mu_i \mu_j. \tag{34}$$

For the illustrations of the correlation matrices, we normalized the covariance matrices to correlation coefficients and marginalized over the irrelevant feature (i.e., for the attended condition [orientation], we averaged over all neurons with the same preferred orientation but different preferred spatial frequencies).

To simulate the results of Ruff and Cohen (2014) (see Fig. 9), we used a population of N neurons with receptive field locations arranged linearly in the range [0, 1]. We assumed that receptive fields have a Gaussian shape with SD 0.5 and a peak firing rate of 20 spikes/s. We placed the two stimuli at 0.35 and 0.65. For simplicity, we assumed that the firing rate in response to the two stimuli presented simultaneously is equal to the average firing rate elicited by the two stimuli presented individually. Because the receptive field locations vary within the population, we treat spatial location as a feature. For the gain profile, we used a Mexican hat:

$$h_i(\psi) = 2\exp(-2(\psi - \phi_i)^2) - \exp(-(\psi - \phi_i)^2/2), \tag{35}$$

which corresponds to an excitatory center with an SD of 0.5 and a suppressive surround with an SD of 1. We set the distribution of the attentional gain to $\langle \beta \rangle = 0.2$ (reasoning that attentional effects are typically stronger when multiple stimuli compete in the receptive field; Treue and Maunsell, 1996) and $\langle \delta\beta^2 \rangle = 0.05$ for both attended and unattended conditions. The difference between the two conditions is only the distribution of attended locations ψ. We assumed the attentional focus to be centered on 0.5 on average, but with a smaller SD for the attended con-

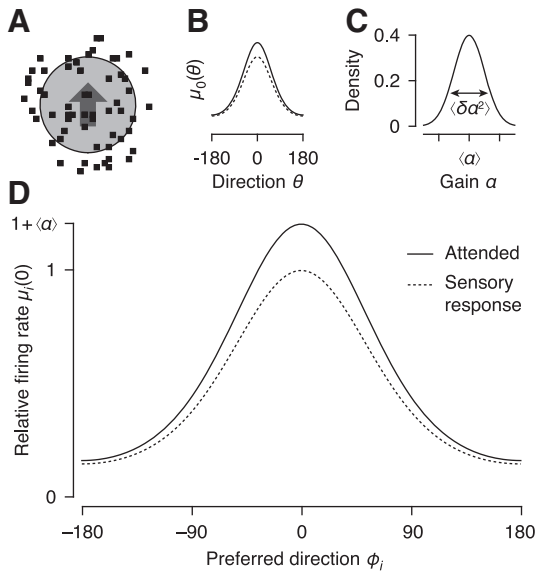

Figure 1. Model of spatial attention. ***A***, Example stimulus. Neurons' receptive fields are assumed to be at the same location (circle). ***B***, Tuning curve under sensory stimulation (dashed line) and with spatial attention directed to the stimulus in the receptive field (solid line). ***C***, Distribution of attentional gain (α). ***D***, Population response of a homogeneous population of neurons under sensory stimulation (dashed line) and with attention directed to the stimulus in the receptive fields (solid line).

dition (0.2) than for the unattended condition (0.5). As before, we obtained the covariance of the gain by numerical integration over the distribution of ψ and the covariance matrix by application of the Law of Total Covariance (see Eq. 34). We used a homogeneous population of 50 neurons for all illustrations and a heterogeneous population of 512 neurons for simulating the relationship between task tuning similarity (TTS) and correlations. TTS was computed as in Ruff and Cohen (2014) as the d' between the responses to the two individual stimuli, assuming Poisson statistics (i.e., variance equal to the average spike count).

Generalized linear population model

Our population model can be recast as a GLM by a simple reparameterization of the stimulus. Consider the log firing rate

$$\log \lambda_i = \alpha + \beta \cos(\psi - \phi_i) + \kappa \cos(\theta - \phi_i) + \gamma_i, \quad (36)$$

where α and β are the spatial and feature gains as before. By representing angles as two-dimensional vectors of unit length (e.g., $\mathbf{x} = [\cos\theta, \sin\theta]^{\mathrm{T}}$), we can rewrite the log firing rate as a linear function of the attentional state and the stimulus:

$$\log \lambda_i = \alpha + \mathbf{k}_i^{\mathrm{T}}\mathbf{b} + \mathbf{k}_i^{\mathrm{T}}\mathbf{x} + \gamma_i. \quad (37)$$

Here, α and $\mathbf{b} = \beta/\kappa \cdot [\cos\psi, \sin\psi]^{\mathrm{T}}$ represent the state of spatial and feature attention, respectively, $\mathbf{x}$ is the stimulus, and $\mathbf{k}_i = \kappa \cdot [\cos\phi_i, \sin\phi_i]^{\mathrm{T}}$ is the neuron's preferred direction. This model is a GLM with Poisson observations and $\log(x)$ as the link function.

Results

Fluctuations in spatial attention

Our goal is to characterize the effect of fluctuating attentional signals on the population response in sensory areas. To simplify the exposition of the basic concepts and results, we start with the simplest possible case: that of spatial attention in a population of neurons with identical receptive field locations (Fig. 1*A*). We assume that neurons encode the direction of motion of a stimulus through bell-shaped tuning curves $f_i(\theta)$ (Fig. 1*B*, dotted line) and that their firing rates are modulated by a common gain e^{α}:

$$\lambda_i = e^{\alpha} f_i(\theta). \quad (38)$$

We assume that α fluctuates from trial to trial and is drawn from a normal distribution with mean $\langle\alpha\rangle$ and variance $\langle\delta\alpha^2\rangle$ (Fig. 1*C*). Using this parameterization $\langle\alpha\rangle = 0$ corresponds to no attentional allocation and a neuronal gain of 1; we refer to this as the sensory response (Fig. 1*D*, dotted line). In contrast, when the stimulus is attended, $\langle\alpha\rangle > 0$ (Fig. 1*D*, solid line). Under this model, the average spike count of a neuron is approximately

$$\langle y_i \rangle \equiv \mu_i \approx (1 + \langle\alpha\rangle) f_i(\theta). \quad (39)$$

Although we use homogeneous neural populations in the figures (all neurons have the same tuning curve up to a preferred direction ϕ_i, i.e., $f_i(\theta) = f(\theta - \phi_i)$), all results in this section hold more generally for arbitrary tuning curves.

Because the attentional state fluctuates from trial to trial, the underlying firing rate also fluctuates. These fluctuations are represented in the model by the variance of the gain term. By applying the Law of Total Covariance, we obtain the spike count variance (Fig. 2*A*):

$$\langle \delta y_i^2 \rangle \approx \mu_i + \langle\delta\alpha^2\rangle \mu_i^2, \quad (40)$$

where $\langle\delta\alpha^2\rangle$ is the variance of the attentional gain. The first term is equal to the average spike count and results from the Poisson process assumption, whereas the second term is quadratic in the firing rate, which results from the multiplicative nature of the fluctuating gain α (Goris et al., 2014), causing the spike count variance to grow more quickly than the mean. Such an expanding mean variance relation has been observed in many experimental studies (Dean, 1981; Tolhurst et al., 1983; Britten et al., 1993; Goris et al., 2014). If the attentional gain does not fluctuate, we recover the Poisson process.

As the neurons' firing rates are comodulated by a common gain, the gain fluctuations induce correlations between the neurons. We find that the resulting covariance matrix C has a simple form. It can be expressed as the sum of a diagonal matrix M and a rank-one matrix (Fig. 2*C*):

$$C = M + \langle\delta\alpha^2\rangle \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}, \quad (41)$$

where $M = \mathrm{Diag}(\boldsymbol{\mu})$ contains the independent variances resulting from the Poisson noise and the second term results from the gain fluctuations. (The assumption of conditional independence could be relaxed without affecting any of the major results qualitatively: the diagonal matrix in the equation above would simply be replaced by an alternative, nondiagonal covariance matrix.)

Hence, the covariance between two neurons is proportional to the product of the firing rates, with the constant of proportionality given by the variance of the attentional gain (Fig. 2*B*). There is no such simple expression for the correlation coefficient, which is more typically quantified in experimental studies. We find that spike count correlations induced by a fluctuating attentional gain increase with firing rates (Fig. 2*D*), as observed in numerous experimental studies (Cohen and Maunsell, 2009; Mitchell et al., 2009; Smith and Sommer, 2013; Ecker et al., 2014). This effect arises because the independent (Poisson) variability is linear in the firing rate, whereas the covariance induced by gain fluctuations is quadratic and therefore dominates for large firing rates. However, although correlations increase with the geometric mean firing rate, there is no simple one-to-one mapping between the two quantities: it also depends on the ratio of the firing rates (Fig. 2*C*). Thus, our analysis suggests that in the presence of gain

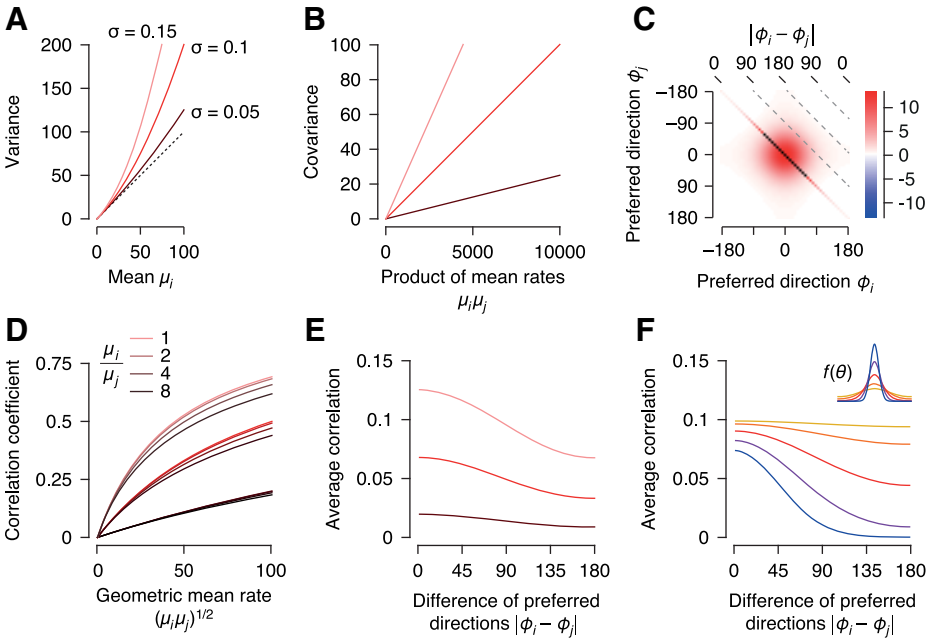

Figure 2. Effect of fluctuations in attentional state on spike count statistics. Solid lines indicate analytical solutions (Eqs. 4–8). Parameter values used here were $\langle\alpha\rangle = 0.1$, $\langle\delta\alpha^2\rangle \in \{0.05^2, 0.1^2, 0.15^2\}$ (dark to light red). ***A***, Spike count variance as a function of mean spike count. Dashed line indicates identity (Poisson process). ***B***, Covariance as a function of product of spike counts. Colors as in ***A***. ***C***, Covariance matrix for $\theta = 0°$ and $\langle\delta\alpha^2\rangle = 0.1^2$. Neurons are ordered by preferred directions. Tuning curves: $f_i(\theta) = \exp(\kappa\cos(\theta - \phi_i) + \gamma)$, $\kappa = 2$, γ chosen so that the average firing rate across all θ is 10 spikes/s. ***D***, Correlation coefficient as function of geometric mean firing rate. The three groups of lines correspond to different levels of $\langle\delta\alpha^2\rangle$ as in ***A***, ***B***. Darker colors within a group indicate increasing ratios μ_i/μ_j. ***E***, Average correlation coefficient (over all directions of motion θ) as a function of difference of the preferred directions of the two neurons. Despite a common gain for all neurons, correlations decay with tuning difference. Tuning parameters as in ***D***. ***F***, Same as in ***E***, but for different tuning widths ($\kappa \in \{0.5, 1, 2, 4, 8\}$, inset, top). The decay of the correlations with the difference of the preferred directions is stronger for narrow tuning curves. Red line corresponds to ***D***, ***E***. Mean firing rate: 10 spikes/s for all tuning widths.

fluctuations covariances are more appropriate to consider when analyzing experimental data than correlation coefficients. Alternatively, it would be appropriate to normalize the covariance by the product of the firing rates $\mu_i\mu_j$ if some kind of normalization is desired.

In addition, the correlation structure induced by gain fluctuations is nontrivial even if all neurons share the same gain (Fig. 2*E*,*F*) (see also Ecker et al., 2014). Because of the nonlinear shape of the tuning function and the nonlinear way the neurons' tuning functions affect spike count correlations, the correlations decrease with increased difference in two neurons' preferred directions (Fig. 2*E*). The slope of the decay depends mainly on the dynamic range of the tuning curve (Fig. 2*F*). If neurons have a high baseline firing rate compared with their peak firing rate, correlations decrease only marginally with preferred direction. In contrast, sharply tuned neurons with close to zero baseline firing rates exhibit strong limited-range structure. This limited-range correlation structure has been observed in numerous experimental studies (Zohary et al., 1994; Bair et al., 2001; Smith and Kohn, 2008; Cohen and Maunsell, 2009; Ecker et al., 2010) and has been hypothesized to reflect shared input among similarly tuned neurons. However, our simple model shows that these seemingly structured correlations can arise from a very simple, nonspecific mechanism: a common fluctuating gain that drives all neurons equally, regardless of their tuning properties.

Fluctuations of feature attention

Feature attention is different from spatial attention in that the sign of the modulation depends on the similarity of the attended direction to the neuron's preferred direction of motion (Fig. 3). Following the feature-similarity gain model (Treue and Martínez Trujillo, 1999), we model feature attention by

$$\lambda_i = e^{\beta h_i(\psi)} f_i(\theta), \tag{42}$$

where we refer to β as the feature gain that controls how strongly the feature ψ (in this case, direction of motion) is attended on the given trial and $h_i(\psi)$ is the gain profile that determines the sign and relative strength of modulation for each neuron depending on the similarity of its preferred direction to the attended direction (Fig. 3*B*).

In this model, we can think of feature attention as a prior on the direction of motion. The attentional term $e^{\beta h_i}$ is a population hill centered on the attended direction of motion. The gain β controls its width and amplitude (the strength of the prior), whereas the profile $h_i(\psi)$ controls its location. Thus, feature attention biases the population response toward the attended direction by enhancing the response of neurons with preferred directions close to the attended direction and suppressing those with opposite preferred directions (Fig. 3*B*). As a result, unlike in the case of spatial attention, the shape of the population response is no longer identical to that of the individual neurons' tuning curve (Fig. 3*D*).

We start by assuming that the subject always attends the same direction (i.e., ψ is constant) and consider the effect of fluctuations in the strength of attention, that is the gain β. We will come back to fluctuations in the attended direction below.

Similar to spatial attention, fluctuations in feature attention lead to overdispersion of the spike counts relative to a Poisson process. This means that the ratio of variance to mean (the Fano factor) is >1. The degree of overdispersion not only increases

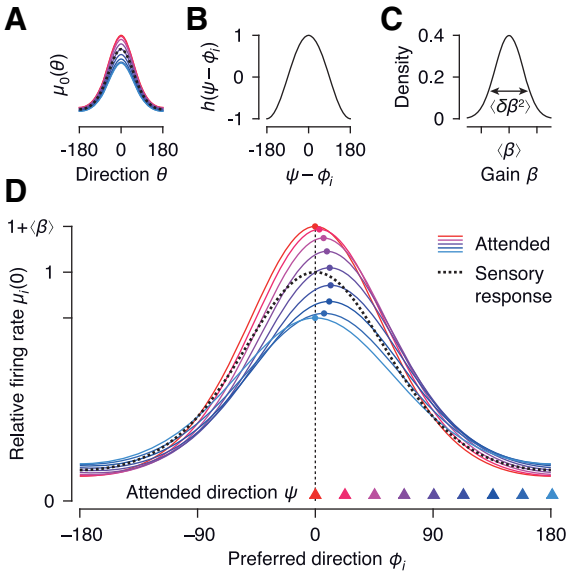

Figure 3. The feature similarity gain model for feature attention. ***A***, Tuning curve of a single neuron under sensory stimulation (black dotted) and with feature attention directed to different directions ranging from preferred (red) to null (blue). The entire tuning curve of the neuron is gain-modulated, and the modulation does not depend on the stimulus θ. ***B***, The gain of a neuron depends on which direction of motion ψ is attended relative to the neuron's preferred direction ϕ_i. ***C***, Distribution of gain (β) fluctuations with mean $\langle\beta\rangle$ and variance $\langle\delta\beta^2\rangle$. ***D***, Population response of a homogeneous population of neurons under sensory stimulation (black dotted) and with attention directed to different directions of motion ranging from 0° (red) to 180° (blue). The stimulus is $\theta = 0$. Curves represent the average response of the neurons as a function of their preferred direction. Attending to a direction of motion biases the population response toward this attended stimulus. Although each neuron's tuning curve is gain-modulated as a whole (***A***), the population response is no longer equal to the individual neurons' tuning curves but instead sharpened/broadened and its peak is moved.

with the neuron's firing rate but also depends on the neuron's preferred direction relative to the attended direction (Fig. 4*A*). Spike counts are most overdispersed at the preferred and the null directions (Fig. 4*A*, red and blue). Moreover, the feature similarity gain model predicts that neurons with preferred directions close to orthogonal to the attended direction should be the least overdispersed (Fig. 4*A*, purple).

As feature attention induces both increases as well as decreases in neuronal gain, the induced correlation structure is different from that induced by spatial attention. However, because each neuron's gain is driven by a common feature gain, the covariance matrix can again be decomposed into a diagonal matrix M plus a rank-one component:

$$C = M + \langle\delta\beta^2\rangle\mathbf{u}\mathbf{u}^{\mathrm{T}}, \tag{43}$$

where $u_i = h_i(\psi)\mu_i(\theta)$. The sign of the covariance is determined by the product of h_i and h_j, which depends on the attended direction and the preferred directions of the two neurons (Fig. 4*B*). The covariance is always positive for two neurons with identical preferred directions, whereas it is always negative for two neurons with orthogonal preferred directions. For any pair of neurons in between, it can be both positive and negative depending on the stimulus (Fig. 4*B*).

As for spatial attention, averaging correlations over multiple stimulus conditions to represent the correlation structure as a function of the neurons' tuning similarity misses much of the underlying structure (Fig. 4*C*): spike count correlations are positively correlated with tuning similarity (Fig. 4*D*), but the stimulus dependence (Fig. 4*C*) is again ignored. As before, the exact shape of the decay depends on the tuning width: for narrow tuning curves, neurons with opposite preferred directions are only weakly anticorrelated, whereas for broad tuning curves, those neurons are strongly anticorrelated (Fig. 4*D*, blue to yellow lines).

So far, we have assumed that the attended direction of motion is constant and only the strength of attention fluctuates from trial to trial. Now we turn to the case where the attended direction itself fluctuates from trial to trial. Assuming that the subject attends on average the correct direction $\langle\psi\rangle = \theta$, but with some variance $\langle\delta\psi^2\rangle$, we find that the covariance matrix can again be written as diagonal plus rank one (Fig. 5; for the derivation, see Materials and Methods):

$$C = M + \langle\delta\psi^2\rangle\frac{\beta^2}{\kappa^2}\boldsymbol{\mu}'\boldsymbol{\mu}'^{\mathrm{T}} \tag{44}$$

Here β is the strength of the attentional modulation, κ the tuning width of the neurons, and $\boldsymbol{\mu}'$ the derivative of the average firing rate with respect to the stimulus. This pattern of correlations is known as differential correlations (Moreno-Bote et al., 2014), as the variability resembles that induced by small changes in the stimulus. This result is indeed expected: as we mentioned above, feature attention can be thought of as a prior on direction of motion, therefore biasing the population response toward the attended direction. If the attended direction changes from trial to trial, this will perturb the population response in the same direction in multineuronal space as small changes in the stimulus itself.

Interestingly, when plotted as a function of tuning similarity, the correlation structure resembles that induced by gain fluctuations (Fig. 5*C*), except for very narrow tuning curves. This finding is quite striking because the correlation matrices look quite different (compare Fig. 4*C* with Fig. 5*B*) and, as we will see below, the two correlation structures have quite dramatically different effects on the population code. However, the correlations induced by fluctuations in the attended direction are substantially weaker than those induced by gain fluctuations (Figs. 2, 4), even if the distribution of attended directions is fairly wide (SD: 10° in Fig. 5).

A second example with relevance to experimental studies is the situation where feature attention fluctuates between two discrete alternatives (Fig. 6). Consider the classic random dot motion discrimination paradigm (Newsome and Paré, 1988), where the subject has to decide whether the net motion in the display is rightward or leftward (Fig. 6*A*). Most interesting to the experimenter are the trials where the net motion is zero (zero coherence). On average, the population response is flat on those trials (Fig. 6*B*, dashed line), as there is no net motion signal in the stimulus. However, on any given trial, the subjects may have expectations about the stimulus that is to come, for example, because of the past stimuli they have observed. They may therefore decide to attend to one direction of motion or the other. As a consequence, in trials where the subject attends to leftward motion, neurons with preferred directions around leftward (rightward) motion will be enhanced (suppressed) and vice versa (Fig. 6*B*, red and blue). That is, the population response will fluctuate between attend-left and attend-right. The covariance structure induced by such fluctuations (Fig. 6*C*) is very similar to that observed before, except that we have to replace the derivative of the response $\boldsymbol{\mu}'$ by the difference between the responses when attending left versus right, $\Delta\boldsymbol{\mu}$:

$$C = M + \Delta\boldsymbol{\mu}\Delta\boldsymbol{\mu}^{\mathrm{T}}. \tag{45}$$

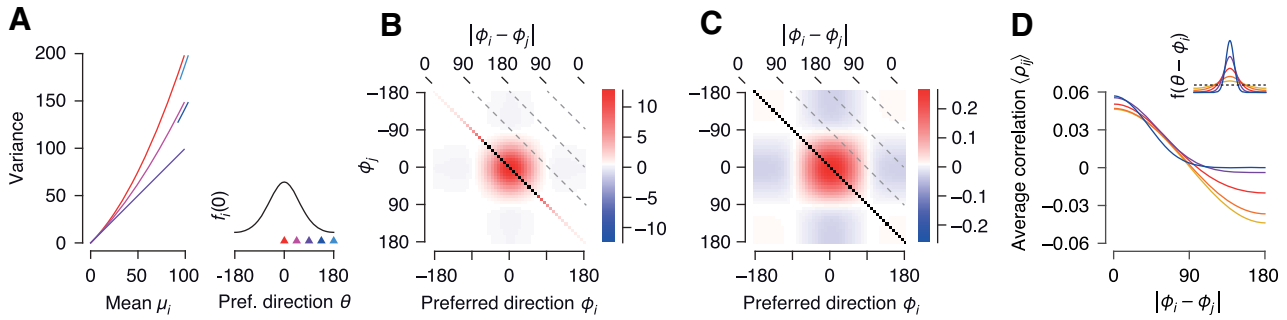

Figure 4. Effect of fluctuations in the feature attention gain on spike count statistics. Parameters here are as follows: $\psi = 0$, $\langle \beta \rangle = 0.1$, $\langle \delta\beta^2 \rangle = 0.1^2$. ***A***, Spike count variance as a function of mean spike count. Colors represent different attended directions relative to the neurons' preferred direction ($\psi - \phi_i$; bottom right, inset, colored triangles). Short blue line segments indicate that the mean-variance relationships in these conditions are identical to those indicated by the red lines right next to them. ***B***, Covariance matrix for stimulus $\theta = 0$. Neurons are ordered by preferred directions. Mean firing rate across the population: 20 spikes/s. ***C***, Same as in ***B***, but the correlation coefficient matrix is shown. ***D***, Dependence of spike count correlations on tuning similarity (difference of preferred directions). Fluctuations in feature attention induce limited range correlations regardless of the shape of the tuning curve. The higher the baseline firing rate, the stronger the negative correlations for neurons with opposite preferred directions. Inset, Different tuning widths used.

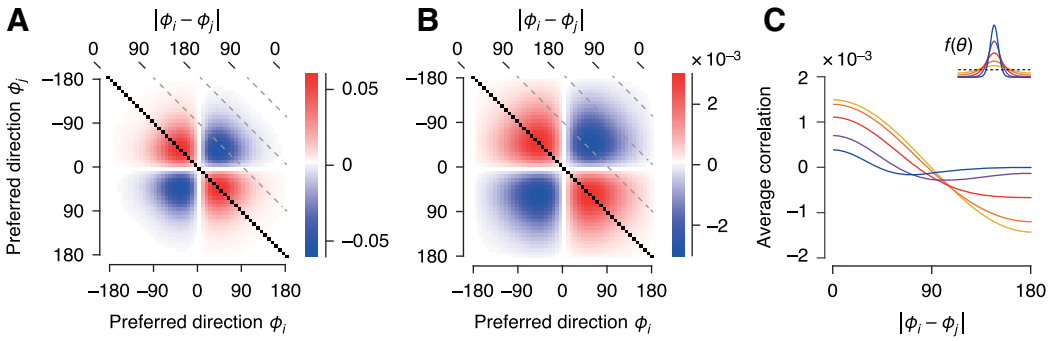

Figure 5. Effect of fluctuations in the attended direction on correlation structure. Parameters here are as follows: $\langle \psi \rangle = \theta = 0$, $\langle \delta\psi^2 \rangle = (10°)^2$, $\beta = 0.1$, mean firing rate across the population: 20 spikes/s. ***A***, Covariance matrix. Neurons are ordered by preferred directions. ***B***, Same as in ***A***, but the correlation coefficient matrix is shown. ***C***, Dependence of spike count correlations on tuning similarity (difference of preferred directions). Fluctuations in the attended direction induce limited range correlations, whose shape depends on the width of the tuning curves. Inset, Different tuning widths used.

An interesting difference to above is that, in this case, the correlation matrix (Fig. 6*D*) is more or less a scaled version of the covariance matrix because the population response is flat; therefore, the normalization does not affect its shape very much.

Effects of attention-induced correlations on population coding

How interneuronal correlations affect the representational accuracy of neuronal populations has been a matter of immense interest (and debate) over the last years, and changes in the correlation structure have been suggested to underlie the improved behavioral performance under attention (Cohen and Maunsell, 2009; Mitchell et al., 2009). Thus, we want to briefly consider how correlations induced by attentional fluctuations affect the coding accuracy of a population code.

Before doing so, we need to make a choice: does the downstream readout have access to the state of attention or not? If it does, the picture is fairly simple: attentional fluctuations do not affect the readout accuracy because the attentional state can be accounted for and there is no additional noise compared with a scenario without attentional fluctuations. The only downside is a potentially more complex readout. In contrast, if we assume that the readout does not have access to the attentional state, the situation becomes more interesting. In this case, the attentional fluctuations act like additional (internally generated) noise, which could impair the readout. In the following, we consider this latter scenario.

To quantify the accuracy of a population code, we use the Fisher information (Kay, 1993) with respect to direction of motion. Fisher information is proportional to the square of d' (Berens et al., 2011), which quantifies the detectability of small changes in a stimulus parameter from the resulting changes in the responses of a population of neurons. For a population of neurons with independent noise, the Fisher information of individual neurons adds up linearly.

We start by considering spatial attention. Because the gain is the same for all neurons, gain fluctuations should not affect the coding accuracy of the population with respect to the direction of the stimulus, which is encoded in the differential activation pattern of the neurons. This is indeed the case. As shown in Materials and Methods, the Fisher information of a population of Poisson neurons whose firing rates are comodulated by a common fluctuating gain is given by

$$J = J_{\text{ind}} - O(1) \approx J_{\text{ind}}, \tag{46}$$

where J_{ind} is the Fisher information of an independent population without fluctuating gain (i.e., $\langle \delta\alpha^2 \rangle = 0$). Thus, unobserved gain fluctuations reduce the information in the population only by a constant term, which is negligible for reasonably large populations (Fig. 7*A*, solid blue line vs circles). This result can be understood intuitively by considering the structure of the covariance matrix (Eq. 41): the dominant eigenvector points in the direction of the neural response μ, which is orthogonal to

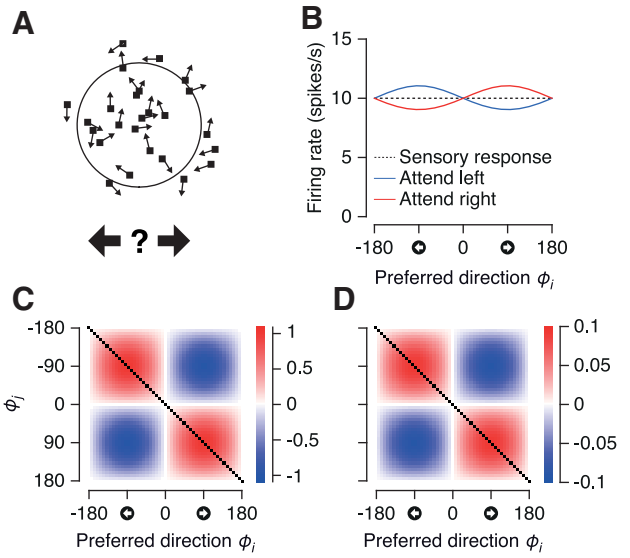

Figure 6. Fluctuations of feature attention in two-alternative forced-choice discrimination task. ***A***, Illustration of zero-coherence random dot stimulus. Arrows attached to dots indicate direction of motion of individual dots. Circle represents neurons' receptive fields. Big arrows and question mark indicate the subject's task: decide whether the net motion is leftward or rightward. ***B***, Population response to zero-coherence stimulus. Dashed line indicates sensory response. Red represents average response when the subject attends to rightward motion. Blue represents average response when the subject attends to leftward motion. ***C***, Covariance matrix if attentional state is unknown to experimenter (as is usually the case in such discrimination tasks). Neurons are ordered by their preferred directions. ***D***, Same as in ***C***, but for correlation coefficient.

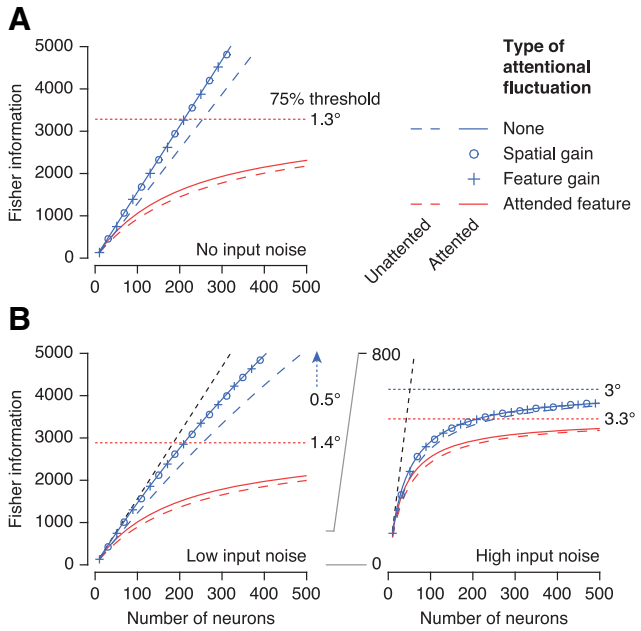

Figure 7. Coding accuracy under unknown fluctuations of attentional state. ***A***, Fisher information as a function of population size in the limit of no input noise. Fluctuations of the attentional gain (circles represent spatial gain; pluses indicate feature gain) do not impair coding accuracy relative to an uncorrelated population without attentional fluctuations (solid blue line). Fluctuations of the attended feature (solid red line) impose an upper bound on the Fisher information (dotted red line). Dashed lines indicate unattended condition with the same level of attentional fluctuations. ***B***, Same as in ***A***, but with input noise. The variance of the input noise was chosen such that it leads to asymptotic 75% thresholds of 0.5° (left) and 3° (right), respectively, in the absence of attentional fluctuations. Dotted lines indicate asymptotic limit of the Fisher information with (red) and without (blue) fluctuations in the attended feature. Black dashed line indicates performance of an uncorrelated population in the absence of input noise (same as in ***A***, blue line).

changes in the response due to changes in the stimulus, μ'. Therefore, gain fluctuations do not impair the readout of the direction of motion. This result concerns only the fluctuations in the gain. The increased gain due to attention still leads to a higher Fisher information in the attended condition compared with the unattended condition (Fig. 7*A*, solid vs dashed blue line).

The same result holds for fluctuations in the feature gain, so long as the attended direction matches the one shown and does not fluctuate from trial to trial (Fig. 7*A*, blue pluses). A fluctuating gain sharpens or broadens the population hill from trial to trial but leaves its peak unchanged. Again, the dominant eigenvector ($u_i = h_i\mu_i$, Eq. 43) points in a direction that is orthogonal to changes in the stimulus (for details, see Materials and Methods).

The situation changes if the focus of attention (i.e., the attended direction) fluctuates from trial to trial. Because feature attention biases the population response toward the attended direction of motion (Fig. 3*D*), such attentional fluctuations induce differential correlations: the dominant eigenvector is the derivative of the neural response with respect to the stimulus, μ'. Therefore, the Fisher information saturates at a finite value (Fig. 7*A*, red lines):

$$J \to \frac{\kappa^2}{\beta^2\langle\delta\psi^2\rangle}. \qquad (47)$$

Thus, for sufficiently large populations, the Fisher information is determined only by the degree of attentional modulation β and the variance of the attended direction $\langle\delta\psi^2\rangle$ relative to the tuning width κ (for derivation, see Materials and Methods). We call the term $\langle\delta\psi^2\rangle\beta^2/\kappa^2$ the effective stimulus variance, as it has exactly the same effect as unobserved variability in the stimulus of the same magnitude.

So far, we have assumed that there is no limit on the amount of information entering the brain. However, in practice, the amount of information is finite due to sensory noise (e.g., in the photoreceptors). In some experimental settings, this sensory noise is likely to be very small (e.g., orientation discrimination tasks with high-contrast gratings), whereas in other situations it can be substantial (e.g., random dot motion paradigms at low coherence). Therefore, we also briefly considered how input noise with variance $\langle\delta\theta^2\rangle$ affects the Fisher information in the presence of attentional fluctuations. The two main results from above also hold in this case: whereas gain fluctuations (both spatial and feature gain) have a negligible effect on the Fisher information (Fig. 7*B*, solid blue line, circles and pluses), fluctuations of the attended direction do impair the code (Fig. 7*B*, red lines).

A few observations are noteworthy. First, when the input noise is small (e.g., 75% threshold of an ideal observer: 0.5°), the results are qualitatively similar to the approximation of no input noise (Fig. 7*B*, left) for population sizes of a few hundred neurons. Second, if the population response is dominated by the amount of input noise (threshold: 3°), we can consider the asymptotic value at which the Fisher information saturates:

$$J \to \frac{\kappa^2}{\kappa^2\langle\delta\theta^2\rangle + \beta^2\langle\delta\psi^2\rangle}. \qquad (48)$$

This value depends only on the tuning width κ, the amount of input noise $\langle\delta\theta^2\rangle$, the degree of attentional modulation β, and the variance of the attended direction $\langle\delta\psi^2\rangle$. Thus, gain fluctuations

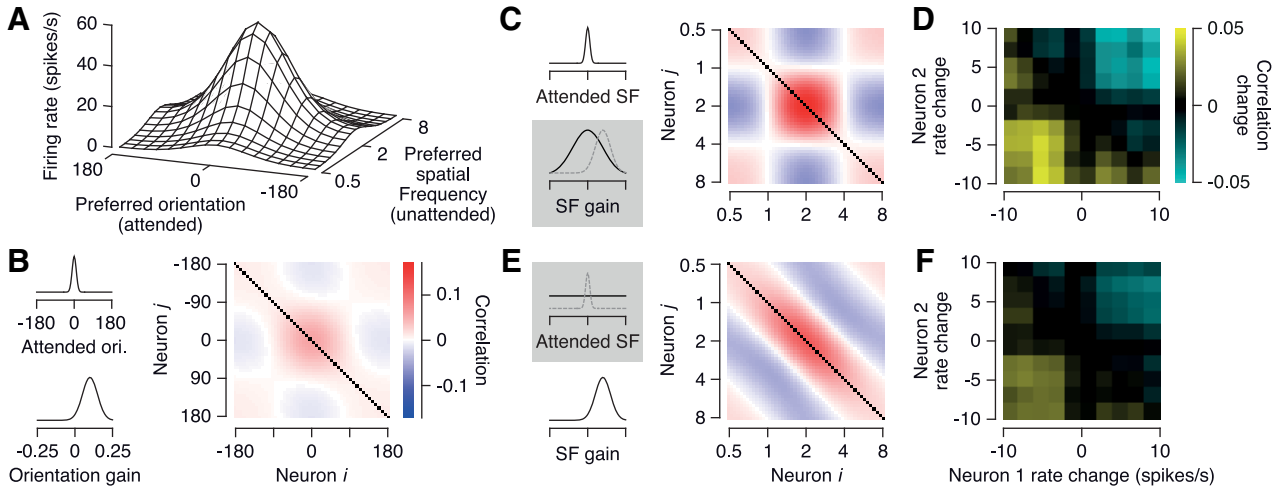

Figure 8. Two possible accounts of the data by Cohen and Maunsell (2011). ***A***, Model population tuned to two feature dimensions: here orientation and spatial frequency. Plots represent the average sensory response of the population to an orientation of 0° and a spatial frequency of 2 cycles/deg. Without loss of generality, we assume the subject attends to orientation (the model is entirely symmetric with respect to the two features). ***B***, Distribution of attended orientations (top left), orientation gain (bottom left), and correlation matrix (right). The correlation matrix is arranged by the neurons' preferred orientation and averaged over neurons with different preferred spatial frequencies. ***C***, Gain variability model for unattended feature. Analogous to ***B***, the distribution of attended spatial frequencies (top left) and spatial frequency gains (bottom left) are shown along with the correlation matrix (right) arranged by preferred spatial frequencies. Here we assume that not attending to spatial frequency reduces the gain and increases its variance (gray box; dashed gray line indicates the distribution of the gain under attention for comparison). ***D***, Changes in correlations as a function of the two neurons' changes in firing rates between the two attention conditions (attending to orientation vs attending to spatial frequency) as reported by Cohen and Maunsell (2011). ***E***, Feature variability model for unattended feature. Same as in ***C***, except that here we assume that not attending to spatial frequency means that a random spatial frequency is attended on each trial (gray box). ***F***, Same as in ***D***, but for the feature variability model.

(both spatial and feature gain) also have no effect in large populations when the Fisher information is bounded by the input noise (Fig. 7*B*, right); the $O(1)$ correction term from Equation 46 disappears. Moreover, in the absence of fluctuations of the attended direction, the Fisher information reduces to $1/\langle\delta\theta^2\rangle$, the inverse variance of the input noise, which is the bound given by the data processing inequality (Moreno-Bote et al., 2014). Third, fluctuations of the attended direction further reduce the Fisher information below the limit imposed by the data processing inequality (Fig. 7*B*, right, red lines).

In summary, attentional gain fluctuations generally do not impair coding accuracy, but fluctuations of the attended feature can have a major effect, in particular when the input noise is small. Importantly, this finding does not apply only to voluntary variability in the attended direction. Even if the animal tries to attend to the same direction on every trial, $\langle\delta\psi^2\rangle$ is non-zero in any realistic scenario because the attended direction ψ is represented by a finite number of neurons in the brain. Therefore, the very existence of an attentional mechanism places a limit on how accurately a stimulus can be represented, and this limit can be substantially lower than that imposed by the information in the feedforward signal (see also Discussion).

A new view on correlated variability under attention

There is ample experimental evidence that attention fluctuates from trial to trial (Cohen and Maunsell, 2010, 2011), and we showed in the previous sections that such fluctuations induce patterns of (correlated) variability that are highly consistent with the reported data on attention (Cohen and Maunsell, 2009; Mitchell et al., 2009; Herrero et al., 2013). Interestingly, in our model, both the magnitude of overdispersion in single neurons' spike counts and the average level of correlations depend on the variance of the attentional gain $\langle\delta\alpha^2\rangle$, which in the model is completely independent of the average modulation $\langle\alpha\rangle$. This observation suggests that the average attentional modulation between an attended and an unattended condition (which can be reliably measured based on average responses) may not predict the level of correlations in either condition because the latter is controlled by an independent variable.

This dissociation between attention effects on firing rates and correlations is indeed a central experimental finding for which our model can account. In many cases, directing spatial attention to a certain location increases the average responses of neurons whose receptive fields represent this location, but reduces their individual and shared variability (Cohen and Maunsell, 2009; Mitchell et al., 2009; Herrero et al., 2013). Thus, if our model is correct, then the data suggest that attention not only increases response gain but also reduces the trial-to-trial fluctuations of the gain.

Attentional fluctuations can also account for more recent experimental results on modulation of correlations under attention, two of which (Cohen and Maunsell, 2011; Ruff and Cohen, 2014) we reproduce with our model in the following.

In the first study, Cohen and Maunsell (2011) investigated how feature attention modulates firing rates and interneuronal correlations. In their paradigm, monkeys have to attend to either the orientation of a grating or its spatial frequency. Their findings resemble the patterns observed for spatial attention: correlations are reduced for neurons whose firing rates increase when attending to orientation compared with attending to spatial frequency, and vice versa (Fig. 8*D*).

To reproduce this pattern of results in our model, we consider a population of neurons that is tuned to both orientation and spatial frequency (Fig. 8*A*; see Materials and Methods). We assume, without loss of generality, that the monkey attends to orientation. Moreover, we assume that he attends to (approximately) the correct value (0°) and that the gain fluctuates moderately from trial to trial (Fig. 8*B*). The resulting correlation structure with respect to the neurons' preferred orientation resembles what we showed earlier (compare Fig. 4*C*). Interestingly, the pattern of results reported by Cohen and Maunsell (2011)

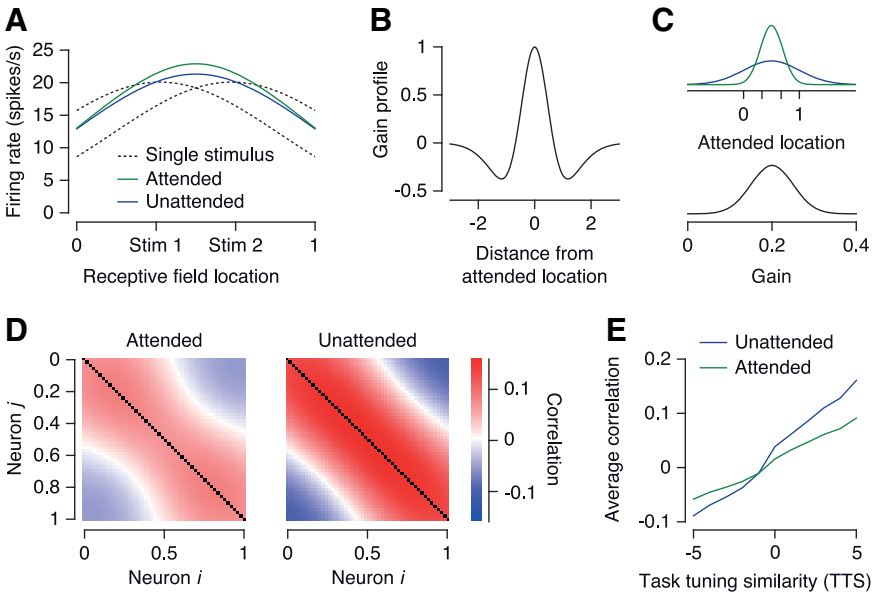

Figure 9. Model of data reported by Ruff and Cohen (2014). ***A***, Model of a population with receptive fields covering a range of locations (here arbitrarily between 0 and 1). Within this range, two small stimuli are shown ("Stim 1" and "Stim 2"), both of which activate all neurons to some extent. Dashed lines indicate sensory population response to individual stimuli (neurons are ordered by their receptive field location). Solid lines indicate response to both stimuli presented simultaneously in both unattended (blue) and attended (green) condition. ***B***, Gain profile for spatial attention. In this task, space is a relevant feature and neurons have different receptive field locations, so we include a gain profile with suppressive surround. ***C***, Distribution of attended locations (top; blue represents unattended; green represents attended) and attentional gain (bottom). Although the distribution of gains remains unaltered in the unattended condition (only the attentional focus fluctuates), the average gain of neurons around the stimulus location increases (see ***A***). ***D***, Correlation matrix for attended (left) and unattended (right) condition. ***E***, Average correlations in attended (green) and unattended (blue) condition as a function of TTS, as reported by Ruff and Cohen (2014).

To reconcile this pattern of results, we model a population of neurons whose receptive field centers cover both stimulus locations. We chose the receptive field sizes large enough such that all neurons respond to both stimuli, albeit with different intensities (Fig. 9*A*). When considering nearby locations, spatial attention behaves much like feature attention in that the central enhancement of the attended location is accompanied by a suppressive surround (Bahcall and Kowler, 1999; Intriligator and Cavanagh, 2001; Müller and Kleinschmidt, 2004; Hopf et al., 2006; Sundberg et al., 2009). We therefore incorporate a center-surround gain profile as in our discussion of feature attention (Fig. 9*B*). Now, if we assume that in the attended condition the spatial focus of attention is less variable than in the unattended condition, we can reproduce the pattern of results obtained by Ruff and Cohen (2014). Specifically, the interaction of fluctuations in the attended feature with the gain profile leads to the unique pattern of correlation changes (Fig. 9*E*): although all neurons increase their firing rates under attention (Fig. 9*A*), pairs with strongly overlapping receptive fields have decreased correlations while pairs with less overlapping receptive fields have increased correlations (Fig. 9*D*,*E*).

(Fig. 8*D*,*F*) can be reproduced by two entirely different scenarios with respect to the unattended feature, spatial frequency. The first possibility is to assume that the attentional gain for spatial frequency fluctuates more strongly from trial to trial, but the attentional focus remains on the stimulus that is shown (albeit with a lower average gain). This hypothesis would lead to essentially the same correlation structure as for orientation, but with higher magnitude (Fig. 8*C*). The second possibility is not to invoke increased gain fluctuations, but instead assume that the focus of attention fluctuates from trial to trial (Fig. 8*E*; i.e., a random spatial frequency is attended). This hypothesis would lead to a different correlation structure with respect to the neurons' preferred spatial frequency, but the same pattern of results (Fig. 8*F*) when plotted as a function of firing rate changes as in Cohen and Maunsell (2011). Resolving which of these two hypotheses (if any) is correct would advance our understanding of how attentional resources are allocated in the brain. Unfortunately, how correlation changes relate to firing rate changes is uninformative in this respect.

This dissociation between attention effects on firing rates and correlations is further supported by a second recent study showing that attention-induced increases in firing rates can be associated with either increased or decreased correlations, depending on how similarly two neurons respond to a pair of nearby gratings in a contrast discrimination task (Ruff and Cohen, 2014). The authors' similarity measure (TTS) depends largely on the degree to which the neurons' receptive fields overlap, such that spike count correlations between neurons with overlapping receptive fields decrease in attended relative to unattended conditions, whereas those between neurons with nonoverlapping receptive fields increase in attended conditions (Fig. 9*E*).

Notably, this model does not only reproduce the qualitative changes of correlations, but also a number of more subtle patterns in the results of Ruff and Cohen (2014): the crossing of the two lines in Figure 9*E* is shifted slightly to negative values; firing rates increase much more modestly for neurons with negative TTS than for those with positive TTS; and the changes in correlation are stronger for neurons with positive TTS. Interestingly, fluctuations in the focus of spatial attention were the only mechanism that we found could account for the observed pattern of correlation changes as a function of TTS. Increased fluctuations of the attentional gain (as above) could not account for the pattern of results in this study. However, from the perspective of a single neuron (or multiple neurons with identical receptive fields), increased fluctuations of the attended features look identical to increased gain variability.

Identifying attentional fluctuations in experimental data

We saw above that fluctuations in attentional state can introduce interesting patterns of correlation in neural activity, which are consistent with the published literature on attention. However, as long as one considers only single neurons and pairwise statistics, any result can be consistent with many hypotheses. For instance, attentional fluctuations induce correlations that depend on firing rates (Fig. 2*C*), but the same result is also predicted by the thresholding nonlinearity of neurons (de la Rocha et al., 2007) and therefore need not result from attentional fluctuations. Similarly, all types of attentional fluctuations considered above lead to correlations that decrease with the difference of two neurons' stimulus preferences (limited range correlations; Figs. 2*E*, 4*D*, 5*C*), but this correlation structure can also arise from shared sensory noise (Shadlen et al., 1998). Finally, changes in the correlation

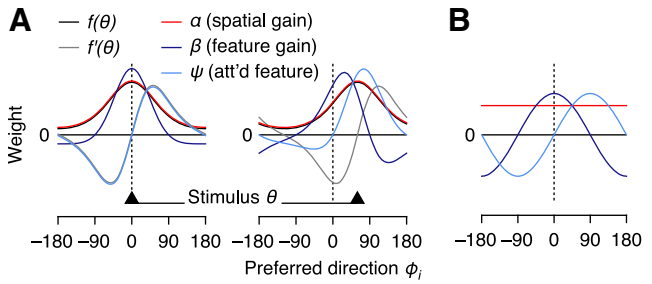

Figure 10. Identifying attentional fluctuations from variability in neuronal population activity. ***A***, The subspace identified by Factor Analysis depends on the stimulus direction. Black triangles represent stimulus direction (left, $\theta = 0°$; right, $\theta = 60°$). Solid lines indicate basis functions (attention axes) corresponding to fluctuations in spatial attention gain (red), feature attention gain (dark blue), and attended direction (light blue); population tuning curve (black) and its derivative (gray). Vertical dashed line indicates (average) attended direction. ***B***, Principal components identified by Exponential Family Principal Component Analysis are independent of the stimulus because the log-link turns a multiplicative modulation into an additive offset, thereby defining stimulus-independent attention axes. Colors are as in ***A***.

structure between attended and unattended conditions could either arise from attention-induced changes in effective connectivity between neurons or, as our model suggests, from the fluctuating attentional state unknown to the experimenter.

How would one go about identifying attentional fluctuations in experimental data? To do so, we have to consider the response patterns of simultaneously recorded populations of neurons rather than just pairwise correlations. In the following, we discuss some predictions our model makes for the structure of the neural population response.

A first approach is suggested by our analyses above: we showed that, in all cases we analyzed, the covariance matrix induced by attentional fluctuations is diagonal plus rank one. Thus, each type of attentional fluctuation is restricted to a one-dimensional subspace (sometimes referred to as the attention axis; Cohen and Maunsell, 2010), which could be inferred from simultaneously recorded neurons by Factor Analysis or perhaps directly measured by appropriately cuing the animal. However, because attentional modulation is multiplicative, this subspace depends on the stimulus: Figure 10*A* shows how the three attention axes defined by attentional fluctuations (corresponding to spatial gain, feature gain, and attended feature) change for different stimulus directions (left: $\theta = 0°$; right: $\theta = 60°$). As these axes are not simply shifted versions of each other, one cannot pool data over multiple stimulus conditions. Moreover, if the attended direction of motion does not match the stimulus direction, the attention axes related to feature attention do not peak at either neurons tuned to the stimulus or the attended direction, but somewhere in between (Fig. 10*A*, right, blue lines, where $\psi = 0°$ and $\psi = 60°$). Thus, it is nontrivial to recover the quantities of interest to the experimenter: the attended feature (direction) and the degree of attention allocated (the gain).

A model that could directly extract attentional gains (spatial and feature gain) and the attended feature would be desirable. Fortunately, it turns out that such models exist and are relatively straightforward to apply. As shown in Materials and Methods, we can convert our model into a GLM by a simple reparameterization of the stimulus. Essentially, both the stimulus and the attentional modulation affect the log firing rate additively and independently. As a consequence, we can infer the linear subspace (attention axes) corresponding to attentional fluctuations from population activity using methods, such as Exponential Family Principal Component Analysis (Collins et al., 2001; Mohamed et al., 2009) or Poisson Linear Dynamical Systems (Macke et al., 2011; Buesing et al., 2012). Moreover, this subspace is independent of the stimulus (Fig. 10*B*). Fluctuations of the spatial attention gain correspond to an additive offset common to all neurons (Fig. 10*B*, red), whereas the subspace spanned by fluctuations in the attended direction and its gain is given by $[\cos\phi_i, \sin\phi_i]$ (Fig. 10*B*, light and dark blue).

Discussion

We have presented a simple model of neuronal responses under attention, which is built on just two key ingredients: that attention acts as a multiplicative gain factor on neuronal responses (Maunsell and Treue, 2006) and that the state of attention fluctuates from trial to trial (Cohen and Maunsell, 2010, 2011). Although both assumptions are fairly uncontroversial, the importance of their combined effects when studying correlations in neuronal population responses has not been fully appreciated. We have shown that such a simple model can account for a range of empirically observed phenomena, such as super-Poisson variability (Ecker and Tolias, 2014; Goris et al., 2014) as well as patterns of (correlated) variability under attention (Mitchell et al., 2009; Cohen and Maunsell, 2010, 2011; Herrero et al., 2013; Ruff and Cohen, 2014).

Our results argue that it is likely that some fraction of variability in the neuronal response can be attributed to fluctuations in behaviorally relevant, internally generated signals, such as attention, rather than shared sensory noise (Nienborg and Cumming, 2009; Ecker et al., 2010, 2014; Ecker and Tolias, 2014; Goris et al., 2014; Haefner et al., 2014). However, exactly what fraction of correlated variability observed in experimental studies can be attributed to such attentional fluctuations remains an empirical question that cannot be answered based on the available published data. We have suggested ways to address this question by identifying attentional fluctuations directly from simultaneous population recordings using latent variable models (Collins et al., 2001; Mohamed et al., 2009; Macke et al., 2011; Buesing et al., 2012; Pillow and Scott, 2012).

We stress that our model incorporates a number of simplifications and therefore cannot capture all aspects of interneuronal correlations. First, we have deliberately ignored any correlations arising from common feedforward inputs or recurrent connectivity, mostly because we expect them to be small for the majority of pairs (Ecker et al., 2010; Renart et al., 2010). However, we expect our analysis to remain valid at least qualitatively even if there are substantial correlations in the data that are due to other sources. Second, by modeling attention on the phenomenological level and treating it as a common gain, we have ignored the question of how such a gain modulation may be implemented in a neural network (Bejjanki et al., 2011) and reduced attentional fluctuations to modulations in one-dimensional subspaces. Although this simplification will miss any changes in the correlation structure that are due to the underlying network mechanisms, we note that there are very few experimental data available to constrain more mechanistic, network-level models. We therefore favored the more simplistic approach, which can already account for a remarkable variety of nontrivial experimental findings. Third, we have considered each type of attentional fluctuation (spatial gain, feature gain, attended feature) individually. However, in practice, all types of fluctuations, as well as many others, are likely to occur at the same time. The combined effects of different types of attentional fluctuations will depend on the correlations between different attentional processes. Although there is some experimental evidence that spatial attention is un-

correlated between hemispheres (Cohen and Maunsell, 2010) and that spatial and feature attention are uncorrelated (Cohen and Maunsell, 2011), the correlations between different attentional processes are likely to depend on the task in general. Because the variability in the population response that is due to attentional fluctuations will lie within the subspace spanned by the individual components, the dependencies within this subspace can inform us about the (in)dependence of different attentional processes.

Further, although it is generally accepted that both spatial and feature attention act as gain-modulating signals (McAdams and Maunsell, 1999; Maunsell and Treue, 2006), they are not the only factors capable of modulating neuronal gain. As such, our results apply more generally to any gain-modulating signal, rather than exclusively to attention. It should also be cautioned that, in certain contexts, the effects of attention may extend beyond gain modulation (e.g., shifting contrast response functions) (Reynolds et al., 2000). Whether such effects can be attributed to differences in stimulus parameters and how such differences interact with attentional and other task strategies that subjects use (Reynolds and Heeger, 2009) or whether certain paradigms engage internal signals in addition to attention is an important empirical question in need of conclusive resolution. Our gain model of attention does not make any assumption about what mechanisms are causing gain fluctuations. Instead, it serves as a parsimonious model that is sufficient to reproduce a number of the main neurophysiological effects described in the attention literature, and accounts for a number of findings on neuronal variability as well that have not been sufficiently appreciated before.

Our model can also be interpreted in the context of the hypothesis that the brain performs probabilistic inference (Mumford, 1991; Lee and Mumford, 2003). In this view, neuronal populations represent not the sensory stimulus itself, but instead the subject's belief about certain features of the stimulus. In other words, the neural response depends not only on the stimulus, but also on the subject's prior, which could be implemented by the attentional gain. Haefner et al. (2014) further developed this idea and proposed a model in which the brain implements Bayesian inference by neural sampling. Their model makes predictions for the correlation structure during discrimination tasks that are very similar to those of our model under fluctuations in the attended feature (Fig. 6). However, the sources of correlation differ between the two models. In our model, attention is treated as a prior, and the correlations between neurons arise from trial-to-trial variability of the prior. In their model, in contrast, differential correlations arise because neural activity represents samples from the posterior. The magnitude of differential correlations is therefore directly related to the width of the subject's posterior, and their timescale is determined by the dynamics of the sampling process. The difference between the two models is seen most clearly if we assume that the prior was constant across trials (i.e., no attentional fluctuations). In this case, their model predictions would not change at all, whereas our model would not predict any correlations. Thus, our model puts the emphasis on the trial-to-trial fluctuations of the prior, which, despite being suboptimal, seem to be present. For instance, it has long been known that there are serial dependencies in subjects' responses (Fernberger, 1920; Senders and Sowards, 1952), which indicate that subjects bias their estimates depending on the past stimuli they have seen, despite the fact that there is no real dependence in the stimuli that are shown. Our model also resonates well with recent behavioral data showing that noise in the prior is an important component for models of human probabilistic inference (Acerbi et al., 2014). Therefore, in our model, the timescale of correlations corresponds to the timescale at which subjects adapt their prior expectations about the stimuli they see. Ultimately, we expect that both processes (the mechanisms of the inference process itself and the variability in the subject's priors from trial to trial) create correlated variability. Separating these aspects through their underlying timescales and/or manipulations in the task contingencies is an interesting avenue for further theoretical and experimental work.

In addition to offering a parsimonious account of neuronal variability and covariability, our model has implications for how we should interpret the effect of attention as it relates to improvements in perceptual performance. Recent studies argue that spatial attention improves behavioral performance primarily by reducing correlations (Cohen and Maunsell, 2009; Mitchell et al., 2009). However, if the reduction of correlations observed under attention is indeed due to a suppression of attentional gain fluctuations, as our model would suggest, this reduction of correlations is irrelevant for the coding accuracy of the population and cannot be the mechanism improving behavioral performance. In terms of Fisher information, the only difference that matters in this case is the increase in response gain, which leads to a proportional increase in Fisher information, at least so long as performance is not limited by input noise.

Although gain fluctuations are irrelevant for coding, this is not true for all types of attentional fluctuations: when the fluctuations in attention occur around a specific feature value rather than in the gain, they introduce differential correlations, a pattern of correlations that leads to information saturation (Moreno-Bote et al., 2014). Thus, our model leads to an interesting, but at the same time puzzling, observation: because the attentional state is represented by a finite number of neurons, there is necessarily some trial-to-trial variability in the attended feature itself. Such fluctuations indeed impair the readout because it cannot have exact (i.e., noiseless) access to the attended feature when implemented in neural hardware. Thus, the precision of the attentional mechanism itself places a limit on how accurately a stimulus can be represented by a sensory population, and this limit can at least in principle be substantially lower than the amount of sensory information entering the brain through the eye.

If attentional fluctuations can limit behavioral performance, one may ask why there should be an attentional mechanism in the first place. We will give two speculative answers in the following.

First, as discussed earlier, we can think of attention as the subject's prior. Using prior information to bias an estimate toward more likely solutions improves the estimate on average (over all possible stimuli) if the distribution of stimuli is nonuniform. In the real world, in situations where the sensory input is noisy or ambiguous and decisions have to be made quickly, such a bias is usually beneficial and outweighs the small extra noise added due to variability in the prior.

Second, the primary goal of attention may not be to improve the sensory representation, as suggested by recent studies characterizing neural responses (Cohen and Maunsell, 2009; Mitchell et al., 2009). In typical laboratory experiments involving two-alternative forced-choice discrimination at perceptual threshold, there is no prior information that could be used to improve performance. The feedforward sensory signal contains all the information available to the organism, so it is unclear where the additional information should come from. However, in the real world, the sensory signal usually contains a lot of nuisance variables, which are irrelevant to the task and thus act as a crucial

noise source. Therefore, the goal of attention may be instead, as suggested by the traditional view (Broadbent, 1958), to select the relevant pieces of information and suppress the irrelevant. This selection is precisely what the gain profile implements: it enhances attended stimuli and suppresses unattended ones. The behavioral improvement due to attention may therefore not be a result of an improved representation at the level of sensory populations but instead result from the fact that the decision is not corrupted by irrelevant information (noise or nuisance) from the distractors.

Notes

Supplemental material for this article is available at http://bethgelab.org/code/ecker2015. It contains MATLAB code to reproduce all figures and numerical simulations in this paper. This material has not been peer reviewed.

References

Acerbi L, Vijayakumar S, Wolpert DM (2014) On the origins of suboptimality in human probabilistic inference. PLoS Comput Biol 10:e1003661. CrossRef Medline

Bahcall DO, Kowler E (1999) Attentional interference at small spatial separations. Vision Res 39:71–86. CrossRef Medline

Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area MT: time scales and relationship to behavior. J Neurosci 21:1676–1697. Medline

Beck J, Bejjanki VR, 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. CrossRef Medline

Bejjanki VR, Beck JM, Lu ZL, Pouget A (2011) Perceptual learning as improved probabilistic inference in early sensory areas. Nat Neurosci 14:642–648. CrossRef Medline

Berens P, Ecker AS, Gerwinn S, Tolias AS, Bethge M (2011) Reassessing optimal neural population codes with neurometric functions. Proc Natl Acad Sci U S A 108:4423–4428. CrossRef Medline

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1993) Responses of neurons in macaque MT to stochastic motion signals. Vis Neurosci 10:1157–1169. CrossRef Medline

Broadbent DE (1958) Perception and communication. Elmsford, NY: Pergamon.

Buesing L, Macke JH, Sahani M (2012) Learning stable, regularised latent models of neural population dynamics. Network 23:24–47. CrossRef Medline

Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci 12:1594–1600. CrossRef Medline

Cohen MR, Maunsell JH (2010) A neuronal population measure of attention predicts behavioral performance on individual trials. J Neurosci 30:15241–15253. CrossRef Medline

Cohen MR, Maunsell JH (2011) Using neuronal populations to study the mechanisms underlying spatial and feature attention. Neuron 70:1192–1204. CrossRef Medline

Collins M, Dasgupta S, Schapire RE (2001) A generalization of principal components analysis to the exponential family. In: Advances in neural information processing systems 14, pp 617–624. Cambridge: MIT.

Dean AF (1981) The variability of discharge of simple cells in the cat striate cortex. Exp Brain Res 44:437–440. Medline

de la Rocha J, Doiron B, Shea-Brown E, Josiæ K, Reyes A (2007) Correlation between neural spike trains increases with firing rate. Nature 448:802–806. CrossRef Medline

Ecker AS, Tolias AS (2014) Is there signal in the noise? Nat Neurosci 17:750–751. CrossRef Medline

Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits. Science 327:584–587. CrossRef Medline

Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. J Neurosci 31:14272–14283. CrossRef Medline

Ecker AS, Berens P, Tolias AS, Bethge M (2012) The correlation structure induced by fluctuations in attention. Cosyne Abstr III-46.

Ecker AS, Berens P, Cotton RJ, Subramaniyan M, Denfield GH, Cadwell CR, Smirnakis SM, Bethge M, Tolias AS (2014) State dependence of noise correlations in macaque primary visual cortex. Neuron 82:235–248. CrossRef Medline

Fernberger SW (1920) Interdependence of judgments within the series for the method of constant stimuli. J Exp Psychol 3:126–150. CrossRef

Goris RL, Movshon JA, Simoncelli EP (2014) Partitioning neuronal variability. Nat Neurosci 17:858–865. CrossRef Medline

Haefner RM, Berkes P, Fiser J (2014) Perceptual decision-making as probabilistic inference by neural sampling. arXiv:1409.0257 [q-bio]. http://arxiv.org/abs/1409.0257.

Herrero JL, Gieselmann MA, Sanayei M, Thiele A (2013) Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. Neuron 78:729–739. CrossRef Medline

Hopf JM, Boehler CN, Luck SJ, Tsotsos JK, Heinze HJ, Schoenfeld MA (2006) Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. Proc Natl Acad Sci U S A 103:1053–1058. CrossRef Medline

Intriligator J, Cavanagh P (2001) The spatial resolution of visual attention. Cogn Psychol 43:171–216. CrossRef Medline

Kay SM (1993) Fundamentals of statistical signal processing, Vol. I: Estimation theory, Ed 1. Englewood Cliffs, NJ: Prentice Hall.

Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am A Opt Image Sci Vis 20:1434–1448. CrossRef Medline

Macke JH, Buesing L, Cunningham JP, Yu BM, Shenoy KV, Sahani M (2011) Empirical models of spiking in neural populations. In: Advances in neural information processing systems 24, p 13501358. Available at: http://papers.nips.cc/book/advances-in-neural-information-processing-systems-24-2011.

Maunsell JH, Treue S (2006) Feature-based attention in visual cortex. Trends Neurosci 29:317–322. CrossRef Medline

McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 19:431–441. Medline

Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron 63:879–888. CrossRef Medline

Mohamed S, Ghahramani Z, Heller KA (2009) Bayesian exponential family PCA. In: Advances in neural information processing systems 22, pp 1089–1096. Available at: http://papers.nips.cc/book/advances-in-neural-information-processing-systems-22-2009.

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

Müller NG, Kleinschmidt A (2004) The attentional 'spotlight's' penumbra: center-surround modulation in striate cortex. Neuroreport 15:977–980. CrossRef Medline

Mumford D (1991) On the computational architecture of the neocortex. Biol Cybern 65:135–145. CrossRef Medline

Newsome WT, Paré EB (1988) A selective impairment of motion perception following lesions of the middle temporal visual area (MT). J Neurosci 8:2201–2211. Medline

Nienborg H, Cumming BG (2009) Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature 459:89–92. CrossRef Medline

Pillow JW, Scott JG (2012) Fully Bayesian inference for neural models with negative-binomial spiking. In: Advances in neural information processing systems 25, pp 1907–1915. Available at: http://papers.nips.cc/book/advances-in-neural-information-processing-systems-25-2012.

Rabinowitz NC, Goris RL, Cohen M, Simoncelli E (2015) Attention stabilizes the shared gain of V4 populations. eLife 4:e08998. CrossRef Medline

Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, Harris KD (2010) The asynchronous state in cortical circuits. Science 327:587–590. CrossRef Medline

Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168–185. CrossRef Medline

Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. Annu Rev Neurosci 27:611–647. CrossRef Medline

Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. Neuron 26:703–714. CrossRef Medline

Ruff DA, Cohen MR (2014) Attention can either increase or decrease spike

count correlations in visual cortex. Nat Neurosci 17:1591–1597. CrossRef Medline

Senders VL, Sowards A (1952) Analysis of response sequences in the setting of a psychophysical experiment. J Psychol 65:358–374. CrossRef Medline

Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci 18:3870–3896. Medline

Shamir M, Sompolinsky H (2006) Implications of neuronal diversity on population coding. Neural Comput 18:1951–1986. CrossRef Medline

Smith MA, Kohn A (2008) Spatial and temporal scales of neuronal correlation in primary visual cortex. J Neurosci 28:12591–12603. CrossRef Medline

Smith MA, Sommer MA (2013) Spatial and temporal scales of neuronal correlation in visual area V4. J Neurosci 33:5422–5432. CrossRef Medline

Sundberg KA, Mitchell JF, Reynolds JH (2009) Spatial attention modulates center-surround interactions in macaque visual area V4. Neuron 61: 952–963. CrossRef Medline

Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res 23: 775–785. CrossRef Medline

Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399:575–579. CrossRef Medline

Treue S, Maunsell JH (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. Nature 382:539–541. CrossRef Medline

Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370:140–143. CrossRef Medline