Attention-related changes in correlated neuronal activity arise from normalization mechanisms

Bram-Ernst Verhoeef1,2 & John H R Maunsell1

Attention-related changes in correlated neuronal activity arise from normalization mechanisms. How attention modifies response correlations in order to flexibly adjust the quality of sensory representations depending on their behavioral relevance for the task at hand14–18. For example, one recent study suggested that attention improves perceptual decisions that are based on the difference in the average activities of different pools of neurons by decreasing correlations within each neuronal pool while increasing correlations between activities of neurons in different pools15. However, it remains uncertain whether attention actively adjusts correlated responses to enhance perceptual behavior or if instead these attention-related changes in response correlations are incidental to attention-related changes in the strength of neuronal responses.

Mounting evidence indicates that the magnitude of attention-related changes in the response strength of neurons is mediated by normalization mechanisms, which can greatly amplify attention-related changes in neuronal responses in some stimulus configurations21–26. Normalization mechanisms operate in many sensory modalities and brain regions, and they underlie a range of neuronal phenomena, such as surround suppression, contrast-response functions and multisensory integration, to name a few27. Yet if and how normalization mechanisms influence correlated neuronal activity is unclear.

We examined how neuronal correlations relate to normalization mechanisms and examined the mechanisms through which attention modifies neuronal correlations. We showed that normalization mechanisms can either increase or decrease response correlations and that attention can increase or decrease response correlations. Notably, our findings indicate that these attention-related increased or decreased neuronal correlations were an indirect consequence of attention-related changes in the inputs into normalization mechanisms. Finally, we showed that similar mechanisms influenced response correlations inside the classical receptive field and the surround, a finding that helps explain previously observed decreases in correlation by attention.

RESULTS

Using chronically implanted microelectrode arrays (Fig. 1a), we recorded from 12,067 multiunits in visual area V4 in the left cerebral hemisphere of two rhesus monkeys (monkey M1, 4,709 units; monkey M2, 7,358 units) while they performed a visual-detection task in which spatial attention was controlled (Online Methods). During task performance, we presented either single or paired stimuli near the receptive fields (RFs) of the recorded neurons, either inside the classical RF or in the surround (Online Methods and Supplementary Fig. 1). During different blocks of trials, monkeys attended to different stimulus locations, with one attended location per block of trials.

Attention was directed toward one of two stimulus locations near the receptive fields of the recorded neurons, either inside the classical RF or in the surround, a finding that helps explain previously observed decreases in correlation by attention.
Normalization mechanisms can increase or decrease spike-count correlations

We first examined how normalization mechanisms affect spike-count correlations in the absence of attention, i.e., correlated fluctuations of neuronal responses across repeated presentations of the same stimuli with attention directed to the other hemifield.

In a mechanistic model, the excitatory and suppressive contributions of each stimulus to the neuronal response are determined by fitting a normalization model to the observed neuronal responses across repeated presentations of the same stimuli.

Normalization mechanisms can increase or decrease spike-count correlations

We found that two aspects of stimulus-related excitation and suppression were critical for understanding the effects of normalization on spike-count correlations. The first is whether the two neurons of a neuron pair prefer the same stimulus. For each neuron, we designated the preferred and nonpreferred stimulus of a stimulus pair as the stimulus that contributed the most and least excitation, respectively. For each neuron pair, we defined a measure of selectivity (Online Methods) that was positive when both neurons of the pair preferred the same stimulus and negative when the two neurons preferred different stimuli, with larger magnitudes reflecting greater selectivity (Fig. 2a). The second critical factor is how strongly the responses of each pair of neurons were suppressed by the nonpreferred stimulus. We defined a measure of nonpreferred suppression (Online Methods) that revealed whether both neurons of a pair received weak (values near 0) or strong (values near 1) suppression from their nonpreferred stimulus (Fig. 2a).

Using the very large set of V4 neuronal pairs, we could examine how selectivity and nonpreferred suppression affected spike-count correlations in the absence of attention to stimuli inside the RF, i.e., during blocks of trials with attention directed far from the RF stimuli. Whenever a single stimulus (preferred or nonpreferred) was presented alone, average spike-count correlations were small and positive for all neuron pairs, regardless of their selectivity and nonpreferred suppression (Fig. 2b). However, adding a second stimulus changed the structure of spike-count correlations across the population of neuron pairs in a specific way: spike-count correlations increased between neuron pairs with the same selectivity but only when both neurons of the pair were strongly suppressed by their nonpreferred stimulus (Fig. 2c–e; main effect of selectivity, $P < 0.0001$; main effect of suppression, $P < 0.0001$; selectivity-suppression interaction, $P < 0.0001$; linear regression, $n = 2,533,424$ correlations resulting from different combinations of neuron and stimulus pairs). Conversely, spike-count correlations decreased between neuron pairs with opposite selectivity but, again, only when both neurons of the pair were strongly suppressed by each neuron's nonpreferred stimulus. Notably, these neuron pairs became negatively correlated (Fig. 2c–e; main effect of selectivity, $P = 0.02$; main effect of suppression, $P = 0.003$; selectivity-suppression interaction, $P < 0.0001$; $n = 1,270,826$ correlations).

In the previous analyses, we treated each pair of neurons as an independent sample for statistical purposes. We also performed an additional analysis on the average results from individual sessions and found similar results. Specifically, for each session, we computed the average difference in spike-count correlations between single- and paired-stimulus presentations within each quadrant of Figure 2e. Across sessions, we observed a significant effect of selectivity, nonpreferred suppression, and the interaction ($P < 0.0001$ for all effects; two-way repeated-measures ANOVA). Similar effects were also observed when analyzing the data from the preferred and nonpreferred stimulus separately (Supplementary Fig. 2).

A simple mechanism provides insights into the dynamics behind these changes in correlation. This mechanism incorporates two mutually suppressive neuronal populations (Fig. 2f). Neurons with similar stimulus preferences belonged to the same neuronal population. For example, neurons 1 and 2 in Population A. Neurons in different neuronal populations had different preferred stimuli; for example, neuron 1 in Population A and neuron 3 in Population B. Thus, the preferred stimulus of Population A was the nonpreferred stimulus of Population B and vice versa. When a pair of stimuli was presented, neurons in both populations became active, while also suppressing neurons in the other population. Suppression from Population B was shared among neurons in Population A, and this shared suppression produced...
Figure 2 Normalization mechanisms determine spike-count correlations. All data in these plots were obtained while monkeys’ attention was directed far from the RF stimuli (locations 3 and 4 in Fig. 1b). (a) Each neuron’s preferred stimulus in a stimulus pair was the stimulus that contributed most excitation (blue arrows). Positive selectivity indices refer to neuron pairs with the same stimulus preference (top, blue arrows). Negative selectivity indices refer to neuron pairs with opposite stimulus preferences (bottom, blue arrows). Values of nonpreferred suppression near 0 indicate that the two neurons of a pair were weakly suppressed by their nonpreferred stimulus (left, thin orange dashed lines); values near 1 indicate strong suppression by their nonpreferred stimulus (right, thick orange dashed lines). For neurons with opposite selectivity (bottom), the preferred stimulus of one neuron is the nonpreferred stimulus of the other neuron. (b) Mean spike-count correlations, indicated by color, as a function of the selectivity and nonpreferred suppression of neuron pairs, measured while a single stimulus was presented alone near the RF. (c) Mean spike-count correlations during paired stimulus presentations. (d) Difference in spike-count correlations between c and b. Plots are based on regularized bilinear interpolation (Online Methods). (e) Mean spike-count correlations (r_sc) computed on the data from four quadrants in the space spanned by selectivity and nonpreferred suppression (quadrants defined as a combination of negative or positive selectivity and nonpreferred suppression < 0.5 or > 0.5). Black, paired-stimulus presentations; gray, single-stimulus presentations. Error bars represent ±1 s.e.m. (f) Pattern of spike-count correlations can be explained by two oppositely tuned neuronal populations, Populations A and B, that mutually suppress each other’s activity. Common suppression, evoked by neurons in Population B (lower two dashed yellow lines), correlates the activities of neurons 1 and 2 in Population A but decorrelates the activities of neurons in different populations (for example, neurons 1 and 3).

a positive correlation between pairs of neurons in population A. Neuron pairs in population A that were more strongly suppressed by neurons in population B (nonpreferred suppression near 1) received more common suppressive inputs and became more positively correlated. The neuron pair consisting of neurons 1 and 2 in Population A might lie near the upper right corner of Figure 2a–e if the suppression they received from Population B was strong or near the upper left corner of Figure 2a–e if that suppression was weak.

While shared suppression resulted in positive correlations between neurons with similar preferences, neurons with opposite selectivity became negatively correlated. This followed from the same mutually suppressive mechanisms: larger responses in Population B more strongly suppressed responses in Population A, causing neuronal responses in population A and B to become negatively correlated. Stronger suppression from the other population (nonpreferred suppression near 1) caused more decorrelation. The neuron pair consisting of neurons 1 and 3 in Figure 2f would lie near the lower right corner in Figure 2a–e if the suppression each received from the other population was strong.

The striking pattern of stimulus-induced changes in spike-count correlations (Fig. 2c–e) suggested that normalization mechanisms strongly influenced the structure of correlated neuronal activity.

Attention-related increased or decreased spike-count correlations arise from normalization mechanisms
Visual attention engages normalization circuitry21–25 and influences neuronal correlations3,8,14–20. We sought to determine whether attention-related changes in spike-count correlations arise from the same normalization mechanism described above. The effects of attention were measured with two stimuli simultaneously presented near the RFs, i.e., the same paired-stimulus configurations and neuron pairs for which selectivity and nonpreferred suppression were measured in Figure 2.

Attention-related changes in spike-count correlations were robust only for neuron pairs with the same stimulus preferences (selectivity
> 0). We first focus on these pairs in Figure 3 (see below for neuron pairs with opposite stimulus preferences). Attending to the preferred stimulus of similarly tuned neurons (‘Attend Preferred’) decreased spike-count correlations relative to when attention was directed far from the RF (i.e., relative to Figure 2c; see Figure 3a,d; main effect of non-preferred suppression, \( P = 0.48 \); main effect of selectivity, \( P < 0.0001 \); selectivity–suppression interaction, \( P < 0.0001 \); linear regression, \( n = 1,266,712 \) correlations).

Attending to the non-preferred stimulus of these neurons (‘Attend Nonpreferred’) increased spike-count correlations compared to when attention was directed far from the RF (Figure 3b,d; main effect of non-preferred suppression, \( P < 0.0001 \); main effect of selectivity, \( P < 0.0001 \); selectivity–suppression interaction, \( P < 0.0001 \); linear regression, \( n = 1,266,712 \) correlations). This was unexpected, because the effect of attention to an RF stimulus is typically described as reducing spike-count correlations.

We next compared the effects of attending to the preferred or nonpreferred stimulus (Figure 3a,b). When shifting attention between the preferred and the nonpreferred stimuli, spike-count correlations modulated most for similarly tuned neuron pairs that were most selective and were most suppressed by their nonpreferred stimulus (Figure 3c,d).

These results were confirmed in a session-by-session analysis. For each session we obtained the average attention modulation of spike-count correlations within each quadrant of Figure 3d. Across sessions there were significant effects of selectivity (\( P = 0.007 \)), nonpreferred suppression (\( P = 0.006 \)) and their interaction (\( P = 0.04 \)). Similar results were obtained when the average responses of pairs of neurons were matched across conditions (Online Methods).

These attention-related changes in spike-count correlations emanated from the same mutually suppressing neuronal populations considered earlier (Figure 3e,f). Attending to the preferred stimulus of neurons with the same selectivity, for example, neurons 1 and 2 in Population A increased their responses (Figure 3e). Attending to the preferred stimulus of neurons in Population A, Population A more strongly suppressed neurons in Population B (Figure 3e), causing responses in Population B to decrease (Figure 3e). Because responses in Population B decreased, the correlating suppressive inputs from Population B into Population A became weaker (Figure 3e).
decreasing the spike-count correlations within Population A, as shown in Figure 3a.

Conversely (Fig. 3f), attending to the nonpreferred stimulus of neurons in Population A increased activity in Population B, because now Population B's preferred stimulus was attended. The increased activity in Population B amplified the common suppression from Population B into Population A (Fig. 3f). This increased common suppression to Population A correlated activity in Population A, as in Figure 3b. Thus, attention-related changes in spike-count correlations stemmed from normalization mechanisms.

No attention-related changes in spike-count correlations for oppositely tuned neurons
For pairs of neurons with opposite stimulus preferences (selectivity \(< 0\)), one neuron's preferred stimulus is the other neuron's nonpreferred stimulus. Attend Preferred and Attend Nonpreferred are not defined for such pairs of neurons (Attend Preferred for one neuron of the pair is Attend Nonpreferred for the other neuron of the pair). We therefore compared conditions in which attention was directed far from the RF to conditions in which attention was directed toward one of the two RF stimuli.

For oppositely tuned pairs of neurons, attending to a RF stimulus had little effect on spike-count correlations relative to when attention was directed far from the RF (Fig. 4a,b). Such small effects of attention on the spike-count correlations of pairs of oppositely tuned neurons were predicted by the normalization model with oppositely tuned mutually suppressive neuronal populations: Attending to the preferred stimulus of neurons in Population A increased the responses of these neurons (Fig. 4c). The stronger responses in Population A increased suppression toward Population B (Fig. 4c), in which responses then decreased. Due to its decreased responses, Population B sent less suppression to Population A (Fig. 4c). Thus, when attending the preferred stimulus of neurons in Population A, Population A sent more correlating suppression to Population B, which would normally result in more negative correlations between neurons in Population A and neurons in Population B (compared to Attend Far). However, because Population B sent less correlating suppression to Population A, these same correlations became more positive (compared to Attend Far). The increased correlating suppression in one direction and the decreased correlating suppression in the other direction acted to cancel each other out. The result was that attention had little overall effect on the response correlations between neurons in Population A and neurons in Population B, i.e., between neurons with opposite stimulus preferences. Taken together, the findings from similarly and oppositely tuned neuron pairs illustrate the heterogeneous effects of attention on spike-count correlations under different neuron and attention conditions.

A stochastic normalization model accounts for the spike-count correlations
To validate the intuitions provided by the model in Figures 2–4, we developed a quantitative stochastic normalization model that formalizes the intuitive model presented thus far (Online Methods). As with the intuitive model, the quantitative model consists of two mutually suppressive and oppositely tuned neuronal populations. Critically, in the quantitative model, attention operates solely by amplifying the excitatory contribution of the attended stimulus to the neuronal response: a gain change. Thus, attention-related changes in spike-count correlations can only appear as an indirect consequence of the normalization mechanism’s response to changes in its input: attention-related increased excitatory inputs amplify the inhibitory outputs from a population, resulting in altered spike-count correlations.

The stochastic normalization model captured all the main trends in the observed correlation structure: increased spike-count correlations for similarly tuned neurons when attending far (Fig. 5a), decreased spike-count correlations for oppositely tuned neurons when attending far (Fig. 5a), decreased correlations when attending to the preferred stimulus of similarly tuned neurons (Fig. 5b), increased correlations when attending to the nonpreferred stimulus of similarly tuned neurons (Fig. 5c) and little attention modulation for oppositely tuned neurons.
neurons (Fig. 5d). Thus, attention-related changes in spike-count correlations can be viewed as a consequence of attention-related changes in the strength of inputs to normalization mechanisms.

**Normalization mechanisms in the classical receptive field and in the surround shape response correlations similarly**

The previous analyses averaged across stimulus configurations with both stimuli contained within the classical receptive field (cRF) and configurations with stimuli in the neuron’s surround. We next distinguished between these stimulus conditions and examined the effects of normalization mechanisms and attention on spike-count correlations inside the cRF and in the surround.

We found that the effects of stimulus selectivity and suppression on spike-count correlations were similar for conditions with two stimuli both presented inside the classical receptive field (cRF–cRF) and stimulus conditions with one stimulus shown inside the cRF and another stimulus presented inside the surround (cRF–surround). In both stimulus configurations, we observed that selectivity and non-preferred suppression increased (similar selectivity) or decreased (opposite selectivity) spike-count correlations (Fig. 6a,b,c: cRF–cRF same selectivity: main effect of selectivity, P < 0.0001; main effect of suppression, P = 0.06; selectivity–suppression interaction, P = 0.0001; linear regression; cRF–cRF different selectivity: main effect of selectivity, P = 0.0001; main effect of suppression, P = 0.76; selectivity–suppression interaction, P < 0.0001; cRF–surround same selectivity: main effect of selectivity, P < 0.0001; main effect of suppression, P = 0.5; selectivity–suppression interaction, P < 0.0001; cRF–surround different selectivity: main effect of selectivity, P = 0.001; main effect of suppression, P = 0.004; selectivity–suppression interaction, P = 0.001). The effects were slightly smaller in the surround condition, as expected due to the weaker suppression from stimuli in the surround compared to stimuli inside the cRF.

Attention also operated similarly in both stimulus configurations, decreasing spike-count correlations when directed to the preferred stimulus of a pair and increasing response correlations when directed to the nonpreferred stimulus (Fig. 6c–f: cRF–cRF Attend Preferred: main effect of selectivity, P < 0.0001; main effect of suppression, P = 0.69; selectivity–suppression interaction, P < 0.0001; linear regression; cRF–cRF Attend Nonpreferred: main effect of selectivity, P < 0.0001; main effect of suppression, P = 0.0003; selectivity–suppression interaction, P = 0.003; cRF–surround Attend Preferred: main effect of selectivity, P < 0.0001; main effect of suppression, P = 0.04; selectivity–suppression interaction, P < 0.0001; cRF–surround Attend Nonpreferred: main effect of selectivity, P < 0.0001; main effect of suppression, P = 0.01; selectivity–suppression interaction, P = 0.005).

Note that for the cRF–surround condition, the surround stimulus was always the nonpreferred stimulus because, by definition, neurons do not respond to stimuli in the surround.
These results agree with a previous study showing that normalization mechanisms and attention operate similarly on firing rates for stimuli inside the cRF and in the surround\textsuperscript{26}. Notably, these analyses also help explain the well-established observation that attention reduces spike-count correlations when shifted to a single stimulus inside the RF\textsuperscript{17}. Specifically, in these experiments, attention shifted between a stimulus in the opposite hemifield and a stimulus in the cRF, a configuration that corresponds to the configuration with one stimulus inside the cRF and one stimulus in the surround for which we observed decreased correlations when the cRF stimulus was attended and which the model explains.

**DISCUSSION**

We showed that shared neuronal activity fluctuations within an area can be understood as arising from normalization mechanisms and provided a simple mechanistic explanation for a heterogeneous set of observations. A relationship between normalization mechanisms and spike-count correlations has been suggested previously\textsuperscript{28–31}, but the precise relationship remained unclear. Our findings show that normalization mechanisms can shape spike-count correlations through suppressive activity that affects the responses of populations of neurons. These suppressive influences can be shared among neurons with similar stimulus preferences, resulting in increased spike-count correlations, or can be antagonistic among neurons with opposite stimulus preferences, creating negative response correlations. Our results indicate that attention can bias suppressive activity as a result of elevating responses in one population of neurons and that this bias can explain why we observed that attention decreased (Attend Preferred) or increased (Attend Nonpreferred) spike-count correlations. A stochastic normalization model, consisting of two mutually suppressive but oppositely tuned neuronal populations, explained all patterns of spike-count correlations.

Previous studies have suggested that spike-count correlations arise from shared inputs\textsuperscript{32} that might take the form of common gain modulations\textsuperscript{33–35} or other shared modulatory signals\textsuperscript{33,35,36}. However, details of the mechanisms that determine these common influences have remained obscure. Our findings show that shared activity fluctuations can be determined by normalization circuits in which suppressive activity can either increase or decrease response correlations. The influence of the suppressive activity on spike-count correlations can be strong: for pairs of neurons with strong preferences for the same stimulus, stimulus-induced suppression can double spike-count correlations, compared to those observed during single-stimulus presentations. On the other hand, for pairs of oppositely tuned neurons, stimulus-induced suppression can turn positive correlations into negative correlations. Thus, suppression in normalization circuits is an important factor in shaping correlated neuronal activity—potentially the dominant source.

Normalization mechanisms underlie a broad range of response properties of neurons in different brain areas, such as contrasts tuning\textsuperscript{37}, suppression of responses to a stimulus by nearby stimuli\textsuperscript{38,39}, multisensory integration\textsuperscript{40} and attention-related modulation of neuronal responses\textsuperscript{21–26}. Given the importance of normalization mechanisms under various sensory and cognitive conditions, it is likely that the relationship between normalization mechanisms and correlated neuronal activity, as reported here, will explain spike-count correlations throughout the brain and for different experimental conditions. The current data provide one example, showing that visual attention uses normalization mechanisms to correlate or decorrelate neuronal activity.

Visual attention engages normalization circuitry by amplifying the excitation and suppression associated with the attended stimulus\textsuperscript{21–26}. We reasoned that if attention-related changes in the rate of firing are influenced by normalization mechanisms, then the effects of attention on spike-count correlations should also follow from normalization mechanisms. We showed that, depending on which stimulus is attended, attention decreased (Attend Preferred) or increased (Attend Nonpreferred) response correlations. These attention-related changes in spike-count correlations followed from the normalization model, as an indirect consequence of amplifying the excitation (gain) associated with the attended stimulus. The attention-related gain change biased the suppression between different neuronal populations in favor of the population preferring the attended stimulus, which in turn modified response correlations. Therefore, these findings provide strong support for normalization models of attention.

Conversely, because attention builds on normalization mechanisms to modify neuronal responses, we could test how perturbations in activity affect correlations in the normalization model of spike-count correlations that we proposed. The findings showed that increased or decreased activity in one population of neurons biased suppressive activity, which led to predictable changes in spike-count correlations. Crucially, in the model, attention is just one factor that can change the response strength of neurons; other factors, like stimulus contrast, are expected to exert similar influences on response correlations.

Previous studies have shown that attending a single stimulus inside the RF decreases spike-count correlations relative to when a stimulus in the opposite hemifield is attended\textsuperscript{7,14,16,17}. Our model predicts such a decrease in correlation. Specifically, the conditions in these experiments correspond to the condition in which one stimulus is shown inside the cRF and another stimulus (the stimulus in the opposite hemifield) in the surround. Our data and model demonstrated that in such configurations, spike-count correlations are lower when the cRF stimulus is attended than when the surround stimulus is attended.

It has been suggested that attention actively modifies spike-count correlations, in a task- and stimulus-dependent way, to improve sensory encoding\textsuperscript{14–18}. How attention might achieve such a complex feat is unknown. Our findings provide a more parsimonious explanation, namely that spike-count correlations are indirectly modulated by the same normalization mechanisms that have been shown to modulate the average responses of individual neurons\textsuperscript{21–26}. Note, however, that our findings do not exclude the possibility that changes in the mean rate effected by normalization are incidental to changes in correlation or the possibility that both mean rate and correlations matter for sensory processing. Our data suggest that mechanisms that are common across different tasks, like normalization mechanisms, contribute to shaping spike-count correlations.

Contrary to our findings, a previous study by Ruff and Cohen found increased spike-count correlations for pairs of neurons with opposite stimulus tuning\textsuperscript{15}. However, there are at least two important differences between the two studies that make direct comparisons difficult. First, the other study defined selectivity based on neuronal responses to stimuli that were not used during the attention task. Specifically, they defined a neuron's selectivity (i.e., spatial selectivity) based on its responses to 100% contrast stimuli at one of two locations (same orientation), as measured during instruction trials in which only one stimulus was presented. During the task, however, stimuli of different contrasts were shown, and these stimuli could be positioned at either of two receptive field positions. It is very likely that which stimulus the neurons preferred (for an individual stimulus presentation) depended on whether the high- or low-contrast stimulus was presented at the more- or less-responsive receptive field position.
For example, a high-contrast stimulus at the less-responsive RF position may elicit a stronger response than the low-contrast stimulus at the more-responsive position. In that case, the neuron would actually prefer the ‘nonpreferred’ stimulus. Thus, stimulus selectivity is likely to have differed across the stimulus configurations they used to measure spike correlations, such that their average results for neurons with opposite stimulus preferences also included pairs of neurons with similar stimulus preferences. In contrast, in our study we always defined stimulus selectivity using the same stimuli (orientation and position) during the attention task (with attention directed away from the RF).

Second, Ruff and Cohen used a different task, one that gave them less control over where monkeys directed their attention. In their study, the monkeys had to compare the contrasts of two nearby stimuli. Consequently, their monkeys were encouraged to either pay attention to both stimuli or to switch their attention between them (across or within trials). Their larger attentional field or variable locus of attention would be expected to contribute to the differences between the studies.

Ruff and Cohen found a negative correlation between normalization and spike-count correlations. However, they did not separate the data by stimulus selectivity. We also found a negative correlation between suppression and spike-count correlations but only for pairs of neurons with opposite stimulus preferences.

Attention can also affect spike-count correlations if the state of attention varies from trial to trial, causing spike rates of neurons representing a stimulus to covary. Our findings did not rely on trial-by-trial changes in the state of attention to explain attention-related changes in spike-count correlations. In the stochastic normalization model, the attention-related gain factor was constant and contributed no variability. Moreover, our data showed that, for a given trial and a given attentional state, attention modulation of correlated neuronal activity differed between pairs of neurons depending on their selectivity and on the strength of their shared suppressive inputs. So, variable attention states were not necessary to account for the observed correlations.

Mutually suppressive neuronal populations with different stimulus tuning have also been proposed as underlying perceptual decision-making. Moreover, normalization mechanisms support diverse response properties across a broad range of brain areas. Thus, it will be interesting to explore how the present results will bear on brain functions well beyond attention.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank D. Freedman for discussions. We thank J. Cone, S. Ghosh, G. Ibus and T. Luo for comments on an earlier version of the manuscript and S. Sheboda for technical assistance. B.-E.V. is supported by a postdoctoral research fellowship from the Flemish Fund for Scientific Research (FWO). This work was supported by NIH grant R01EY009911.

AUTHOR CONTRIBUTIONS

B.-E.V. and J.H.R.M. designed the experiments, performed the surgeries and wrote the paper. B.-E.V. performed the experiments and analyzed the data.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Kastner, S. & Ungerleider, L.G. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341 (2000).

2. Anton-Erxleben, K. & Carrasco, M. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nat. Rev. Neurosci.* 14, 188–200 (2013).

3. Moran, J. & Desimone, R. Selective attention gates visual processing in the extrastriate cortex. *Science* 229, 782–784 (1985).

4. Martinez-Trujillo, J. & Treue, S. Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron* 35, 365–370 (2002).

5. McAdams, C.J. & Maunsell, J.H. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19, 431–441 (1999).

6. Luck, S.J., Chelazzi, L., Hillyard, S.A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42 (1997).

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

8. Barun, J.K., Lau, B. & Salzman, C.D. Reward expectation differentially modulates behavioral and activity in visual area V4. *Nat. Neurosci.* 18, 1656–1663 (2015).

9. Mitchell, J.F., Sundberg, K.A. & Reynolds, J.H. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55, 131–141 (2007).

10. Teue, S. & Maunsell, J.H. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382, 539–541 (1996).

11. Teue, S. & Martinez Trujillo, J.C. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579 (1999).

12. McAdams, C.J. & Maunsell, J.H.R. Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron* 23, 765–773 (1999).

13. Martinez-Trujillo, J.C. & Teue, S. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* 14, 744–751 (2004).

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

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

16. Mitchell, J.F., Sundberg, K.A. & Reynolds, J.H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63, 879–888 (2009).

17. Herrera, J.L., Gieselmann, M.A., Sanaye, M. & Thiele, A. Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron* 78, 723–739 (2013).

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

19. Zénon, A. & Krauzlis, R.J. Attention deficits without cortical neuronal deficits. *Nature* 489, 434–437 (2012).

20. Poort, J. & Roelfsema, P.R. Noise correlations have little influence on the coding of selective attention in area V1. *Cereb. Cortex* 19, 543–553 (2009).

21. Lee, J. & Maunsell, J.H.R. A normalization model of attentional modulation of single unit responses. *PLoS One* 4, e6451 (2009).

22. Reynolds, J.H. & Heeger, D.J. The normalization model of attention. *Neuron* 61, 168–185 (2009).

23. Boynton, G.M. A framework for describing the effects of attention on visual responses. *Vision Res.* 49, 1129–1143 (2009).

24. Ni, A.M., Ray, S. & Maunsell, J.H.R. Tuned normalization explains the size of attention modulations. *Neuron* 73, 803–813 (2012).

25. Ghose, G.M. Attentional modulation of visual responses by flexible input gain. *J. Neurophysiol.* 101, 2089–2106 (2009).

26. Verhoef, B.-E. & Maunsell, J.H. Attention operates uniformly throughout the classical receptive field and the surround. *Elife* 5, e17256 (2016).

27. Carandini, M. & Heeger, D.J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62 (2011).

28. Tripp, B.P. Decorrelation of spiking variability and improved information transfer through feedforward divisive normalization. *Neural Comput.* 24, 867–894 (2012).

29. Snyder, A.C., Morais, M.J., Kohn, A. & Smith, M.A. Correlations in V1 are reduced by stimulation outside the receptive field. *J. Neurosci.* 34, 11222–11227 (2014).

30. Ruff, D.A., Alberts, J.J. & Cohen, M.R. Relating normalization to neuronal populations across cortical areas. *J. Neurophysiol.* 116, 1379–1386 (2016).

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

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

33. Ecker, A.S. et al. State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82, 235–248 (2014).
34. Goris, R.L.T., Movshon, J.A. & Simoncelli, E.P. Partitioning neuronal variability. Nat. Neurosci. 17, 858–865 (2014).
35. Lin, I.C., Okun, M., Carandini, M. & Harris, K.D. The nature of shared cortical variability. Neuron 87, 644–656 (2015).
36. Okun, M. et al. Diverse coupling of neurons to populations in sensory cortex. Nature 521, 511–515 (2015).
37. Heeger, D.J. Normalization of cell responses in cat striate cortex. Vis. Neurosci. 9, 181–197 (1992).
38. Cavanaugh, J.R., Bair, W. & Movshon, J.A. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. J. Neurophysiol. 88, 2530–2546 (2002).
39. Coen-Cagli, R., Kohn, A. & Schwartz, O. Flexible gating of contextual influences in natural vision. Nat. Neurosci. 18, 1648–1655 (2015).
40. Ohshiro, T., Angelaki, D.E. & DeAngelis, G.C. A normalization model of multisensory integration. Nat. Neurosci. 14, 775–782 (2011).

41. Cohen, M.R. & Maunsell, J.H.R. A neuronal population measure of attention predicts behavioral performance on individual trials. J. Neurosci. 30, 15241–15253 (2010).
42. Ecker, A.S., Denfield, G.H., Bethge, M. & Tolias, A.S. On the structure of neuronal population activity under fluctuations in attentional state. J. Neurosci. 36, 1775–1789 (2016).
43. Ditterich, J., Mazurek, M.E. & Shadlen, M.N. Microstimulation of visual cortex affects the speed of perceptual decisions. Nat. Neurosci. 6, 891–898 (2003).
44. Verhoef, B.E., Vogels, R. & Janssen, P. Inferotemporal cortex subserves three-dimensional structure categorization. Neuron 73, 171–182 (2012).
45. Rust, N.C., Mante, V., Simoncelli, E.P. & Movshon, J.A. How MT cells analyze the motion of visual patterns. Nat. Neurosci. 9, 1421–1431 (2006).
46. Olsen, S.R., Bhandawat, V. & Wilson, R.I. Divisive normalization in olfactory population codes. Neuron 66, 287–299 (2010).
ONLINE METHODS

Surgical procedures. Two male rhesus monkeys (M1 and M2, Macaca mulatta, both 9 kg, 7 and 10 years old) were pair-housed on a standard 12:12-h light-dark cycle and given food *ad libitum*. Before training, each animal was implanted with a head post. Following completion of the behavioral training (~7 months), we implanted a 10 × 10 array of microelectrodes into area V4 of the left cerebral hemisphere, on the prelunate gyrus. Before surgery, animals were given buprenorphine (5.0 μg/kg, intramuscular) and fluimixin (1.0 mg/kg, intramuscular) as analgesics and a prophylactic dose of an antibiotic (Baytril, 5.0 mg/kg, intramuscular). They were then sedated with ketamine (15 mg/kg, intramuscular) and xylazine (2 mg/kg, intramuscular) and given atropine (50 μg/kg, intramuscular) to reduce salivation. Anesthesia was administered again 1.5 h into surgery; buprenorphine and fluimixin were given for 48 h postoperatively. All procedures were approved by the Institutional Animal Care and Use Committee of Harvard Medical School.

Spatial attention task. As described in detail previously, we trained monkeys to perform a visual-detection task in which spatial attention was manipulated. Each trial started when the monkey fixated on a small spot in a virtual 1.5° square fixation window in the center of the video display for 240–700 ms. Eye movements were tracked using an infrared camera (EyeLink 1000), sampling binocularly at 500 Hz. The duration of the fixation period was randomly drawn from a uniform distribution. Following fixation, a sequence of stimuli was presented, in which each stimulus presentation lasted 200 ms and was separated from other presentations by 200–1,020 ms interstimulus intervals. The durations of the interstimulus intervals were randomly drawn from an exponential distribution (τ = 200 ms). During the interstimulus interval, only a gray screen with the fixation dot was shown. The stimulus presentations were short to prevent animals from adjusting their attention within a stimulus presentation in response to the number or orientation of stimuli presented.

On each trial, stimuli appeared at two locations near the RFs of neurons, but the two locations differed between blocks of trials. One stimulus location (the middle location; location 1 in Supplementary Fig. 1a,b) never varied, but in different blocks of trials the second stimulus location was offset either clockwise (location 2 in Supplementary Fig. 1a) or counterclockwise (location 3 in Supplementary Fig. 1b). All stimulus locations were equidistant from the fixation point, and stimulus locations 2 and 3 were equidistant from stimulus location 1.

On each stimulus presentation within a trial, we presented a total of one, two or no stimuli at the two stimulus locations near the neurons’ RFs. Each stimulus could be one of two orthogonal orientations. Each session, the stimulus orientation and location were optimized for a randomly selected unit, so that different orientations and locations were used across sessions. A representative set of nine possible stimulus combinations (for a particular orientation pair) is shown in Supplementary Fig. 1c. Using these different stimulus combinations, we used a normalization model to measure how the excitatory and suppressive contributions from different stimuli were combined into a neuronal response (see below).

Each stimulus location near the neurons’ RFs (stimulus locations 1–3 in Supplementary Fig. 1a,b) had a corresponding and equally eccentric stimulus location on the opposite side of the fixation point (for example, stimuli near Far in Fig. 1b and Supplementary Fig. 1a,b). As outlined below, we instructed monkeys to direct their attention to one stimulus location, either near or far from the RFs. Using this approach, we not only measured how attention modulated correlated neuronal responses when directed to different stimuli near the neurons’ RFs but also measured stimulus selectivity and suppression with attention directed far from the neurons’ RFs.

When Gabor pairs were presented near the neurons’ RFs, their centers were separated by a median of 2.3° (range: 1.6–4.8°) and always separated by at least 6 Gabor s.d. (mean Gabor σ = 0.45°; range, 0.17°–0.50°). With such interstimulus spacing, two stimuli could fall within the RFs of a V4 neuron. We recorded simultaneously from several units and therefore could not optimize the stimuli for most neurons. Depending on the precise size and locations of the neurons’ RF, the two stimuli could both fall in the cRF (cRF–cRF condition; Fig. 6) or one stimulus could fall in the cRF and the other in the surround (cRF–surround; Fig. 6; the condition with two stimuli in the surround did not elicit responses and was not further analyzed).

Subjects were required to detect a faint white spot (target; Supplementary Fig. 1e). The target appeared at one of the four stimulus locations (see above) during one stimulus presentation within a trial. The target never appeared on the first stimulus presentation of a trial but could occur with equal probability on any other stimulus presentation (range: 2–8). Two to five percent of the trials contained no target, and the monkey was rewarded for maintaining fixation until the trial ended. Targets were presented in the center of the Gabor stimuli to encourage the monkeys to confine their attention to a restricted part of visual space near the cued stimulus location.

Task difficulty was manipulated by varying the target strength, defined as the opacity of the target (range of alpha-transparency values: 0.06–0.28). During each session we used six different target strengths (Supplementary Fig. 1e). The monkey was rewarded with a drop of juice for making a saccade to the target location within 350 ms of its appearance.

Attention was cued to a single location throughout each block of ~150 trials. Before the start of each block, the monkey performed three to five instruction trials, in which stimuli were presented at a single (cued) location. The instruction trials cued the monkey to attend to that location during subsequent trials, in which stimuli could occur at all four locations. No spatial cuing was provided after the instruction trials were completed.

Within a block of trials, the target appeared at the cued location in 91% of the trials (valid trials; black circle in Supplementary Fig. 1d). In the remaining 9% of the trials (invalid trials) the target appeared at one of the three other (uncued) stimulus locations with equal probability (yellow and green circles in Supplementary Fig. 1d). We used a single target strength for all invalid trials, as this allowed us to obtain reliable estimates of behavior at the unattended locations despite the small number of invalid trials. Using invalid trials, we compared performances for stimuli at attended and at unattended locations. For both monkeys, the attention cue greatly affected behavioral performance in the task: targets were much more likely to be detected at a cued location than at an uncued location, even when the uncued location was adjacent to the cued location (Supplementary Fig. 1e).

Adaptation or repetition suppression/enhancement causes neuronal responses to a stimulus to decrease/enhance depending on the recent stimulus history. The magnitude of adaptation-related response changes may depend on the identity of the preceding stimulus47. Such stimulus-dependent adaptation can potentially bias spike-count correlations, which are calculated across different repetitions of a stimulus with possibly different preceding stimuli. We mitigated adaptation-related correlation biases by keeping the recent stimulus history constant within a recording session, i.e., by ensuring that each stimulus was always preceded by the same stimulus. For each daily recording session, we used a new and randomly ordered circular stimulus sequence (imagine placing the different stimuli on a rotating carousel). Each trial started at a random position within the sequence and subsequently the sequence progressed by the random number of stimulus presentations for that trial. This procedure was not employed during training sessions, and stimuli containing the target were not counted as part of the sequence.

Visual stimulation and recordings. Stimuli were presented on a gamma-corrected cathode-ray tube (CRT) display with a 100-Hz frame rate (1,024 × 768 pixels). Monkeys viewed the display from a distance of 57 cm. Stimuli consisted of full-contrast, achromatic, odd-symmetric static Gabor stimuli (0.6–2.2 cycles per degree; one spatial frequency per daily session) presented on a gray background (42 cd/m²), rendered online using custom software. The Gabor stimuli were truncated at 3 s.d. from their center.

We recorded neuronal activity using a 10 × 10 array of microelectrodes (Blackrock Microsystems; impedances: 0.3–1.2 MΩ at 1 kHz; 1-mm electrodes; 0.4 mm between adjacent electrodes), chronically implanted into area V4 of the left cerebral hemisphere of each monkey. The data presented here are from 130 daily sessions of recording (monkey M1; 52 sessions; monkey M2; 78 sessions).

At the beginning of each recording session, we mapped the RFs and optimized the stimulus parameters (position, orientation) for one randomly selected unit. The RFs of the units were located in the lower right quadrant at an average eccentricity of 3° for monkey M1 and 4° for monkey M2.

Statistics. We included only neuronal data from stimulus presentations from validly cued correct trials. We excluded invalidly cued trials, incorrect trials,
instruction trials, trials with no target, the first stimulus presentation of a trial (on which no target could occur) and stimulus presentations with a target. Units were included in the analyses if they responded significantly above baseline to any single Gabor presented at any stimulus location in the Attend Far condition (ANOVA; $\alpha = 0.05$). Responses in the Attend Far condition were obtained by averaging the firing rates from the conditions in which attention was directed to either of the two stimulus locations furthest away from the center of the neuron’s RF (Fig. 1b). Similar results were obtained for each monkey, so the data from both monkeys were combined.

We examined the relationship between normalization mechanisms and spike-count correlations. Normalization mechanisms determine how neurons combine the suppressive and excitatory contributions triggered by each stimulus into one response$^{27,38}$. Because we only observed neuronal responses, we needed a way to disentangle these different contributions to the neuronal response. We used a previously described divisive normalization model to estimate, for each stimulus, the strength of its excitatory and suppressive contribution to the neuronal response$^{26}$. This model successfully captures neuronal responses in all stimulus and attention conditions. Its basic form is given by:

$$R_{i,j} = \frac{L_i + L_j}{\alpha_i + \alpha_j + \sigma}$$ (1)

where $R_{i,j}$ is the neuronal response to a Gabor pair consisting of component Gabors 1 and 2. $L_i$ and $L_j$ are the excitatory contributions associated with each component Gabor. The $\alpha_i$ and $\alpha_j$ parameters determine the suppressive contribution of each component Gabor. Parameters $\alpha_i$ and $\alpha_j$ are each associated with one RF location and do not vary with the orientation of the stimuli shown at those locations. For simplicity, no contrast terms appear in equation (1) because the Gabors were always presented at full contrast; when a single stimulus is presented in isolation, the $L$ and $\alpha$ terms from the other stimuli are set to zero. Directing attention toward the first ($R_{i,\text{att}2}$; equation (2)) or second ($R_{j,\text{att}1}$; equation (3)) RF location has a multiplicative effect on the parameters corresponding to the attended RF location. This is described by the $\beta$ parameter in equations (2) and (3):

$$R_{i,\text{att}2} = \frac{\beta L_i + L_j}{\beta \alpha_i + \alpha_j + \sigma}$$ (2)

$$R_{j,\text{att}1} = \frac{L_i + \beta L_j}{\alpha_i + \beta \alpha_j + \sigma}$$ (3)

The model was fit to each unit’s responses to all stimulus conditions, including conditions with single Gabors or Gabor pairs near the RF and conditions in which attention was directed toward stimulus locations near the RF or far from it. All parameters were constrained to be non-negative. The model was fit by minimizing the sum of squared error using a simplex optimization algorithm (‘minisearch’ in Matlab; MathWorks). The model was fitted to 36 attention-and-stimulus combinations with ten parameters (see our previous study$^{26}$ for additional details). Across all units, the model’s median percentage of explained variance was 83%.

The model parameters provide information about each stimulus’ suppressive and excitatory contribution to a unit’s response. For each of the presented Gabor pairs, we quantified the relative excitatory and suppressive contribution of the component Gabors to each unit’s response, using two indices. First, we defined the selectivity of a unit for the two component Gabors of a Gabor pair as:

$$\text{Selectivity} = \frac{L_i - L_j}{L_i + L_j}$$

where Gabor 1 is the preferred stimulus, i.e., $L_1 > L_2$.

Similarly, we defined for each Gabor pair the nonpreferred suppression index as

$$\text{nonpreferred suppression} = \frac{\alpha_2}{\alpha_1 + \alpha_2}$$

where $\alpha_1$ and $\alpha_2$ correspond to the preferred and nonpreferred stimuli, respectively.

Selectivity and relative-suppression indices were computed for each unit ($n = 12,067$ units) and all different Gabor pairs (consisting of Gabors of different orientations and presented at different positions; Supplementary Fig. 1a–c).

To simplify the presentation of the results, in Figure 2 we excluded from analysis the 27% of unit pairs in which one unit received strong suppression from its preferred stimulus and the other unit received strong suppression from its nonpreferred stimulus. However, similar results were obtained for this subset of unit pairs (Supplementary Fig. 4).

Spike-count correlations were computed as Pearson correlation coefficients between the spike counts of pairs of units. For this purpose, spike counts were obtained from repeated presentations of the same stimulus in the same attention condition and were measured in the interval from 50 ms to 300 ms after stimulus onset.

Spike-count correlations were computed for pairs of units, where each unit of a pair had its own selectivity and relative suppression index. For each index, the two indices of a correlation pair were combined using the geometric mean of both units’ indices (for example, $\text{Selectivity}_{\text{pair}1,2} = \sqrt{\text{Selectivity}_{\text{unit1}} \times \text{Selectivity}_{\text{unit2}}}$). Figures 2 and 3 are based on these combined indices. The combined selectivity index was subsequently signed: positive selectivity indices correspond to pairs of units with the same stimulus preference (i.e., both units prefer the same component Gabor of a Gabor pair); negative selectivity indices correspond to pairs of units with opposite stimulus preferences (for example, unit 1 prefers Gabor 1 while unit 2 prefers Gabor 2 of a Gabor pair). Selectivity was computed per Gabor pair and could depend on either spatial or orientation selectivity. The nonpreferred suppression index ranges from 0 to 1. Values near 0 indicate that each unit of a correlation pair is weakly suppressed by its nonpreferred stimulus relative to suppression by its preferred stimulus. Values near 1 indicate that each unit of a correlation pair is strongly suppressed by its nonpreferred stimulus relative to the suppression by its preferred stimulus.

Similar results were obtained when equating the average response strength of pairs of neurons across conditions. For this purpose we matched the response-strength distributions (eight bins covering the range between the maximum and minimum response strength) between stimulus (Fig. 2e) or attention (Fig. 3d) conditions for each quadrant in the space spanned by selectivity and nonpreferred suppression. The quadrants are those used in Figures 2e, 3d and 4b. Here response strength is defined by the geometric mean of the average neuronal responses of the two units of a correlation pair. Response-strength distributions were matched by randomly removing spike-count correlations.

Due to the chronic nature of our recordings, it is possible that some units were resampled across days. Because we adjusted the orientations and locations of the stimuli each day for a randomly selected unit, any such resampling would have rarely involved identical stimulus configurations.

A stimulus location was considered within the cRF if the unit responded significantly (ANOVA; $\alpha = 0.05$) to any single stimulus (of either orientation) presented at that location, measured in the Attend Far condition. A stimulus location was considered to be within the surround of a unit if the unit did not increase its firing rate significantly to any single stimulus (of either orientation; $n > 36$ trials per stimulus) presented at that location, measured in the attend Far condition. Units for which a surround location was measured did respond significantly to at least one of the stimuli when it was presented inside the cRF instead of the surround.

The plots in Figures 2b–d, 3a–c and 4a were obtained using regularized (regularizing the gradient; smoothness $= 1$) bilinear interpolation on the observed spike-count correlations from all Gabor pairs and all unit pairs$^{49}$. The data covered virtually the entire space spanned by the nonpreferred suppression and selectivity indices (Supplementary Fig. 5).

No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to or larger than those reported in previous publications$^{14}$.

Where linear regression and ANOVA analyses where used, the distribution of the residuals closely approximated a normal distribution (i.e., comparing the empirical residual distribution to the best-fitted Gaussian distribution), but they did statistically deviate from normality (Kolmogorov-Smirnov test), as expected from the large sample sizes employed in the experiment. All statistical tests were two-tailed. Supplementary Figure 6 shows the box-and-whisker plots for Figures 2e, 3d and 4b.
Data collection but not analysis was performed blind to the conditions of the experiments. No animals were excluded from the analysis. A Supplementary Methods Checklist is available.

**Model.** The intuitive model in Figures 2f and 3e,f was formalized using a Poisson process with a rate governed by a system of coupled differential equations. Specifically, each neuron i generated spikes from a Poisson process with a random, time-varying mean spike rate λ(t). The mean spike rate of the neurons on a given trial evolved over time according to

\[ \tau \frac{d\lambda}{dt} = -\lambda + S(N\lambda + E) \]  

(4)

Here, \( \lambda \) is a vector containing the spike rates \( \lambda_i(t) \) of each neuron \( i \). \( \tau \) is a time constant (i.e., 50 ms), S is a sigmoidal transfer function (i.e., logistic function) stabilizing the model, \( N \) is a normalization matrix capturing the coupling of the spike rates of the individual neurons thereby inducing spike-count correlations between the neurons, and \( E \) is a vector containing the excitatory inputs \( E_i \) to each neuron.

In the steady state, \( \frac{d\lambda}{dt} = 0 \), this model is akin to a multivariate normalization model, \( \lambda = E(I - N)^{-1} \), where \((I - N)^{-1}\) functions as the normalizing denominator and I is the identity matrix; and we used the fact that most of the dynamics in \( \lambda \) occurs on the linear part of S, so that S can be ignored.

For each trial, the excitatory input was assumed to be randomly distributed according to a Gaussian distribution, i.e., \( E = N(\overline{E},\sigma_E^2) \). The \( \overline{E} \) vector contains the mean excitatory inputs to each neuron and depends on each neuron’s stimulus tuning (see below). The term \( \sigma_E \) is a covariance matrix containing all zeroes except on the diagonal where values are identical (i.e., 0.03). So, the excitatory inputs \( E_i \) across neurons are independent of each other, given the mean \( \overline{E} \).

Populations A and B contained neurons with different stimulus preferences. For the simulations, neurons in different populations had different orientation preferences (preferred orientation for Population A: 45°; for Population B: 135°) with circular Gaussian (von Mises distribution) tuning curves. Neurons within a population had the same tuning curves. As in the experiment described in the results, we presented the model with pairs of stimuli such that the two stimuli in the stimulus pair had orthogonal orientations. Depending on where the stimulus orientations fell on the tuning curves of each neuron, stimuli elicited different amounts of excitation. Each stimulus of a stimulus pair contributed excitation, which was summed to obtain \( \overline{E} \), the average response of neuron i to that stimulus pair. A range of stimulus conditions was employed, consisting of stimulus pairs of different orthogonal orientations (covering 0°–180°) and resulting in different excitations. Selectivity was computed for pairs of neurons according to the definition given above, using the difference in excitatory drive from each stimulus.

Attending to a stimulus corresponded to multiplying the excitatory drive associated with the attended stimulus with a constant gain factor (i.e., 1.1). Aside from this gain change, attention exerted no other direct influence on the model parameters, only indirect influences that followed from the model dynamics.

For simplicity the simulations described in the text were performed using four neurons, two in Population A and two in Population B, but the results do not depend on the number of neurons. Within a population, none of the neurons’ spike rates were coupled to each other. Between populations, neurons’ spike rates were negatively coupled to each other, thus creating mutual suppression between neuronal populations. So, the matrix N consisted of 0 (within-populations) or negative (between-populations) coupling values. A range of negative coupling values was employed (0 to –3.5) to simulate different amounts of nonpreferred suppression. Nonpreferred suppression indices were obtained by normalizing the coupling values to the range 0–1.

Once a stimulus is present, the evoked excitation \( E \) elevates the responses of neurons depending on their stimulus tuning. Thus the excitation \( E \) will increase the rate \( \lambda \) of the Poisson process (equation (4)). This rate is multiplied by the normalization matrix \( N \) in equation (4), resulting in an increased inhibition toward the neurons in the other population. The elements of the normalization matrix \( N \) are the only parameters that describe the interaction between neurons in different populations. However, the strength of the interaction crucially depends on the strength of the stimulus-driven excitation: increased (or decreased) excitation will result in more (or less) inhibition toward the other population.

Using this model, we simulated 10,000 trials per condition, i.e., for each combination of nonpreferred suppression, selectivity and position of attention. Equation (4) was numerically solved using the Runge-Kutta method. For each condition, we computed the spike-count correlation (Pearson correlation) based on the spikes simulated from the model in the interval from 50 ms to 300 ms after stimulus onset.

**Code availability.** Computer code for running the experiments is available from https://github.com/MaunsellLab/Lablib-Public-05-July-2016.git. Further code is available from the corresponding author upon reasonable request.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon reasonable request.

47. Verhoef, B.E., Kayaert, G., Franko, E., Vangeneugden, J. & Vogels, R. Stimulus similarity-contingent neural adaptation can be time and cortical area dependent. J. Neurosci. 28, 10631–10640 (2008).
48. D’Errico, J. Surface fitting using gridfit. Matlab Central File Exchange http://www.mathworks.com/matlabcentral/fileexchange/8998 (2005).