Neuron

Article

Laminar Organization of Attentional Modulation in Macaque Visual Area V4

Highlights

- Attention engages the V4 laminar circuit in selective ways
- Superficial broad-spiking cells show increased rate and reliability with attention
- Low-frequency coherence decreases with attention in superficial and input layers
- Correlations in V4 are highest in the input layer

Authors

Anirvan S. Nandy, Jonathan J. Nassi, John H. Reynolds

Correspondence

nandy@snl.salk.edu

In Brief

Attention is a critical component of perception. Here, Nandy et al. examine the laminar organization of attentional modulation in sensory cortex. They find layer- and cell-class-specific differences in the laminar cortical circuit in area V4.

Nandy et al., 2017, Neuron *93*, 235–246
January 4, 2017 © 2017 Elsevier Inc.
http://dx.doi.org/10.1016/j.neuron.2016.11.029

CellPress

Laminar Organization of Attentional Modulation in Macaque Visual Area V4

Anirvan S. Nandy,[1,3,*] Jonathan J. Nassi,[1,2] and John H. Reynolds[1]
[1]Systems Neurobiology Laboratories, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[2]Present address: Inscopix Inc., Palo Alto, CA 94303, USA
[3]Lead Contact
*Correspondence: nandy@snl.salk.edu

http://dx.doi.org/10.1016/j.neuron.2016.11.029

SUMMARY

Attention is critical to perception, serving to select behaviorally relevant information for privileged processing. To understand the neural mechanisms of attention, we must discern how attentional modulation varies by cell type and across cortical layers. Here, we test whether attention acts non-selectively across cortical layers or whether it engages the laminar circuit in specific and selective ways. We find layer- and cell-class-specific differences in several different forms of attentional modulation in area V4. Broad-spiking neurons in the superficial layers exhibit attention-mediated increases in firing rate and decreases in variability. Spike count correlations are highest in the input layer and attention serves to reduce these correlations. Superficial and input layer neurons exhibit attention-dependent decreases in low-frequency (<10 Hz) coherence, but deep layer neurons exhibit increases in coherence in the beta and gamma frequency ranges. Our study provides a template for attention-mediated laminar information processing that might be applicable across sensory modalities.

INTRODUCTION

Spatial attention is a critical component of our perceptual system. It mediates the enhancement of task-relevant signals, suppression of distractor signals, and reduction of noise among sensory neurons. Traditional single-unit electrophysiology has provided key insights into the probable mechanisms of attentional filtering of sensory information by measuring the attentional modulation of signals in sensory areas, such as visual area V4 (see Knudsen, 2007; Reynolds and Chelazzi, 2004 for review). These include modulation of mean firing rate (McAdams and Maunsell, 1999; Reynolds et al., 2000), increased reliability in the firing of individual neurons (Mitchell et al., 2007), and reduction in co-variability among pairs of neurons (Cohen and Maunsell, 2009; Mitchell et al., 2009), all of which are thought to improve the signal-to-noise ratio of neurons encoding the sensory stimulus. These changes in neuronal response are thought to result from feedback signals generated in attentional control centers (such as in the pre-frontal and parietal cortices; see Squire et al., 2013 for a review), which impinge upon the neural circuits of the sensory cortices.

The mammalian sensory cortex with its columnar organization (Mountcastle, 1997) is a six-layered (‘‘laminar’’) structure with a canonical circuit organization composed of excitatory and local inhibitory interneurons that have distinct patterns of projection within and between layers and to other cortical and sub-cortical areas (Callaway, 1998; Douglas and Martin, 2004, 2007). For example, input layer neurons project locally to superficial and deep layers; superficial layer neurons project to higher-order visual areas and also locally to superficial and deep layers; deep layer neurons project primarily to sub-cortical nuclei that are involved in motor control. This anatomical organization is a repeated circuit motif that is replicated throughout the sensory neocortex (see Harris and Mrsic-Flogel, 2013 for a review). Understanding the functional role of this circuit could provide a template for canonical information processing principles in the neocortex.

To understand the cortical mechanisms of attention, it is therefore critical to investigate the laminar organization of attentional modulation of sensory information. Previous attempts to record from cortical columns in V4 and obtain reliable layer estimates have proved to be challenging, since dorsal V4 straddles a narrow gyrus, with only a narrow strip of cortex (~5 mm) parallel to the calvarium (Gattass et al., 1988). We overcame this challenge by removing the opaque dura mater and replacing it with a transparent silicone-based artificial dura (Figure 1A). This allowed us to precisely target laminar probes at specific cortical sites under visual guidance through a microscope and thereby record neuronal signals with reliable estimates of laminar location.

Given that the cortical layers exhibit different projection patterns, it is important to know how attentional modulation varies across layers. One possibility is that attention serves to improve signals in superficial layer neurons that project to higher cortical areas. Alternatively, attentional modulation might be strongest among deep layer neurons that project to sub-cortical nuclei involved in motor control. A third possibility is that modulation could alter local processing: primarily modulating the input layers that project locally within the cortical column.

RESULTS

We recorded neuronal responses from well-isolated single units, multi-unit activity, and local field potentials (LFPs) using linear

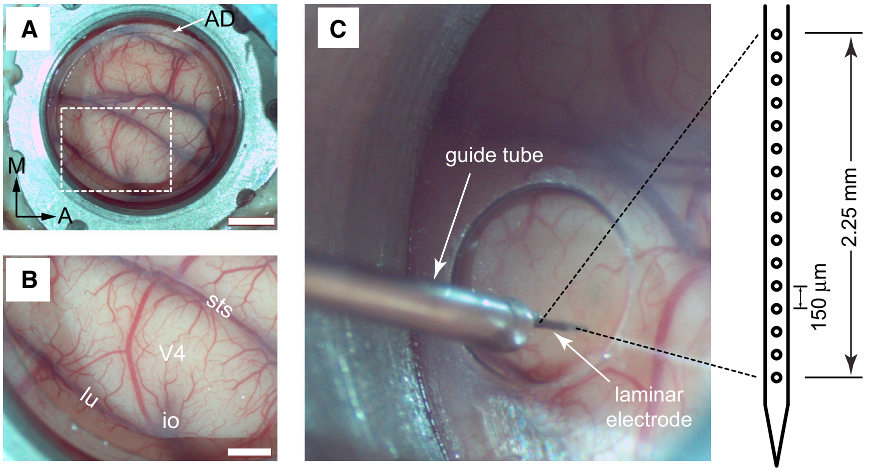


Figure 1. Laminar Electrophysiology through an Artificial Dura

(A) An artificial dura (AD) chamber is shown over dorsal V4 in the right hemisphere of monkey A. The native dura mater was resected and replaced with a silicone-based artificial dura, thereby providing an optically clear window into the cortex. Scale bar, 5 mm.

(B) An enlarged view of the boxed region in (A) clearly shows the sulci and the microvasculature. sts, superior temporal sulcus; lu, lunate sulcus; io, inferior occipital sulcus. Area V4 lies on the prelunate gyrus between the superior temporal and lunate sulci. Scale bar, 2 mm.

(C) Electrophysiology setup: a plastic stabilizer with a circular aperture is secured in place inside the chamber such that the aperture is centered over the pre-lunate gryus. A 16-channel linear array electrode (electrode spacing 150 µm) is positioned over the center of the gyrus and lowered into the cortex under microscopic guidance. The microvasculature pattern was used as a reference to target different cortical sites across recording sessions.

array electrodes oriented normal to the surface of area V4 of two rhesus macaques. This was achieved using artificial dura chambers (see Experimental Procedures; Figure 1). We took advantage of the optical clarity of the artificial dura to target 16-channel linear array electrodes to cortical sites near the center of the prelunate gyrus, where the cortex is maximally flat with respect to the calvarium. This resulted in penetrations that were perpendicular to the cortical surface, as indicated by excellent alignment of receptive fields through the entire depth of cortex (Figure 2D; Figure S1A) We used current source density (CSD) analysis (Mitzdorf, 1985) to estimate the boundaries between different cortical layers. The CSD, defined as the second spatial derivative of local field potential signals, produces a map of local current sinks and sources down the cortical depth as a function of time (Figure 2B; Figure S1B). This allows us to identify the superficial (Layers 1–3), input (Layer 4), and deep (Layers 5 and 6) layers of the cortex (Bollimunta et al., 2008; Schroeder and Lakatos, 2009; Schroeder et al., 1998). The location of the earliest current sink followed by a reversal to current source was identified as the input layer. The superficial and deep layers had complementary sink-source patterns (source followed by sink). Figure 2C shows line traces averaged across all channels classified by layer, illustrating the average sink-source pattern differences between layers. Isolated single units and multi-unit clusters (see Experimental Procedures) were then assigned to one of the three identified laminar compartments: superficial, input, or deep. The high quality of unit isolation allowed us to differentiate between narrow-spiking (putative interneurons) and broad-spiking (putative pyramidal) neurons among the majority of neurons recorded across all cortical layers (Figure 3; see Experimental Procedures) (Mitchell et al., 2007).

To investigate the laminar organization of attentional modulation, we recorded neuronal responses in two monkeys trained to perform an attention-demanding orientation change detection task (see Experimental Procedures; Figure 4). Attention was cued to one of two spatial locations. In the “attend-in” condition, the monkeys were instructed to covertly attend to a spatial location within the area of receptive field (RF) overlap of the neurons recorded throughout the V4 cortical column while maintaining fixation at a central fixation spot. In the “attend-away” condition, attention was cued to an equally eccentric location across the vertical meridian. During each trial, a sequence of oriented Gabor stimuli (baseline orientation optimized for each recording session) simultaneously flashed on and off at both spatial locations (200 ms on, variable 200–400 ms inter-stimulus intervals). At an unpredictable time (minimum 1 s, maximum 5 s), one of the two stimuli (95% probability at cued location; 5% probability at uncued location; “foil trials”) briefly changed in orientation (200 ms), and the monkey was rewarded for making a saccade to the location of orientation change (Figure 4A). If no change occurred within 5 s, the monkey was rewarded for holding fixation (“catch trial”). We controlled task difficulty by varying the degree of orientation change and thereby obtained behavioral performance curves (psychometric functions) for each recording session (Figure 4B). Impaired performance (Figure 4B, square symbol) and slower reaction times (Figure 4C, square symbol) were observed for the foil trials, indicating that the monkey was indeed using the spatial cue in performing the task.

While the monkey was performing the attention task, we simultaneously recorded neuronal data from V4 cortical columns (Figure S2). We analyzed the data from visually responsive single units (SUs, n = 274) and multi-unit activity (MUA, n = 217) (see Inclusion Criteria in Experimental Procedures; Figure S3; Table S1). To quantify the effects of attentional modulation, we focused our analyses on correct trials, where we had the best behavioral evidence that the monkey was deploying attention as instructed by the cue (see Experimental Procedures).

We find that attention positively modulates firing rate across all cortical layers (Figure 5A, SU and MUA, all distributions are significantly greater than zero, $p \ll 0.01$) but that the magnitude varies significantly by layer, with its greatest effects in the input layers. Attentional modulation of mean firing rate is significantly

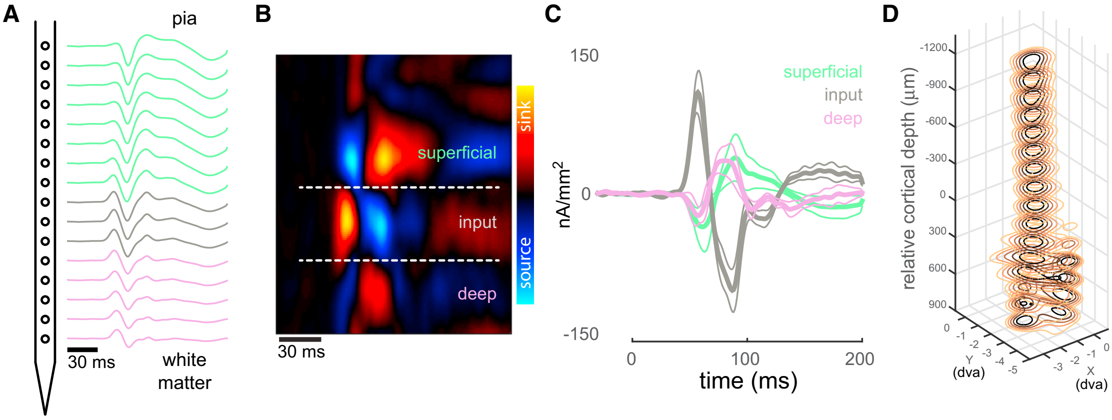


Figure 2. Laminar Identification and Receptive Fields in a V4 Cortical Column

(A) Stimulus-triggered local field potentials (LFPs) obtained by flashing 30 ms high-contrast ring stimuli in the receptive field of a V4 cortical column. LFP traces averaged across all stimulus repeats are shown color coded as being part of the superficial (green), input (gray), or deep (pink) layers. Layer assignment was done after current source density (CSD) analysis.

(B) Current source density calculated as the second spatial derivative of the stimulus-triggered LFPs and displayed as a colored map. The x axis represents time from stimulus onset; the y axis represents cortical depth oriented such that the pial surface is at the top and the white matter is at the bottom. Red hues represent current sink; blue hues represent current source. The input layer is identified at the first current sink followed by a reversal to current source. The superficial and deep layers have the opposite sink-source pattern. The CSD map has been spatially smoothed for visualization.

(C) CSD represented as line traces averaged across all channels that are part of a layer (same color coding convention as in A). Upward deflections represent current sink; downward deflections represent current source. Mean ± SEM.

(D) Stacked contour plots show spatial receptive fields (RFs) mapped along each contact point in the laminar probe. The spatial receptive fields were obtained by applying reverse correlation to the LFP power evoked by sparse pseudo-random sequences of Gabor stimuli. The RFs are well aligned, indicating perpendicular penetration down a cortical column. Zero depth represents the center of the input layer as estimated from the CSD.

larger in the input layer compared to the superficial and deep layers ($p_{input \leftrightarrow superficial} \ll 0.01, p_{input \leftrightarrow deep} = 0.01$). Next, we examined cell-class-specific rate modulation for broad- and narrow-spiking single units. A two-way ANOVA of the rate modulation indices with the factors "unit" (narrow and broad) and "layer" (superficial, input, and deep) revealed a significant main effect of both factors ($F_{unit} = 12.33, p = 0.0005$; $F_{layer} = 7.39, p = 0.008$), but no significant interactions between the factors. Further analysis reveals that the broad-spiking population is significantly modulated across all cortical layers ($p_{superficial} = 0.01, p_{input} \ll 0.01, p_{deep} \ll 0.01$), whereas the narrow-spiking population is significantly modulated in the input layer ($p_{superficial} = 0.25, p_{input} = 0.003, p_{deep} = 0.35$) (Figure 5B).

We calculated trial-to-trial variability in individual units by estimating the Fano factor (trial-to-trial spike count variance divided by the mean). We find that the broad-spiking population in the superficial layers exhibits significant reduction in Fano factor due to attention ($p < 0.01$; Figure 6, top panel). This reduction cannot be attributed to differences in firing rate between the two attention conditions (Figure S5C). The broad-spiking population in the input layer also exhibits a trend in reduction of variability due to attention, but the reduction is not statistically significant ($p = 0.06$; Figure 6, middle panel).

We next examined the laminar profile of spike count correlations among simultaneously recorded single units in V4. A three-way ANOVA of spike count correlations with the factors "attention" (attend-in and attend-away), "epoch" (inter-stimulus period and stimulus-evoked period), and "layer" (superficial, input, and deep) revealed a significant main effect of all three factors ($F_{attention} = 4.03, p = 0.04$; $F_{epoch} = 36.17, p \ll 0.01$; $F_{layer} = 32.85, p \ll 0.01$), but no significant interactions between the factors. Upon detailed analysis, we find two surprising results (Figure 7). First, correlated variability in V4 is highest in the input layer, resulting in an inverted "U" profile. This is true for both the inter-stimulus period ($p_{input \leftrightarrow superficial} \ll 0.01$, $p_{input \leftrightarrow deep} \ll 0.01$; Figure 7A, blue-bordered bars) and the stimulus-evoked period ($p_{input \leftrightarrow superficial} = 0.005, p_{input \leftrightarrow deep} \ll 0.01$; Figure 7B, blue-bordered bars) in the attend-away condition. The same inverted U profile holds for the pre-stimulus period in the attend-in condition ($p_{input \leftrightarrow superficial} \ll 0.01$, $p_{input \leftrightarrow deep} \ll 0.01$; Figure 7A, red-bordered bars). This is in contrast to the laminar profile of correlated variability in V1, where co-variability is highest in the superficial and deep layers and lowest in the input layer, resulting in a U-shaped profile across cortical layers (Hansen et al., 2012; Smith et al., 2013). If the laminar profile of correlated variability had a canonical organization in the visual cortex (Hansen et al., 2012), we would have expected to find a similar organization in V4. The second surprising finding is that attention reduces spike count correlations predominantly in the presence of a stimulus in the input layer ($p = 0.03$; Figure 7B). This is in contrast to expectations set up by prior studies which find strong evidence that deployment of attention reduces correlated variability among V4 neurons (Cohen and Maunsell, 2009; Mitchell et al., 2009). These two studies estimated on theoretical grounds that this decorrelation accounted for about 80% of the perceptual benefit due to attention. Since these signals presumably affect perception by transmitting improved signals to other brain areas, a natural expectation would be that decorrelation would be pronounced in the output layers of the cortex. The higher co-variability that

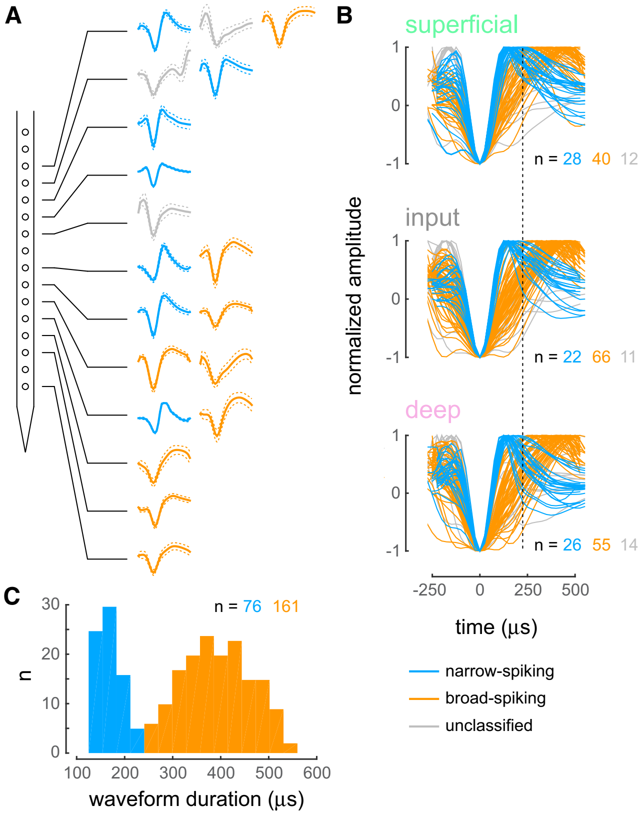


Figure 3. Single Units Recorded in V4 Cortical Columns

(A) Example recording session in monkey C depicting single-unit waveforms (mean ± SEM) isolated along the cortical column. Blue waveforms correspond to narrow-spiking putative interneurons, orange waveforms correspond to broad-spiking putative excitatory units, and gray waveforms are single units that could not be classified as narrow or broad (see Experimental Procedures).

(B) Average action potential waveforms for all 274 visually responsive single units in our population split by layer and neuronal class (blue, narrow; orange, broad; gray, unclassified). Waveform heights have been normalized for ease of comparison between the narrow- and broad-spiking units. The reference line (dotted) at 225 μs was used to classify the waveforms into the narrow and broad categories (see Experimental Procedures).

(C) Distribution of action potential waveform widths (trough-to-peak duration) for all narrow- and broad-spiking units collapsed across layers. The distribution is clearly bimodal ($p = 0.012$, Hartigan's dip test).

we observe in the input layer cannot be attributed to differences in firing rate across layers (Figure S5A) nor can it be attributed to differences in cortical distance between recorded pairs of neurons across layers (Figure S5B). Moreover, our results are consistent across monkeys (Figure S6).

To investigate the laminar aspects of co-variability further, we examined the spike-spike coherence (SSC) among pairs of simultaneously recorded single units. The SSC is a frequency-resolved measure of the degree to which the spiking activity of one unit fluctuates with that of a second unit. We chose SSC over spike-field coherence (SFC) analyses used in some other studies (Buffalo et al., 2011; Chalk et al., 2010; Fries et al., 2001) to allow direct comparison of the present results with those of Mitchell et al. (2009). Moreover, the short window of analyses did not lend itself well to evaluating SFC. Consistent with the modulation of spike count correlation, we find that the previously reported reduction in low-frequency SSC (Mitchell et al., 2009) due to attention is strongest in the input layer with significant, but weaker, modulation in superficial layers (Figure 7C). This attention-dependent modulation is better appreciated as a modulation index, where we see a larger negative modulation in the input layer (Figure 7D, middle panel; $p \ll 0.01$) compared to the superficial layers for frequencies less than 10 Hz (Figure 7D, top panel; $p = 0.05$). Interestingly, the deep layer exhibits a different pattern of modulation. Here, we see an increase in SSC due to attention for frequencies above 10 Hz (Figure 7C, lower panel), with positive modulation in the beta (15–25 Hz; $p \ll 0.01$) and gamma (>30 Hz; $p = 0.003$) frequency bands (Figure 7D, lower panel). That is, in the deep layers, attention *increases*, rather than decreases, coherence.

To gain insight into the possible neural mechanisms underlying the observed variation in spike count correlations across layers, we examined how correlations change as a function of the strength of excitatory (E) and inhibitory (I) connections in a conductance-based model of spiking neurons (see Experimental Procedures). This investigation was driven by the observation that spiking activity was highest in the input layer in both the broad- and narrow-spiking populations (Figure S4), suggesting a strongly coupled E-I network in that layer. We set up networks of E and I units that were mutually coupled (Figure 8A) and performed simulations that generated spiking activity in the network in response to a step input (Figure 8B). We calculated spike count correlations across repeated simulations of the network. We find that the strength of the feedback loop between the E and the I populations (W_{EI}, W_{IE}) is a critical factor in determining the strength of correlated activity in such a network. Holding the self-excitation and self-inhibition parameters fixed $(W_{EE} = 16, W_{II} = -1)$, we see that a strong inhibitory feedback loop (larger absolute values of W_{EI} and W_{IE}) leads to higher correlated activity in the network (Figure 8C). A strong inhibitory feedback loop acts as a common signal that drives the correlated activity in the network. This result is robust and holds for a wide range of values of W_{EE} and W_{II}, especially when inhibition is high in the network (low absolute values of W_{II}; Figure S7B). The pattern of results also holds for different network connection probabilities (Figure S7C), although spike count correlations decrease with increasing sparseness. This suggests a model in which the local circuit in the input layer in V4 is a tightly coupled E-I network leading to stronger correlated activity in this layer. On the other hand, the E-I local circuits in the superficial and deep layers are weakly coupled leading to weaker correlations in these layers (Figure 8D).

DISCUSSION

We find layer- and cell-class-specific differences in attentional modulation of mean firing rate, Fano factor, spike count correlations, and spike-spike coherence in area V4 of the macaque. Attention increases firing rates across all cortical layers, but this modulation is highest in the input layer. Broad-spiking units are significantly modulated across all layers; narrow-spiking units are significantly modulated in the input layer. Further, broad-spiking units in the superficial layer show a significant reduction in variability due to attention. Since the superficial layers project to downstream cortical areas, an elevation in firing rate and reduction in spiking variability among these units suggest their functional role in improving the signal quality of encoded signals under attention.

It is important to note that the method we have used to distinguish putative excitatory and inhibitory neurons (based on spike-waveform duration) does not unambiguously distinguish interneurons from pyramidal neurons. Although a majority of pyramidal cells do have broad action potential widths (McCormick et al., 1985; Nowak et al., 2003; Povysheva et al., 2006), there is evidence that pyramidal tract neurons in the motor cortex can fire “thin” spikes (Vigneswaran et al., 2011). Similarly, although parvalbumin-positive interneurons with the morphology of basket cells and chandelier cells, which make up 75% of interneurons in the primate, have narrow action potentials, the remaining 25% of interneurons have broad action potentials (Cauli et al., 1997; Connors and Gutnick, 1990; Kawaguchi, 1995; Kawaguchi and Kubota, 1997).

Consistent with earlier studies, we find that correlated variability among neurons in area V4 is weaker than in the primary visual cortex (Smith and Sommer, 2013) and that the presence of a stimulus reduces spike count correlations (Figures 7A and 7B) (Kohn and Smith, 2005). However, one of the surprising results of our study is the inverted-U laminar profile of spike count correlations in V4, which is in contrast with the U-shaped profile in V1 (Hansen et al., 2012; Smith et al., 2013). This suggests that there might not be a canonical laminar organization of correlated variability in the sensory neocortex. The higher levels of correlated activity in the input layer could be attributed to two factors. Part of the correlated activity could be inherited as common inputs from the supra-granular layers of V1. Neurons in the supra-granular layers of V1 have the highest levels of correlated variability (Hansen et al., 2012; Smith et al., 2013) and these layers include neurons that project to the parafoveal regions of V4 (Ungerleider et al., 2008) from which our data were collected. This higher level of correlated activity could also reflect the dynamics of a strongly coupled local E-I network in the input layer, as suggested by our data (Figure S4) and explored in our model (Figure 8; Figure S7). Our model suggests a possible active mechanism for reducing these spike count correlations in the superficial and deep layers by weakly coupled local E-I networks in these layers. According to this model, correlated variability results from local patterns of cortical circuit connectivity. In contrast to earlier work demonstrating extremely low correlations in recurrent network models where pre-synaptic activity is instantaneously integrated (Renart et al., 2010), our conductance-based model exhibits correlations that are consistent with our data.

Prior studies (Cohen and Maunsell, 2009; Mitchell et al., 2009) estimated on theoretical grounds that the reduction in correlated

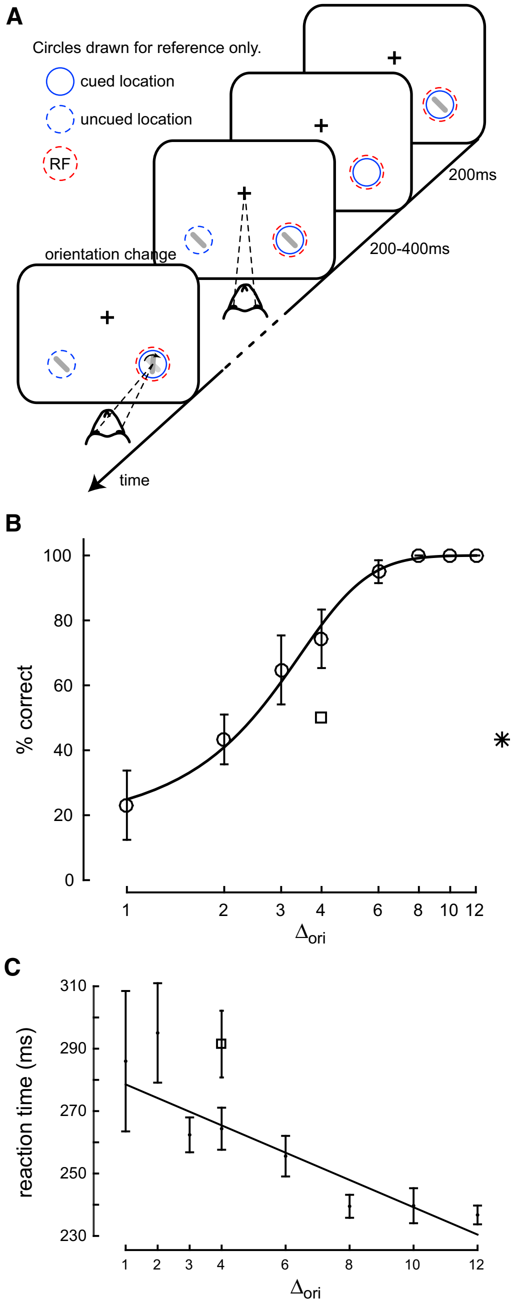


Figure 4. Attention Task and Behavior

(A) While the monkey maintained fixation, two oriented Gabor stimuli (schematized as oriented bars) flashed on and off simultaneously at two spatial locations: one at the RF overlap region of the recorded V4 column and the other at a location of equal eccentricity across the vertical meridian. The monkey was cued to covertly attend to one of the two locations. At an unpredictable time, one of the two stimuli changed in orientation. The monkey was rewarded for making a saccade to the location of orientation change at either location (95% probability of change at cued location; 5% probability at uncued location [foil trials]). If no change occurred (catch trials), the monkey was rewarded for maintaining fixation.

(B) Example behavioral session showing performance (hit rate) as a function of task difficulty (size of orientation change). Square symbol, foil trial performance. Asterisk, catch trial performance. Error bars are SD obtained by a jackknife procedure and corrected for the number of jackknives (20). The data have been fitted with a smooth logistic function.

(C) Reaction time as a function of task difficulty. The data have been fitted with a linear regression line. Performance is degraded and reaction times are higher for the foil trials, indicating that the animal was indeed deploying attention to the spatially cued location.

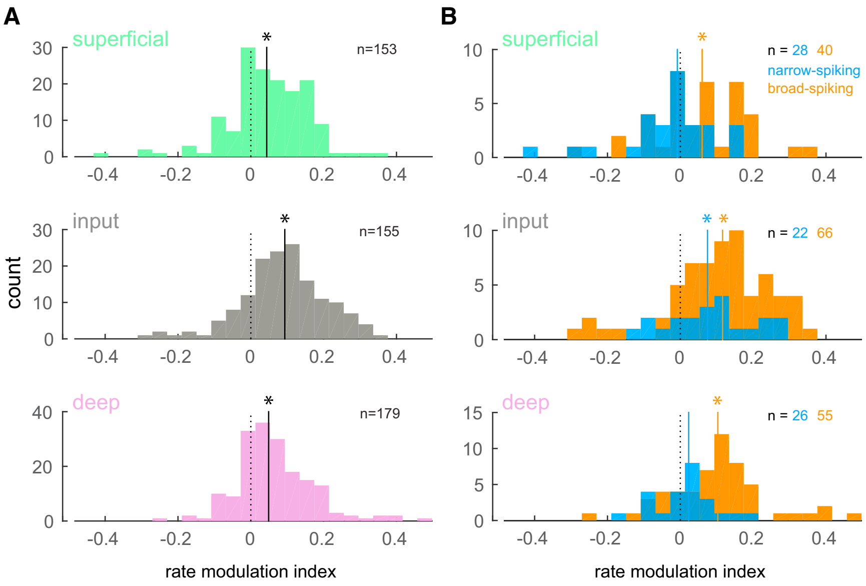


Figure 5. Attention-Mediated Firing Rate Modulation Is Strongest in the Input Layer

(A) Histograms depict the distribution of firing rate modulation indices (see Experimental Procedures) across layers for all units (single units and multi-unit activity) in a particular layer. Solid vertical lines depict the median values of each distribution. All distributions are significantly greater than zero ($p \ll 0.01$). Rate modulation for the input layer is significantly higher than the other two layers ($p_{input \leftrightarrow superficial} \ll 0.01, p_{input \leftrightarrow deep} = 0.01$).

(B) Rate modulation indices for the narrow- and broad-spiking single-unit populations (narrow, blue; broad, orange). The broad-spiking units exhibit significant positive rate modulation due to attention across all layers ($p_{superficial} = 0.01, p_{input} \ll 0.01, p_{deep} \ll 0.01$). The narrow-spiking units exhibit significant modulation in the input layer ($p_{superficial} = 0.25, p_{input} = 0.003, p_{deep} = 0.35$). Asterisks in (A) and (B) indicate statistically significant modulations.

variability accounted for a large fraction of the perceptual benefit due to attention. For these reductions to benefit decision making in other parts of the brain, this reduction in correlated variability would need to be localized to the output layers of the cortical column in V4. Neurons in the output layers of V4 transmit signals to downstream areas for further processing and decorrelation among these neurons would be expected to improve the signal to noise of the neural code. To the contrary, while we do find modest changes in gain and correlation in the superficial layers, we find that attention reduces correlated variability primarily in the input layer. This unexpected result forces us to reconsider the functional role of decorrelation in mediating the perceptual benefit due to attention. Rather than serving to directly improve the signal-to-noise ratio of the transmitted neural code, our results suggest that decorrelation in V4 is a local computation that serves to remove correlations from the inputs received from the earlier visual cortices. Further studies are needed to examine whether such a reduction plays a causal role in behavior.

A Framework of the V4 Laminar Circuit

Microstimulation of frontal eye field (FEF) neurons produces improvements in perception of targets appearing among distracters and also causes changes in V4 responses that mirror the effects of attention in V4 (Armstrong et al., 2006; Armstrong and Moore, 2007; Moore and Armstrong, 2003; Moore and Fallah, 2001). Together, these findings implicate FEF feedback to V4 as a key source of attentional modulation. These afferents originate mainly from superficial layer cells in FEF (Barone et al., 2000; Markov et al., 2011; Pouget et al., 2009). Our finding that attention significantly modulated the firing rate of broad-spiking neurons across all layers is consistent with anatomical evidence that FEF projections terminate in all cortical layers in V4 and that the labeled synapses of these projections are predominantly excitatory (Anderson et al., 2011). Direct anatomical pathways from FEF to inhibitory neurons in V4 are rare, making up only 4% of the total number of FEF synapses (Anderson et al., 2011). We thus speculate that the larger attentional modulation of putative parvalbumin-expressing interneurons (narrow-spiking units) as compared to broad-spiking units (Mitchell et al., 2009) and that we observe in our data in the input layer (Table S1; Figure S4) are likely mediated indirectly via excitatory neurons. These E-I interactions may mediate shifts in the local network to a decorrelated state (Figure 8C, arrow and dotted white rectangle). Such a shift, if most prominent in the input layer, would contribute to the suppression of input correlations to V4. Anderson et al. (2011) also found that individual FEF axons terminating in the V4 input layer formed sprays of collaterals, raising the possibility that these collaterals may play a prominent role in mediating attentional modulation. Finally, the differential modulation of coherence that we observe in our data is in agreement with the recent proposal that alpha and gamma band activity characterize feedback and feedforward processing, respectively (van Kerkoerle et al., 2014), although we observe elevated gamma band activity only in the deep layers.

Relationship to Prior Studies of Attention

Our data are consistent with prior studies in several aspects but differs from them in important aspects of the laminar organization. Consistent with Mitchell et al. (2007), we found that narrow-spiking neurons had higher firing rates than broad-spiking neurons, though this difference was only apparent in the input layer (compare Figure S4, middle panel, to Figure 2C of Mitchell et al., 2007). Data in Mitchell et al. (2007) were collected with single tungsten electrodes, which might have biased the data toward units with higher firing rates. Our more unbiased exploration with linear array electrodes reveal that overall firing rates are higher in the input layer with the rate differences between broad- and narrow-spiking units consistent with Mitchell et al. (2007). However, the superficial and deep layers with lower firing rates overall do not have this cell-class-specific difference (Figure S4). The reduction in variability in our data is modest compared to

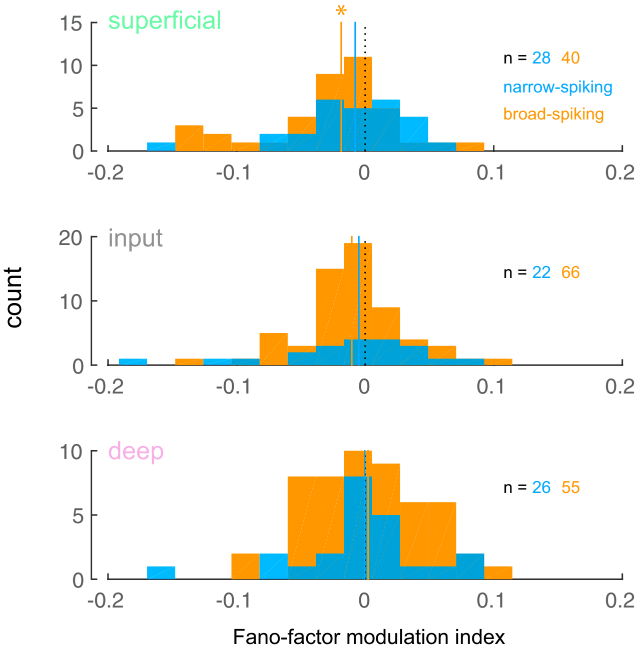


Figure 6. Attention Reduces Trial-to-Trial Variability among Broad-Spiking Cells in the Superficial Layer

Histograms depict the distribution of Fano factor modulation indices (see Experimental Procedures) across layers for the narrow- and broad-spiking single-unit populations (narrow, blue; broad, orange). The broad-spiking population in the superficial layer exhibits significant reduction in Fano factor (i.e., reduced trial-to-trial variability) due to attention ($p < 0.01$). Although the broad-spiking units in the input layer also exhibit a reduction in variability, this reduction is not significant ($p = 0.06$). Asterisk indicates statistically significant modulation.

Mitchell et al. (2007). This could be due to differences in task and measurement durations between the two studies—in the stimulus tracking task of Mitchell et al. (2007), the stimuli paused in the RF for 1 s, whereas in our task, the stimuli were briefly flashed for 200 ms. Our experimental paradigm was essentially similar to Cohen and Maunsell (2009), who used flashed gabors in an orientation discrimination task and reported a modest decrease in variability, consistent with the present findings.

Mitchell et al. (2009) found a strong reduction in low-frequency correlated variability due to attention. We find a similar reduction in the superficial layers and the input layer, with the effect being larger in the input layer. However, we also find an attention-mediated increase in coherence in the beta and gamma frequency range in the deep layers, consistent with earlier reports in V4 (Fries et al., 2001). Gamma band activity has been correlated with perceptual performance (Womelsdorf et al., 2006), while beta band activity has been implicated in motor behavior and top-down signaling (see Wang, 2010 for a review). Activity in both bands have also been implicated in modulating the efficiency of the oculomotor system underlying saccadic eye movements (Bartlett et al., 2011; Drewes and VanRullen, 2011).

Cohen and Maunsell (2009) report a decrease in spike count correlations with attention, which is consistent with our finding in the input layer. However, we do not find a significant attention-mediated change in spike count correlations in the superficial and deep layers. Cohen and Maunsell (2009) recorded with array electrodes (“UTAH” arrays) with 1-mm-long electrodes and might have sampled primarily from the input layer of V4 (Bjornsson et al., 2006).

Our coherence results are somewhat at odds with those reported by (Buffalo et al., 2011). They report a decrease in alpha band coherence and no appreciable change in gamma band coherence in the deep layers. We find decreases in low-frequency band coherence in the superficial and input layers and an increase in beta/gamma-band coherence in the deep layers. This discrepancy could be due to differences in experimental paradigms. In the Buffalo et al. (2011) study, monkeys were attending to a high-contrast, slowly drifting grating to detect a change in velocity. A second difference is the temporal window used for coherence analyses: Buffalo et al. (2011) used the sustained epoch with constant visual stimulation ignoring the first 300 ms after stimulus onset for calculating SFC; in our study, we used the stimulus-evoked period 60–260 ms after stimulus onset for calculating SSC among single units only. These differences, along with the fact that we used smaller, lower-contrast stimuli, are probable reasons why we do not see a peak in higher frequencies as in other studies (Buffalo et al., 2011; Chalk et al., 2010; Fries et al., 2001). A third key difference is in the methods used for layer estimation: layer assignment in Buffalo et al. (2011) was based upon estimated penetration depth of single tungsten electrodes, while in the current study, we have used the current source density profile to determine layer identity.

CONCLUSION

We find that different aspects of attention-mediated modulation of neuronal activity—increase in firing rate and decreases in variability and co-variability—are functionally segregated in different layers and among different cell classes within a cortical column in V4. This suggests a division of labor that serves to improve both the fidelity of transmitted information and the quality of local computation. Our study provides a template for how attention influences laminar information processing that might be applicable across sensory modalities.

EXPERIMENTAL PROCEDURES

Surgical Procedures

Surgical procedures have been described in detail previously (Nassi et al., 2015; Ruiz et al., 2013). In brief, an MRI-compatible, low-profile titanium chamber was placed over the pre-lunate gyrus on the basis of preoperative MRI imaging in two rhesus macaques (right hemisphere in monkey A, left hemisphere in monkey C). The native dura mater was then removed, and a silicone-based, optically clear artificial dura (AD) was inserted, resulting in an optical window over dorsal V4 (Figure 1). All procedures were approved by the Institutional Animal Care and Use Committee and conformed to NIH guidelines.

Electrophysiology

At the beginning of each recording session, a plastic insert with an opening for targeting electrodes was lowered into the chamber and secured. This served to stabilize the recording site against cardiac pulsations. Neurons were recorded from cortical columns in dorsal V4 using 16-channel linear array electrodes (“laminar probes”; Plexon, Plexon V-probe). The laminar probes were mounted on adjustable x-y stages attached to the recording chamber and

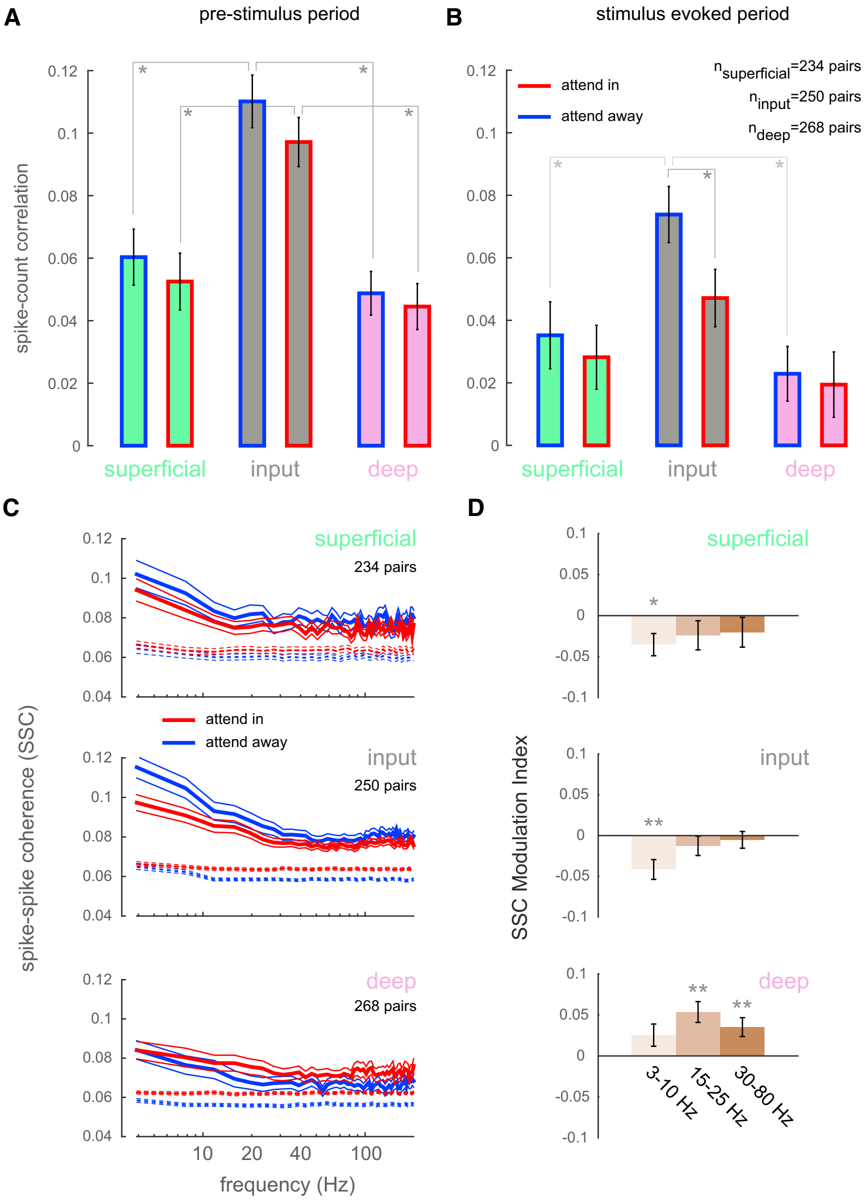


Figure 7. Attention Reduces Correlated Variability Most Prominently in the Input Layer

(A and B) Spike count correlations (r_{SC}) between pairs of simultaneously recorded single units within a layer are shown as a function of attention condition for two different temporal windows: an inter-stimulus period 200 ms before stimulus onset (A) and a stimulus-evoked period 60–260 ms after stimulus onset (B). A noteworthy aspect is that the laminar profile of r_{SC} in V4 has an inverted U shape, with the highest levels of correlated variability in the input layer. This is in contrast to the laminar profile reported in upstream V1, which has a U shape with correlations highest in superficial and deep layers (Hansen et al., 2012). Correlated variability is reduced in the stimulus-evoked period across all layers. Attention significantly reduces correlated variability in the stimulus-evoked period in the input layer $(p = 0.03)$.

(C) Spike-spike coherence (SSC)—a frequency resolved measure of the degree to which the spiking activity of one unit fluctuates with that of a second unit—among pairs of simultaneously recorded single units is plotted as a function of layer and attention condition. The dotted lines represent the baseline coherence that is expected solely due to trends in firing time locked to stimulus identity. Attention reduces low-frequency SSC in the superficial and input layers. The reduction is most prominent in the input layer. Attention increases high-frequency SSC in the deep layer. Mean ± SEM.

(D) Same data as in (C), but plotted as a modulation index: $(SSC_{in} - SSC_{away})/(SSC_{in} + SSC_{away})$ for the indicated frequency bands.

positioned over the center of the pre-lunate gyrus under visual guidance through a microscope (Zeiss) (Figure 1C). This ensured that the probes were maximally perpendicular to the surface of the cortex and thus had the best possible trajectory to make a normal penetration down a cortical column. Across recording sessions, the probes were positioned over different sites along the center of the gyrus in the parafoveal region of V4 with receptive field eccentricities between 2 and 7 degrees of visual angle (dvas). Care was taken to target cortical sites with no surface micro-vasculature and, in fact, the surface micro-vasculature was used as reference so that the same cortical site was not targeted across recording sessions. The probes were advanced using a hydraulic microdrive (Narishige) to first penetrate the AD and then through the cortex under microscopic visual guidance. Probes were advanced until the point that the top-most electrode (toward the pial surface) registered LFP signals. At this point, the probe was retracted by about 100–200 μm to ease the dimpling of the cortex due to the penetration. This procedure greatly increased the stability of the recordings and also increased the neuronal yield in the superficial electrodes.

The distance from the tip of the probes to the first electrode contact was either 300 μm or 700 μm. The inter-electrode distance was 150 μm, thus negating the possibility of recording the same neural spikes in adjacent recording channels. Neuronal signals were recorded extracellularly, filtered, and stored using the Multichannel Acquisition Processor system (Plexon). Neuronal signals were classified as either multiunit clusters or isolated single units using the Plexon Offline Sorter program. Single units were identified based on two criteria: (1) if they formed an identifiable cluster, separate from noise and other units, when projected into the principal components of waveforms recorded on that electrode and (2) if the inter-spike interval (ISI) distribution had a well-defined refractory period. Single units were classified as either narrow-spiking (putative inter-neurons) or broad-spiking (putative pyramidal cells) based on methods described in detail previously (Mitchell et al., 2007). Specifically, only units with waveforms having a clearly defined peak *preceded* by a trough were potential candidates. The distribution of trough-to-peak duration was clearly bimodal (Figure 3C, Hartigan's dip test, p = 0.012) (Hartigan and Hartigan, 1985). Units with trough-to-peak duration less than 225 μs were classified as narrow-spiking units; units with trough-to-peak duration greater than 225 μs were classified as broad-spiking units (Figure 3; blue, narrow; orange, broad). Units that did not have the prototypical biphasic shape were unclassified (Figure 3, gray waveforms) and only included in some of the analyses where neuronal class identity was not considered.

Data were collected over 32 sessions (23 sessions in monkey A and 9 in monkey C), yielding a total of 413 single units (128 narrow-spiking, 209

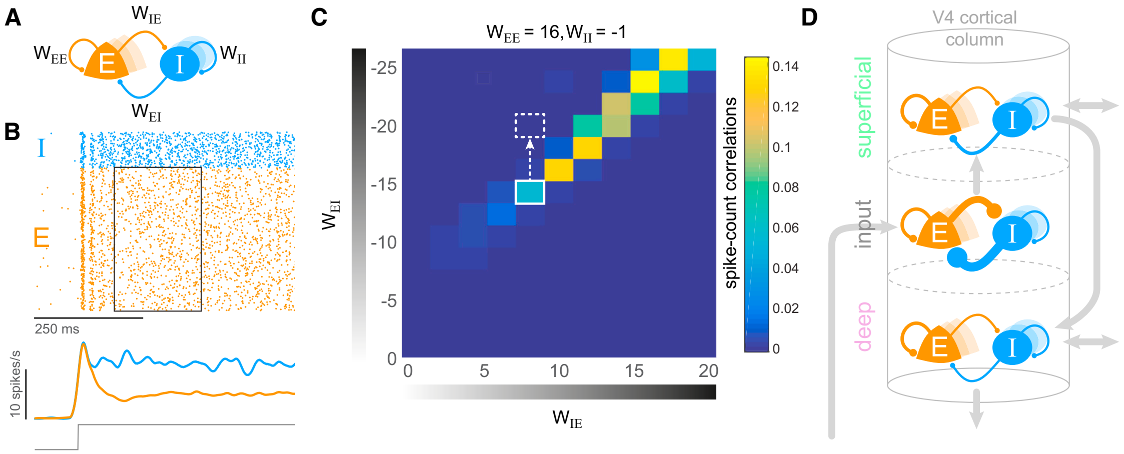


Figure 8. A Computational Model of Differentially Coupled E-I Networks Explains the Laminar Profile of Correlated Activity

(A) Schematic of a local conductance-based E-I network with mutually coupled excitatory (E) and inhibitory (I) units. W_{EE}, self-excitation among E units; W_{II}, self-inhibition among I units; W_{IE}, excitation provided by E units to I units; W_{EI}, inhibition provided by I units to E units.

(B) Simulation of a network of 800 E and 200 I units $(W_{EE} = 16, W_{II} = -1, W_{IE} = 8, W_{EI} = -14)$. The raster plot shows the spiking activity for all units in the model (blue, E; orange, I) in response to a step input. The middle traces show the population-spiking rate of the E and I units. The box indicates a 200 ms window used for calculating spike count correlations among the units.

(C) Spike count correlations as a function of different values of the E-I coupling parameters W_{IE} and W_{EI} while holding W_{EE} and W_{II} fixed. The shaded bars indicate the direction of stronger coupling. Spike count correlations were calculated across 2,000 repeats of identical stimulation to the network and averaged over four consecutive and non-overlapping 200 ms windows (first window indicated by box in B). The solid white rectangle corresponds to the model parameters used for the simulation in (B). A strong inhibitory feedback loop (larger absolute values of W_{EI} and W_{IE}) increases correlated activity. Conversely, weaker inhibitory cross-coupling de-correlates the network. The dotted white rectangle depicts a possible shift to a low-correlation regime due to attention (see Discussion).

(D) Proposed E-I local circuits in a V4 cortical column. The empirical data and the model simulations suggest a strongly coupled E-I network in the input layer leading to stronger correlations in this layer. Weakly coupled E-I networks in the superficial and deep layers lead to weak correlations in these layers. The widths of the links indicate coupling strength. The gray arrows indicate the primary information flow pathways in a canonical columnar circuit.

broad-spiking, and 76 unclassified) and 296 multi-unit clusters. Per session unit yield was considerably higher in monkey C compared to monkey A, resulting in a roughly equal contribution of both monkeys toward the population data.

Task, Stimuli, and Inclusion Criteria

Stimuli were presented on a computer monitor placed 57 cm from the eye. Eye position was continuously monitored with an infrared eye tracking system (ISCAN ETL-200). Trials were aborted if eye position deviated more than 1° (dva) from fixation. Experimental control was handled by NIMH Cortex software (http://www.cortex.salk.edu/).

Receptive Field Mapping

At the beginning of each recording session, neuronal RFs were mapped using subspace reverse correlation in which Gabor (eight orientations, 80% luminance contrast, spatial frequency 1.2 cycles/degree, Gaussian half-width 2°) or ring (80% luminance contrast) stimuli appeared at 60 Hz while monkeys maintained fixation. Each stimulus appeared at a random location selected from an 11 × 11 grid with 1° spacing in the appropriate visual quadrant. Spatial receptive maps were obtained by applying reverse correlation to the evoked LFP signal at each recording site. For each spatial location in the 11 × 11 grid, we calculated the time-averaged power in the stimulus-evoked LFP (0–200 ms after each stimulus flash) at each recording site. The resulting spatial map of LFP power was taken as the spatial RF at the recording site. For the purpose of visualization, the spatial RF maps were smoothed using spline interpolation and displayed as stacked contours plots of the smoothed maps (Figure 2D; Figure S1A). All RFs were in the lower visual quadrant (lower left in monkey A and lower right in monkey C) and with eccentricities between 2 and 7 dvas.

Current Source Density Mapping

In order to estimate the laminar identity of each recording channel, we used a CSD mapping procedure (Mitzdorf, 1985). Monkeys maintained fixation while 100% luminance contrast ring stimuli were flashed (30 ms), centered at the estimated RF overlap region across all channels. The size of the ring was scaled to about three-quarters of the estimated diameter of the RF. CSD was calculated as the second spatial derivative of the flash-triggered LFPs (Figure 2A). The resulting time-varying traces of current across the cortical layers can be visualized as CSD maps (Figure 2B; Figure S1B; maps have been spatially smoothed with a Gaussian kernel for aid in visualization). Red regions depict current sinks in the corresponding region of the cortical laminae; blue regions depict current sources. The input layer (Layer 4) was identified as the first current sink followed by a reversal to current source. The superficial (Layers 1–3) and deep (Layers 5 and 6) layers had opposite sink-source patterns. LFPs and spikes from the corresponding recording channels were then assigned to one of three layers: superficial, input, or deep.

Attention Task

In the main experiment, monkeys had to perform an attention-demanding orientation change detection task (Figure 4A). While the monkey maintained fixation, two achromatic Gabor stimuli (orientation optimized per recording session, spatial frequency 1.2 cycles/degree, 6 contrasts randomly chosen from a uniform distribution of luminance contrasts, $c = [10\%, 18\%, 26\%, 34\%, 42\%, \text{and } 50\%]$) were flashed on for 200 ms and off for a variable period chosen from a uniform distribution between 200 and 400 ms. One of the Gabors was flashed at the receptive field overlap region and the other at a location of equal eccentricity across the vertical meridian. At the beginning of a block of trials, the monkey was spatially cued ("instruction trials") to covertly attend to one of these two spatial locations. During these instruction trials, the stimuli were only flashed at the spatially cued location. At an unpredictable time (minimum 1 s, maximum 5 s, mean 3 s), one of the two stimuli changed in orientation. The monkey was rewarded for making a saccade to the location of orientation change. The monkey was rewarded for only those saccades where the saccade onset time was within a window of 100–400 ms after the onset of the orientation change. The orientation change occurred at the cued location with 95% probability and at the uncued location with 5% probability (foil trials). We controlled task difficulty by varying the degree of orientation change (Δ_{ori}), which was randomly chosen from one of the following: 1°, 2°, 3°, 4°, 6°, 8°, 10°, and 12°. The orientation change in the foil trials was fixed at 4°. These foil trials allowed us to assess the extent to which the monkey was using the spatial cue, with the expectation that there would be an impairment in performance (Figure 4B) and slower reaction times (Figure 4C) compared to the case in which the change occurred at the cued location. If no change occurred before 5 s, the monkey was rewarded for maintaining fixation (catch trials; 13% of trials). We will refer to all stimuli at the baseline orientation as "non-targets" and the stimulus flash with the orientation change as the "target."

Inclusion Criteria

Of the 413 single units and 296 multi-unit clusters, we included only a subset of neurons that were visually responsive for further analysis. For each neuron, we calculated its baseline firing rate for each attention condition (attend into RF [attend in or "IN"], attend away from RF [attend away or "AWAY"]) from a 200 ms window before a stimulus flash. We also calculated the neuron's contrast response function for both attention conditions (Figure S3). This was calculated as the firing rate over a window between 60 and 200 ms after stimulus onset and averaged across all stimulus flashes (restricted to non-targets) of a particular contrast separately for each attention condition. A neuron was considered visually responsive if any part of the contrast response curves exceeded the baseline rate by 4 SDs for both attention conditions. This left us with 274 single units (76 narrow-spiking, 161 broad-spiking, and 37 unclassified) and 217 multi-unit clusters for further analysis.

Data Analysis

Behavioral Analysis

For each orientation change condition Δ_{ori}, we calculated the hit rate as the ratio of the number of trials in which the monkey correctly identified the target with a saccade over the number of trials in which the target was presented. The hit rate as a function of Δ_{ori} yields a behavioral psychometric function (Figure 4B). Psychometric functions were fitted with a smooth logistic function (Palamedes MATLAB toolbox). Error bars were obtained by a jackknife procedure (20 jackknives, 5% of trials left out for each jackknife). Performance for the foil trials were calculated similarly as the hit rate for trials in which the orientation change occurred at the uncued location (Figure 4B, square symbol). Performance for the catch trials was calculated as the fraction of trials in which the monkey correctly held fixation for trials in which there was no orientation change (Figure 4B, asterisk).

Firing Rate Modulation Index

As in the contrast response functions, we calculated the average firing rate across all non-target flashes of a particular contrast (time window: 60–260 ms after stimulus onset) separately for each attention condition $(FR_{in}(c), FR_{away}(c))$. We only included stimulus flashes from correct trials (hit trials in which the monkey correctly detected a target or correct catch trials) for this and all subsequent analyses. The attentional rate modulation index (Figure 5) was calculated as the average modulation index across all contrast levels:

$$AMI_{rate} = \left\langle \frac{FR_{in}(c) - FR_{away}(c)}{FR_{in}(c) + FR_{away}(c)} \right\rangle$$

Fano Factor Modulation Index

Trial-to-trial variability was estimated by the Fano factor, which is the ratio of the variance of the spike counts across trials over the mean of the spike counts. The Fano factor was calculated over non-overlapping 20 ms time bins. The average Fano factor over a time window between 122 and 260 ms after non-target flash onset was calculated separately for each attention condition (FF_{in}, FF_{away}) to determine the attentional Fano factor modulation index (Figure 6):

$$FFMI_{rate} = \frac{FF_{in} - FF_{away}}{FF_{in} + FF_{away}}$$

We examined the degree to which neuronal firing was correlated among pairs of simultaneously recorded units in two different ways.

Spike Count Correlations

We calculated the Pearson correlation of spike counts across trials for every pair of simultaneously recorded single units where both units in the pair were in the same cortical layer (superficial, input, and deep). In order to remove the influence of confounding variables, like stimulus strength, we Z scored spike counts using the mean and SD for repetitions of each stimulus type. Ordered pairs of Z scored spike counts were collapsed across contrast conditions, and the Pearson correlation was calculated from these ordered pairs. This was done for each attention condition and also for two different counting windows: an inter-stimulus 200 ms period before non-target flash onset (Figure 7A) and a stimulus-evoked period between 60 and 260 ms after non-target flash onset (Figure 7B). We only considered those pre-stimulus periods where the inter-stimulus interval was greater than 500 ms (in other words, the interval between onset of the stimulus and the offset of the previous stimulus was greater than 300 ms), so as to minimize artifacts due to stimulus offset.

Spike-Spike Coherence

We computed the coherence between simultaneously recorded, single-unit pairs within a cortical layer using multi-taper methods (Mitra and Pesaran, 1999) over the 200 ms window between 60 and 260 ms after non-target flash onset. Spike trains were tapered with a single Slepian taper, giving an effective smoothing of 5 Hz for the 200 ms window (NW = 1, K = 1). Magnitude of coherence estimates depend on the number of spikes used to create the estimates (Zeitler et al., 2006). To control for differences in firing rate, we adopted a rate-matching procedure similar to Mitchell et al. (2009). In order to obtain a baseline for the coherence expected solely due to trends in firing time-locked to stimulus onset, we also computed coherence in which trial identities were randomly shuffled (Figure 7C).

Computational Model

We set up a conductance-based model of N_{E} excitatory and N_{I} inhibitory neurons with all-to-all connectivity with the following synaptic weights (Figure 8A):

$$\text{E to E}: \ w_{\mathrm{EE}} = \frac{W_{\mathrm{EE}}}{N_{\mathrm{E}}}; \text{I to I}: \ w_{\mathrm{II}} = \frac{W_{\mathrm{II}}}{N_{\mathrm{I}}}; \text{E to I}: \ w_{\mathrm{IE}} = \frac{W_{\mathrm{IE}}}{N_{\mathrm{E}}}; \text{I to E}: \ w_{\mathrm{EI}} = \frac{W_{\mathrm{EI}}}{N_{\mathrm{I}}}$$

We simulated models of $N_{\mathrm{E}} = 800$ excitatory and $N_{\mathrm{I}} = 200$ inhibitory spiking units. The spiking units were modeled as Izhikevich neurons (Izhikevich, 2003) with the following dynamics:

$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + I$$

$$\frac{du}{dt} = a(bv - u)$$

$$\text{if } v = 30mV, \text{ then } v \leftarrow c \text{ and } u \leftarrow u + d$$

v is the membrane potential of the neuron and u is a membrane-recovery variable. I is the current input to the neuron (synaptic and injected DC currents). The parameters a, b, c, and d determine intrinsic firing patterns and were chosen as follows:

$$\text{Excitatory units(regular spiking)}: \ a = 0.02, b = 0.2, c = -65, d = 8$$

$$\text{Inhibitory units(fast spiking)}: \ a = 0.1, b = 0.2, c = -65, d = 2$$

Presynaptic spikes from excitatory units generated fast (AMPA) and slow (NMDA) synaptic currents, while presynaptic spikes from inhibitory units generated fast GABA currents:

$$I_{syn} = \sum\nolimits_i g_{\mathrm{AMPA}}(t)(v(t) - V_{\mathrm{AMPA}}) + \sum\nolimits_j g_{\mathrm{NMDA}}(t)(v(t) - V_{\mathrm{NMDA}}) + \sum\nolimits_k g_{\mathrm{GABA}}(t)(v(t) - V_{\mathrm{GABA}})$$

where $V_{\mathrm{AMPA}} = 0, V_{\mathrm{NMDA}} = 0,$ and $V_{\mathrm{GABA}} = -70$ are the respective reversal potentials (mV). The synaptic time courses $g(t)$ were modeled as a difference of exponentials (Figure S7A):

$$g(t) = \frac{1}{\tau_d - \tau_r}\left[exp\left(-\frac{t - \tau_l}{\tau_d}\right) - exp\left(-\frac{t - \tau_l}{\tau_r}\right)\right]$$

where $\tau_l, \tau_r,$ and τ_d are the latency, rise, and decay time constants with the following parameter values (Brunel and Wang, 2003): AMPA: τ_l = 1 ms, τ_r = 0.5 ms, τ_d = 2 ms; NMDA: τ_l = 1 ms, τ_r = 2 ms, τ_d = 80 ms; GABA: τ_l = 1 ms, τ_r = 0.5 ms, τ_d = 5 ms. The NMDA to AMPA ratio was chosen as 0.45 (Myme et al., 2003).

The network was stimulated by a DC step current $(I_{DC} = 4)$ of duration 1.5 s (Figure 8B). Synaptic noise was simulated by drawing from a normal distribution $(I_{syn-noise} \sim \mathcal{N}(\mu = 0, \sigma = 3))$. We calculated spike count correlations from *Z* scored spike counts within 200 ms counting windows after the initial transient response (results were averaged across four such consecutive non-overlapping windows) across 2,000 repeats of the stimulation. Spike count correlations for different choices of network parameters $(W_{EE}, W_{II}, W_{EI}, W_{IE})$ were reported as the average across all possible pairs of excitatory spiking units (Figure 8C; Figure S7).

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2016.11.029.

AUTHOR CONTRIBUTIONS

A.S.N., J.J.N., and J.H.R. designed the experiments. A.S.N. collected and analyzed the data. A.S.N. developed the computational model. A.S.N., J.H.R., and J.J.N. wrote the manuscript.

ACKNOWLEDGMENTS

This research was supported by NIH R01 EY021827 to J.H.R. and A.S.N., the Gatsby Charitable Foundation to J.H.R., fellowship from the NIH T32 EY020503 training grant and the NARSAD Young Investigator Grant to A.S.N., The Salk Institute Excellerators Fellowship Program and the NARSAD Young Investigator Grant to J.J.N., and a NEI core grant for vision research P30 EY019005 to The Salk Institute. We would like to thank Ed Callaway, Amanda Casale, Bryan Hansen, Monika Jadi, and Jude Mitchell for helpful discussion; Rob Teeuwen for assistance with animal training; and Catherine Williams and Mat LeBlanc for excellent animal care. We would like to dedicate this to the memory of John C. Curtis.

Received: August 1, 2016
Revised: October 20, 2016
Accepted: November 8, 2016
Published: December 15, 2016

REFERENCES

Anderson, J.C., Kennedy, H., and Martin, K.A.C. (2011). Pathways of attention: synaptic relationships of frontal eye field to V4, lateral intraparietal cortex, and area 46 in macaque monkey. J. Neurosci. *31*, 10872–10881.

Armstrong, K.M., and Moore, T. (2007). Rapid enhancement of visual cortical response discriminability by microstimulation of the frontal eye field. Proc. Natl. Acad. Sci. USA *104*, 9499–9504.

Armstrong, K.M., Fitzgerald, J.K., and Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. Neuron *50*, 791–798.

Barone, P., Batardiere, A., Knoblauch, K., and Kennedy, H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. J. Neurosci. *20*, 3263–3281.

Bartlett, A.M., Ovaysikia, S., Logothetis, N.K., and Hoffman, K.L. (2011). Saccades during object viewing modulate oscillatory phase in the superior temporal sulcus. J. Neurosci. *31*, 18423–18432.

Bjornsson, C.S., Oh, S.J., Al-Kofahi, Y.A., Lim, Y.J., Smith, K.L., Turner, J.N., De, S., Roysam, B., Shain, W., and Kim, S.J. (2006). Effects of insertion conditions on tissue strain and vascular damage during neuroprosthetic device insertion. J. Neural Eng. *3*, 196–207.

Bollimunta, A., Chen, Y., Schroeder, C.E., and Ding, M. (2008). Neuronal mechanisms of cortical alpha oscillations in awake-behaving macaques. J. Neurosci. *28*, 9976–9988.

Brunel, N., and Wang, X.-J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. J. Neurophysiol. *90*, 415–430.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., and Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. Proc. Natl. Acad. Sci. USA *108*, 11262–11267.

Callaway, E.M. (1998). Local circuits in primary visual cortex of the macaque monkey. Annu. Rev. Neurosci. *21*, 47–74.

Cauli, B., Audinat, E., Lambolez, B., Angulo, M.C., Ropert, N., Tsuzuki, K., Hestrin, S., and Rossier, J. (1997). Molecular and physiological diversity of cortical nonpyramidal cells. J. Neurosci. *17*, 3894–3906.

Chalk, M., Herrero, J.L., Gieselmann, M.A., Delicato, L.S., Gotthardt, S., and Thiele, A. (2010). Attention reduces stimulus-driven gamma frequency oscillations and spike field coherence in V1. Neuron *66*, 114–125.

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

Connors, B.W., and Gutnick, M.J. (1990). Intrinsic firing patterns of diverse neocortical neurons. Trends Neurosci. *13*, 99–104.

Douglas, R.J., and Martin, K.A.C. (2004). Neuronal circuits of the neocortex. Annu. Rev. Neurosci. *27*, 419–451.

Douglas, R.J., and Martin, K.A.C. (2007). Mapping the matrix: the ways of neocortex. Neuron *56*, 226–238.

Drewes, J., and VanRullen, R. (2011). This is the rhythm of your eyes: the phase of ongoing electroencephalogram oscillations modulates saccadic reaction time. J. Neurosci. *31*, 4698–4708.

Fries, P., Reynolds, J.H., Rorie, A.E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science *291*, 1560–1563.

Gattass, R., Sousa, A.P., and Gross, C.G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. J. Neurosci. *8*, 1831–1845.

Hansen, B.J., Chelaru, M.I., and Dragoi, V. (2012). Correlated variability in laminar cortical circuits. Neuron *76*, 590–602.

Harris, K.D., and Mrsic-Flogel, T.D. (2013). Cortical connectivity and sensory coding. Nature *503*, 51–58.

Hartigan, J.A., and Hartigan, P.M. (1985). The dip test of unimodality. Ann. Stat. *13*, 70–84.

Izhikevich, E.M. (2003). Simple model of spiking neurons. IEEE Trans. Neural Netw. *14*, 1569–1572.

Kawaguchi, Y. (1995). Physiological subgroups of nonpyramidal cells with specific morphological characteristics in layer II/III of rat frontal cortex. J. Neurosci. *15*, 2638–2655.

Kawaguchi, Y., and Kubota, Y. (1997). GABAergic cell subtypes and their synaptic connections in rat frontal cortex. Cereb. Cortex *7*, 476–486.

Knudsen, E.I. (2007). Fundamental components of attention. Annu. Rev. Neurosci. *30*, 57–78.

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

Markov, N.T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M.A., Giroud, P., Ercsey-Ravasz, M., Pilaz, L.J., et al. (2011). Weight consistency specifies regularities of macaque cortical networks. Cereb. Cortex *21*, 1254–1272.

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

McCormick, D.A., Connors, B.W., Lighthall, J.W., and Prince, D.A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. J. Neurophysiol. *54*, 782–806.

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

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

Mitra, P.P., and Pesaran, B. (1999). Analysis of dynamic brain imaging data. Biophys. J. *76*, 691–708.

Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol. Rev. *65*, 37–100.

Moore, T., and Armstrong, K.M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. Nature *421*, 370–373.

Moore, T., and Fallah, M. (2001). Control of eye movements and spatial attention. Proc. Natl. Acad. Sci. USA *98*, 1273–1276.

Mountcastle, V.B. (1997). The columnar organization of the neocortex. Brain *120*, 701–722.

Myme, C.I.O., Sugino, K., Turrigiano, G.G., and Nelson, S.B. (2003). The NMDA-to-AMPA ratio at synapses onto layer 2/3 pyramidal neurons is conserved across prefrontal and visual cortices. J. Neurophysiol. *90*, 771–779.

Nassi, J.J., Avery, M.C., Cetin, A.H., Roe, A.W., and Reynolds, J.H. (2015). Optogenetic activation of normalization in alert macaque visual cortex. Neuron *86*, 1504–1517.

Nowak, L.G., Azouz, R., Sanchez-Vives, M.V., Gray, C.M., and McCormick, D.A. (2003). Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. J. Neurophysiol. *89*, 1541–1566.

Pouget, P., Stepniewska, I., Crowder, E.A., Leslie, M.W., Emeric, E.E., Nelson, M.J., and Schall, J.D. (2009). Visual and motor connectivity and the distribution of calcium-binding proteins in macaque frontal eye field: implications for saccade target selection. Front. Neuroanat. *3*, 2.

Povysheva, N.V., Gonzalez-Burgos, G., Zaitsev, A.V., Kröner, S., Barrionuevo, G., Lewis, D.A., and Krimer, L.S. (2006). Properties of excitatory synaptic responses in fast-spiking interneurons and pyramidal cells from monkey and rat prefrontal cortex. Cereb. Cortex *16*, 541–552.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K.D. (2010). The asynchronous state in cortical circuits. Science *327*, 587–590.

Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. Annu. Rev. Neurosci. *27*, 611–647.

Reynolds, J.H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. Neuron *26*, 703–714.

Ruiz, O., Lustig, B.R., Nassi, J.J., Cetin, A., Reynolds, J.H., Albright, T.D., Callaway, E.M., Stoner, G.R., and Roe, A.W. (2013). Optogenetics through windows on the brain in the nonhuman primate. J. Neurophysiol. *110*, 1455–1467.

Schroeder, C.E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. *32*, 9–18.

Schroeder, C.E., Mehta, A.D., and Givre, S.J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb. Cortex *8*, 575–592.

Smith, M.A., and Sommer, M.A. (2013). Spatial and temporal scales of neuronal correlation in visual area V4. J. Neurosci. *33*, 5422–5432.

Smith, M.A., Jia, X., Zandvakili, A., and Kohn, A. (2013). Laminar dependence of neuronal correlations in visual cortex. J. Neurophysiol. *109*, 940–947.

Squire, R.F., Noudoost, B., Schafer, R.J., and Moore, T. (2013). Prefrontal contributions to visual selective attention. Annu. Rev. Neurosci. *36*, 451–466.

Ungerleider, L.G., Galkin, T.W., Desimone, R., and Gattass, R. (2008). Cortical connections of area V4 in the macaque. Cereb. Cortex *18*, 477–499.

van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., and Roelfsema, P.R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc. Natl. Acad. Sci. USA *111*, 14332–14341.

Vigneswaran, G., Kraskov, A., and Lemon, R.N. (2011). Large identified pyramidal cells in macaque motor and premotor cortex exhibit “thin spikes”: implications for cell type classification. J. Neurosci. *31*, 14235–14242.

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. Physiol. Rev. *90*, 1195–1268.

Womelsdorf, T., Fries, P., Mitra, P.P., and Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. Nature *439*, 733–736.

Zeitler, M., Fries, P., and Gielen, S. (2006). Assessing neuronal coherence with single-unit, multi-unit, and local field potentials. Neural Comput. *18*, 2256–2281.

Neuron, Volume *93*

Supplemental Information

Laminar Organization of Attentional Modulation in Macaque Visual Area V4

Anirvan S. Nandy, Jonathan J. Nassi, and John H. Reynolds

Supplementary Information: Laminar organization of attentional modulation in macaque visual area V4

Anirvan S. Nandy[1], Jonathan J. Nassi[1,2] & John H. Reynolds[1]

[1]Systems Neurobiology Laboratories

[2]Present address: Inscopix Inc., Palo Alto, CA 94303

The Salk Institute for Biological Studies, La Jolla, CA 92037

Corresponding author: A.S.N. (nandy@snl.salk.edu)

SUPPLEMENTARY FIGURE LEGENDS

Supp Fig 1. Example receptive fields and CSD maps. Related to Fig 2. (**A**) Stacked contour plots (same format as Fig 2D) show spatial receptive fields along V4 cortical columns in three different recording sessions. Zero depth represents the center of the input layer as estimated from the CSD. (**B**) Current source density maps for three example recording sessions (same format as in Fig 2B). The reference bar represents the duration of the 30ms flashes used to obtain the CSD maps. Red hues=current sink; Blue hues=current source.

Supp Fig 2. Example trial in the attention experiment. Related to Figs 4,5. An example trial with neuronal signals is shown in the attend-in condition. The time axis is referenced to the appearance of the fixation spot. Shown are the eye-position traces (vertical and horizontal), LFPs, spikes (vertical ticks; orange=broad-spiking single unit, gray=MUA) and population PSTH. The gray boxes represent stimulus presentation epochs. In this particular trial, a series of 6 non-target flashes were followed by a target stimulus flash after which the monkey correctly detected the orientation change (exaggerated here for illustration) by making a saccade to the target.

Supp Fig 3. Contrast response functions used to determine visual responsiveness of identified units. Related to Fig 5. Contrast response functions – spike-rate (spikes/s) as a function of stimulus contrast – are shown for 20 units (16 single units, 4 multi-unit clusters) identified in a single recording session in Monkey A. The dotted lines represent background firing-rate. The dashed lines are 4 standard deviations above baseline. A unit was considered as visually responsive, if the contrast response functions exceeded this threshold in both the attention conditions. Single units are identified as either broad- or narrow-spiking; multi-unit clusters are indicated as MUA. Units are arranged from most superficial (u1) to the most deep (u20). Also indicated are the layer assignments of the units.

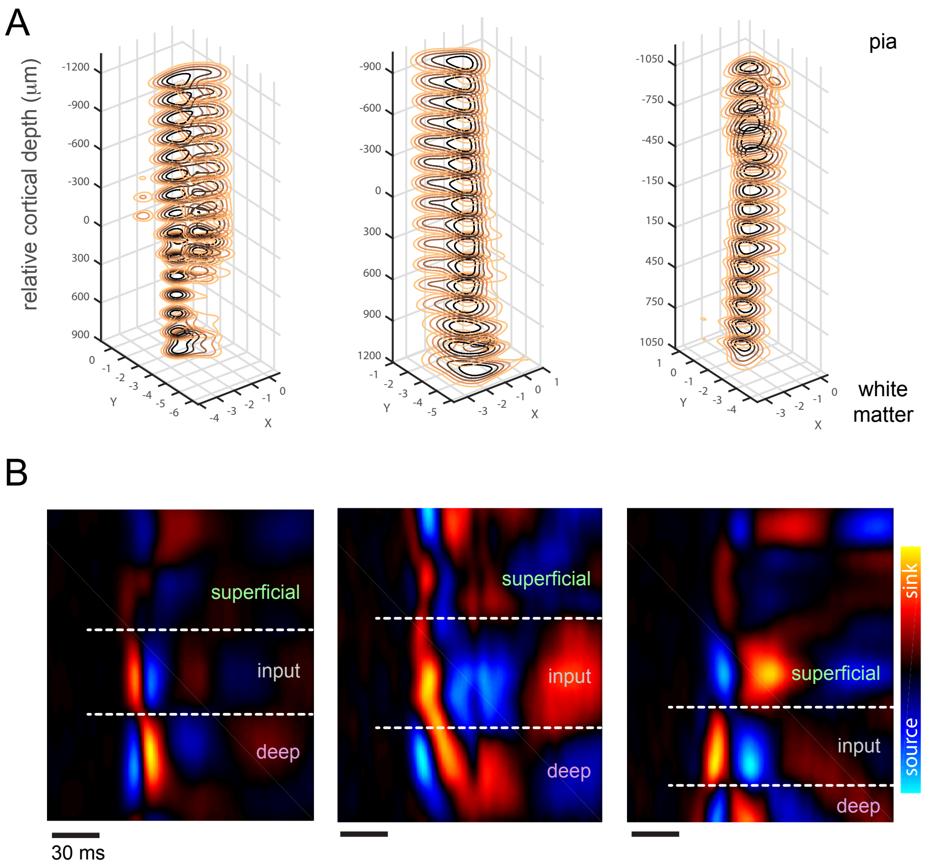
Supp Fig 4. Population PSTH of narrow- and broad-spiking units. Related to Figs 5,8. *Left column*: average firing rate (spikes/s) of narrow- (blue) and broad-spiking (orange) units by layer and attention condition (solid=attend in; dashed=attend away) See also Supp Table 1. *Right column*: same format, but with firing rates normalized to those of the broad-spiking units in the input layer in the attend-away condition. The narrow-spiking units have a much higher firing rate than the broad-spiking units in the input layer. The superficial narrow-spiking units have a higher rate than the superficial broad-spiking units, but their firing rates are much lower than the input layer narrow-spiking units. In the deep layers, the firing rates of the narrow and broad units are not significantly different.

Supp Fig 5. Control analyses for spike-count correlations and variability. Related to Figs 5-7. (**A**) Spike-count correlations for pairs of simultaneously recorded units within a layer are shown as a function of mean response of the pair for two different temporal windows: an inter-stimulus period 200ms before stimulus onset (left panel) and a stimulus-evoked period 60-260ms after stimulus onset (right panel). Correlated variability is highest in the input layer despite mean matching. The firing rates were binned for the analysis (4 bins for the inter-stimulus period, 6 bins for the stimulus evoked period). (**B**) The inverted U-shaped laminar profile of spike-count correlations ($p_{\text{input}\leftrightarrow\text{superficial}} = 0.05, p_{\text{input}\leftrightarrow\text{deep}} \ll 0.01$) and the significant reduction in spike-count correlations due to attention in the input layer ($p = 0.03$) are preserved for pairs of units that were recorded from the same or from immediately adjacent electrodes. (**C**) Mean spike-count variance for broad-spiking neurons in the superficial layer, sorted according to mean spike-count. Measures of spike-count variance and mean were derived over a time window between 122-260ms after non-target flash onset. Mean +/- s.e.m. in all plots.

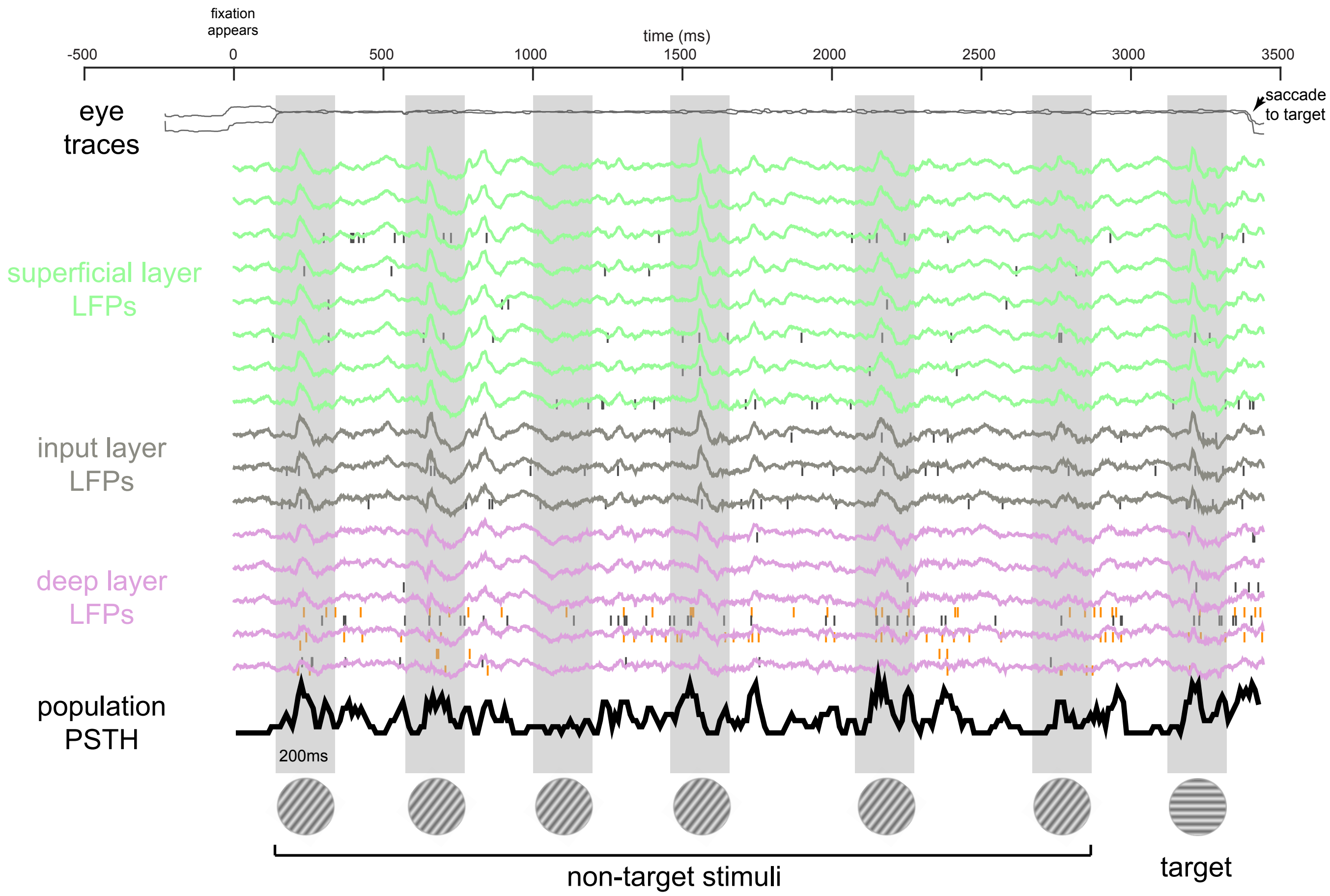
Supp Fig 6. Consistency of results across subjects. Related to Figs 5-7. Left column: Monkey A; Right column: Monkey C. (**A**) Histograms depict the distribution of firing rate modulation indices across layers for all units (single units, multi-unit activity) in a particular layer. Solid vertical lines depict the median values of each distribution. Same format as Fig 5A. (**B**) Histograms depict the distribution of Fano-factor modulation indices for the narrow- and broad-spiking single unit populations in the superficial layer (narrow=blue, broad=orange). Same format as Fig 6. (**C**) Spike count correlations (r_{SC}) between pairs of simultaneously recorded single units within a layer are shown as a function of attention condition for the stimulus-evoked period 60-260ms after stimulus onset. Same format as Fig 7B. Mean +/- s.e.m. in all plots. Star symbols indicate statistically significant differences.

Supp Fig 7. Impact of E-I network connectivity parameters on spike-count correlations. Related to Fig 8. (**A**) *Left panel*: Same format as Fig 8A. *Right panel*: traces show the time course of AMPA, NMDA and GABA conductances. (**B**) Same format as in Fig 8C. Each heat-map shows spike-count correlations as a function of different values of the E-I coupling parameters W_{IE} and W_{EI}, while holding W_{EE} and W_{II} fixed. The shaded bars indicate the direction of stronger coupling. A strong inhibitory feedback loop (larger absolute values of W_{EI} and W_{IE}) increases correlated activity. This is especially so when inhibition is high in the network (low absolute values of W_{II}). (**C**) Spike-count correlations for different network connection probabilities. *Left panel*: fully connected network; *Middle panel*: network with connection probability of 80%; *Right panel*: network with 50% connection probability.

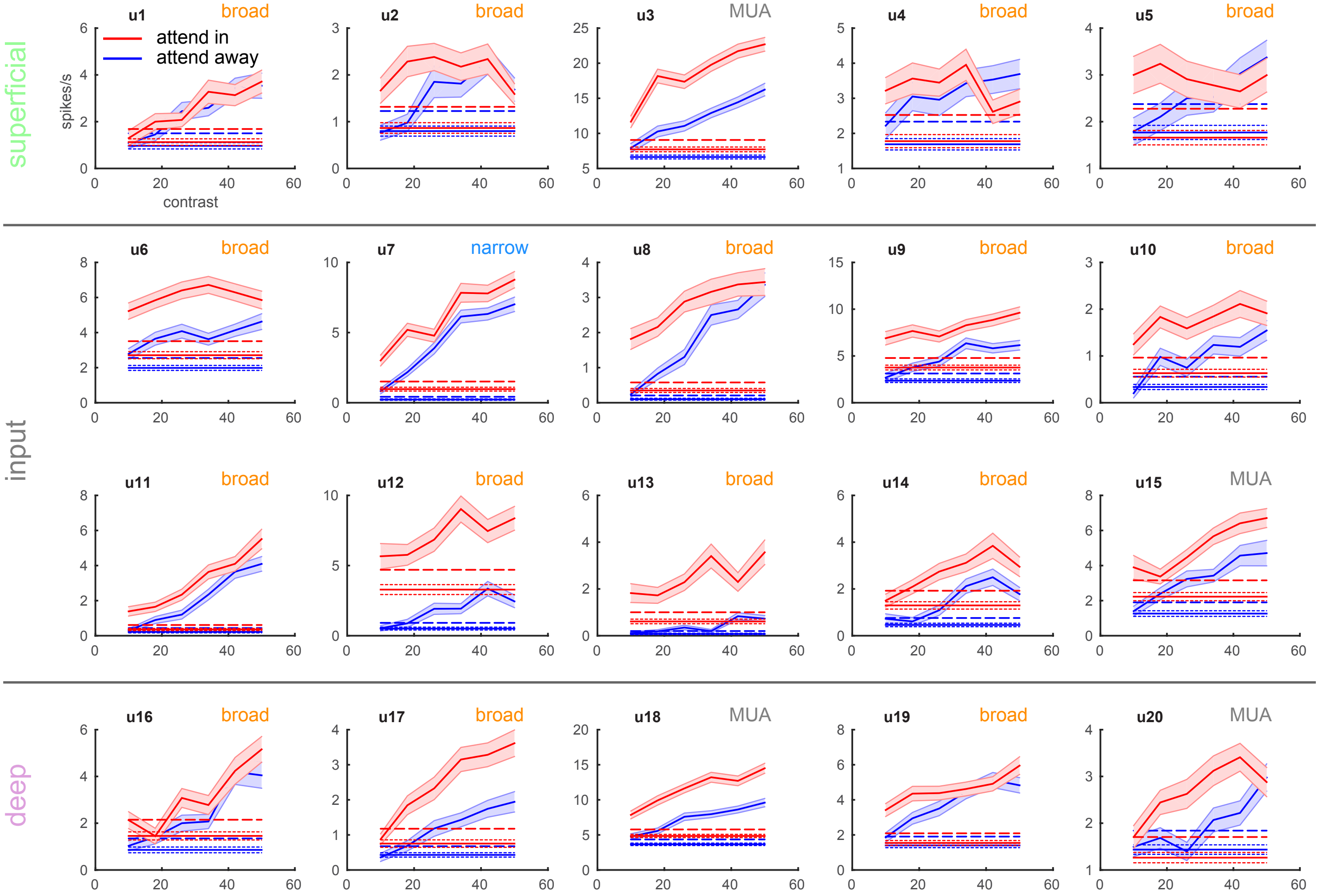
Supplementary Figure-1 NANDY



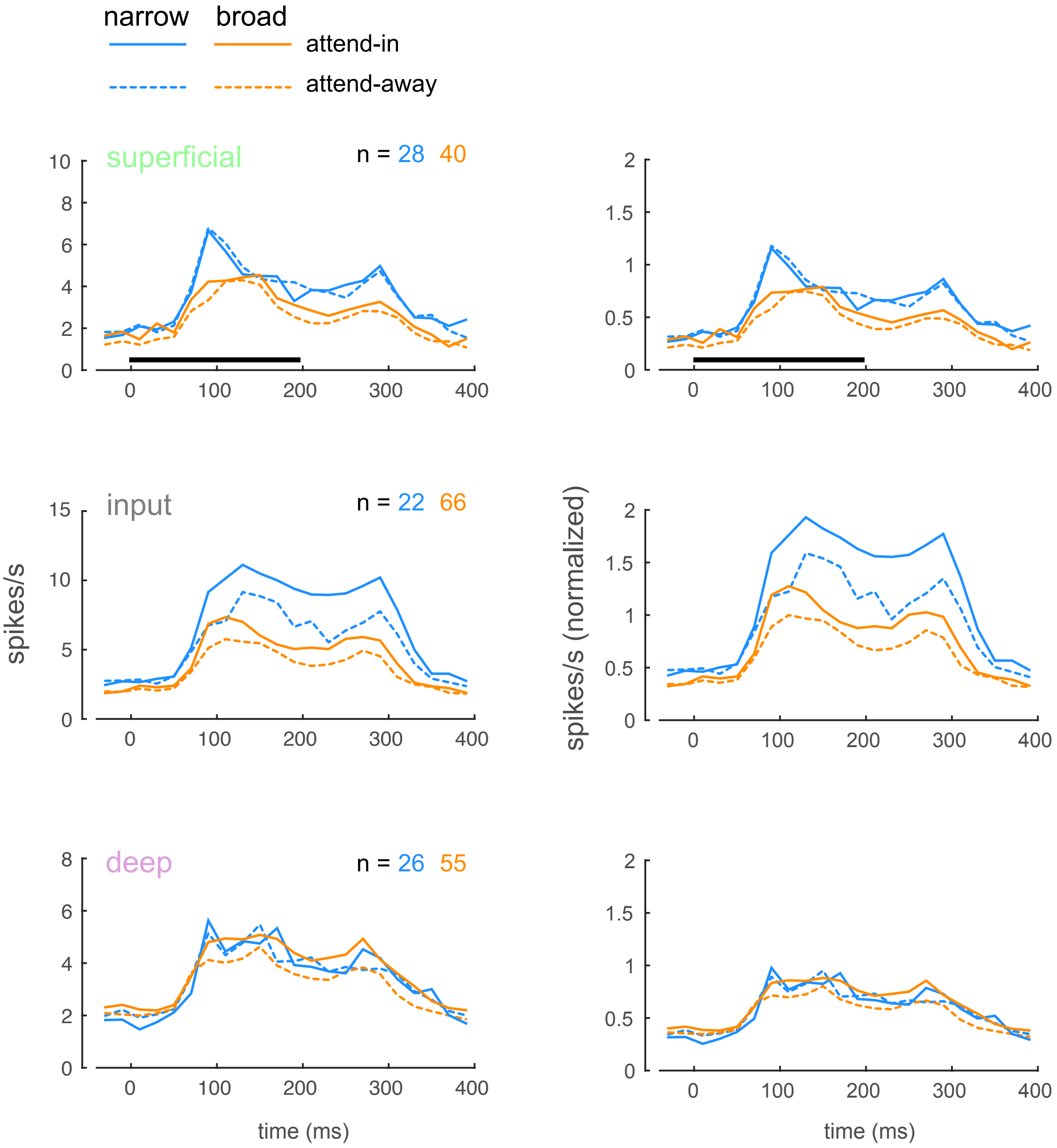
Supplementary Figure-2 NANDY



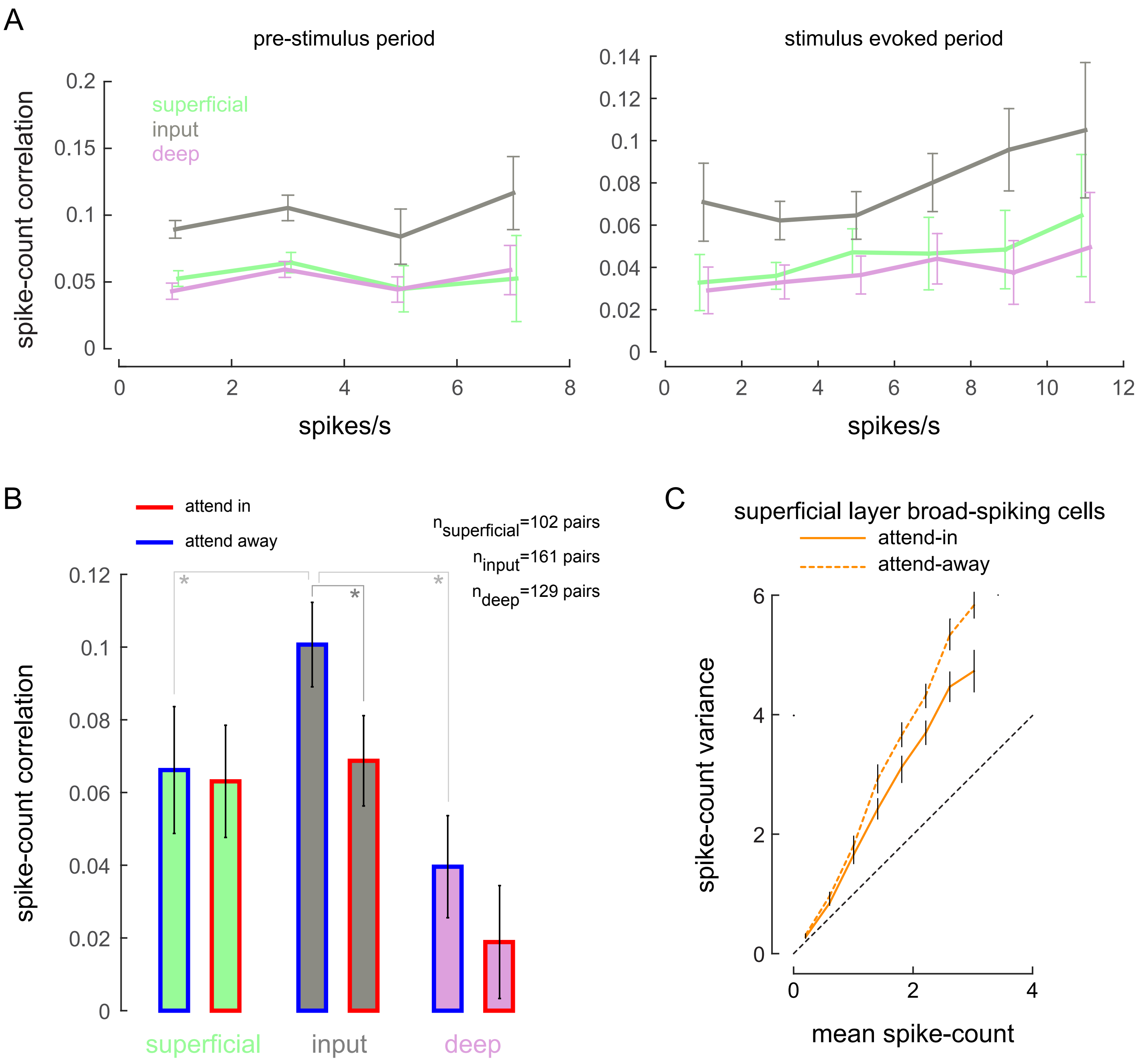
Supplementary Figure-3 NANDY



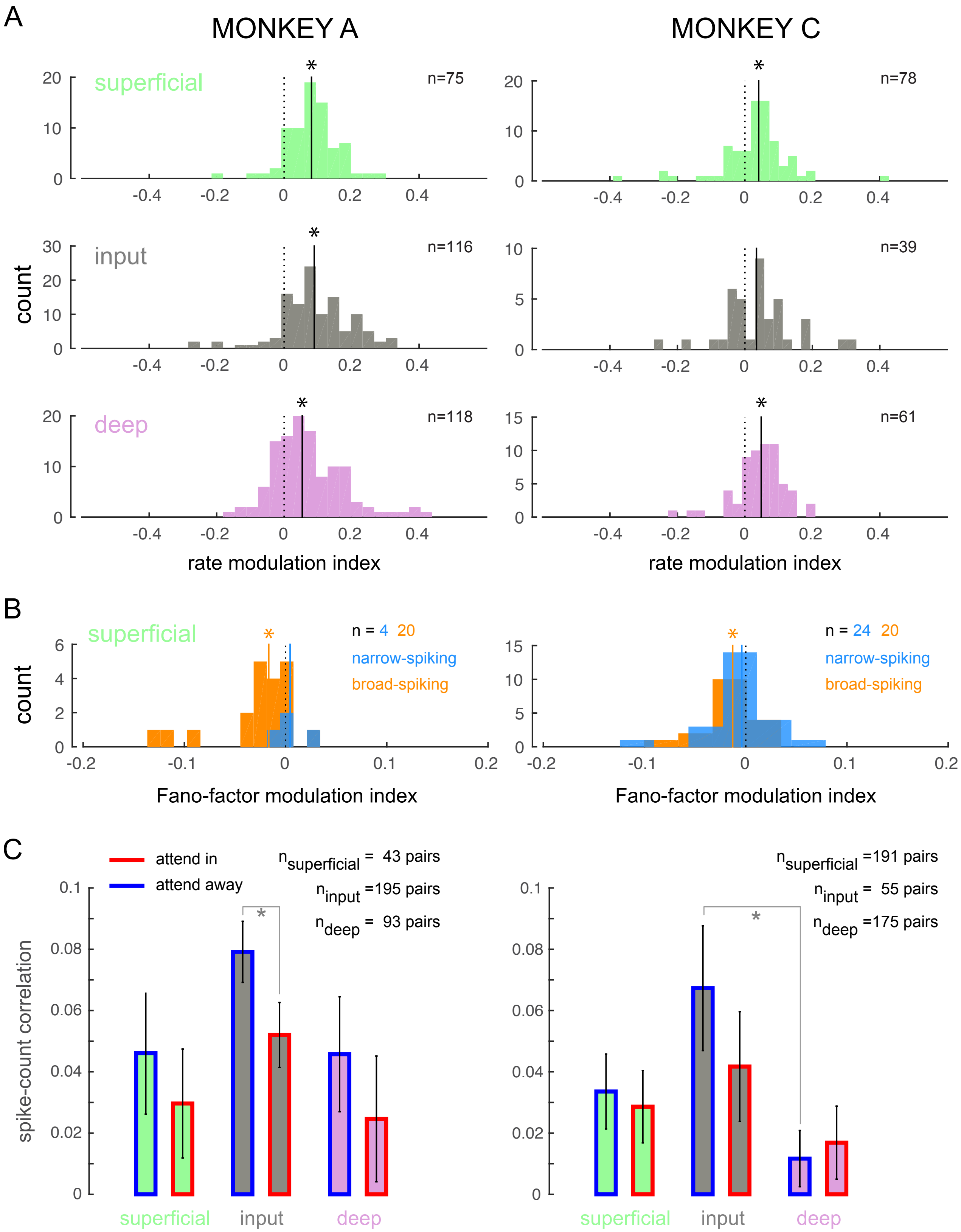
Supplementary Figure-4 NANDY



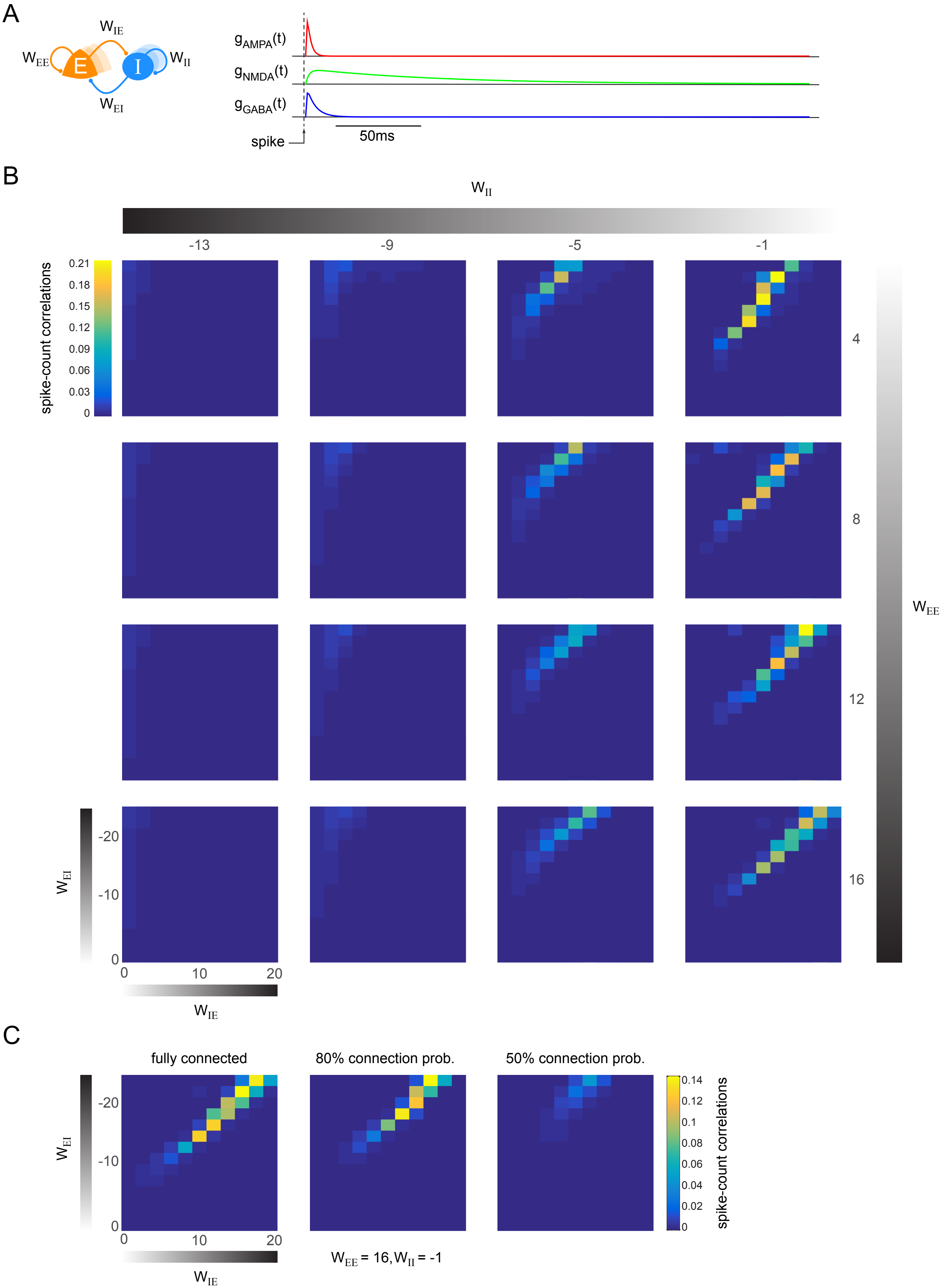
Supplementary Figure-5 NANDY



Supplementary Figure-6 NANDY



Supplementary Figure-7 NANDY



Supplementary Table 1. Empirical values of firing rate (spikes/s), Fano Factor and spike-count correlations (r_{SC}). Related to Figs 5-7. The firing rate is reported as the average spike rate within a time-window 60-260 after non-target flash (34% luminance contrast) onset. The Fano Factor is reported as the average over a time window between 122-260ms after non-target flash onset. The spike-count correlations were calculated from the stimulus evoked period 60-260ms after non-target flash onset. Narrow=narrow-spiking putative interneurons; Broad=broad-spiking putative excitatory neurons; All=single units + multi-unit activity

			Attend AWAY		Attend IN		%
			mean	s.e.m.	mean	s.e.m.	change
Firing Rate	Superficial	Narrow	4.55	0.76	4.46	0.81	-2.05
		Broad	3.13	0.43	3.57	0.51	13.83
		All	4.69	0.40	5.16	0.45	9.96
	Input	Narrow	7.07	2.40	9.24	3.53	30.73
		Broad	4.63	0.51	5.73	0.64	23.75
		All	5.66	0.46	6.89	0.62	21.82
	Deep	Narrow	4.31	0.94	4.29	0.90	-0.32
		Broad	3.85	0.58	4.52	0.66	17.32
		All	5.64	0.51	6.27	0.54	11.17
Fano Factor	Superficial	Narrow	1.21	0.06	1.20	0.07	-0.50
		Broad	1.21	0.05	1.13	0.05	-6.49
	Input	Narrow	1.10	0.03	1.08	0.03	-1.44
		Broad	1.27	0.04	1.25	0.04	-1.94
	Deep	Narrow	1.05	0.07	1.04	0.07	-0.63
		Broad	1.23	0.05	1.24	0.06	0.50
Spike-Count Correlations	Superficial		0.035	0.01	0.028	0.01	-19.93
	Input		0.074	0.009	0.047	0.009	-36.18
	Deep		0.023	0.009	0.019	0.01	-15.17