ORIGINAL ARTICLE

Spiking Suppression Precedes Cued Attentional Enhancement of Neural Responses in Primary Visual Cortex

Michele A. Cox[1], Kacie Dougherty[1], Geoffrey K. Adams[2], Eric A. Reavis[3,4], Jacob A. Westerberg[1], Brandon S. Moore[1], David A. Leopold[5] and Alexander Maier[1]

[1]Department of Psychology, College of Arts and Science, Vanderbilt Vision Research Center, Center for Integrative and Cognitive Neuroscience, Vanderbilt University, Wilson Hall, 111 21st Ave S, Nashville, TN 37240, USA, [2]Center for Translational Social Neuroscience, Emory University, Atlanta, GA 30322, USA, [3]Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA, [4]Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA, USA and [5]Section on Cognitive Neurophysiology and Imaging, Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, 49, Convent Drive, Bethesda, MD, USA

Address correspondence to Alexanader Maier. Email: alex.maier@vanderbilt.edu

Abstract

Attending to a visual stimulus increases its detectability, even if gaze is directed elsewhere. This covert attentional selection is known to enhance spiking across many brain areas, including the primary visual cortex (V1). Here we investigate the temporal dynamics of attention-related spiking changes in V1 of macaques performing a task that separates attentional selection from the onset of visual stimulation. We found that preceding attentional enhancement there was a sharp, transient decline in spiking following presentation of an attention-guiding cue. This disruption of V1 spiking was not observed in a task-naïve subject that passively observed the same stimulus sequence, suggesting that sensory activation is insufficient to cause suppression. Following this suppression, attended stimuli evoked more spiking than unattended stimuli, matching previous reports of attention-related activity in V1. Laminar analyses revealed a distinct pattern of activation in feedback-associated layers during both the cue-induced suppression and subsequent attentional enhancement. These findings suggest that top-down modulation of V1 spiking can be bidirectional and result in either suppression or enhancement of spiking responses.

Key words: cortical column, feedback, inhibition, macaque, microcircuit

Introduction

Primate visual function is characterized by the capacity to direct attention to a position in visual space that is different from the point of eye gaze fixation (Posner 2016). Intracranial recordings have demonstrated that spiking responses to attended stimuli at the earliest stage of cortical visual processing, the primary visual cortex (V1), differ substantially from spiking responses to unattended stimuli. While early intracranial studies found little to no attentional modulation of V1 neurons (Moran and Desimone 1985; Luck et al. 1997), later studies

using smaller stimuli and more challenging tasks revealed consistent, if modest, response gain with spatial attention (Motter 1993; Posner and Gilbert 1999; McAdams et al. 2005; Thiele et al. 2009). In addition to these spike rate changes, attentional modulation of V1 responses also affects the correlational structure and shared variance between local neural populations (Siegel et al. 2008; Poort and Roelfsema 2009; Bosman et al. 2012; Herrero et al. 2013; Ruff and Cohen 2016). Furthermore, attentional modulation in V1 seems to be exclusively spatial in nature, having virtually no effect on the featural tuning of V1 neurons (Haenny and Schiller 1988; McAdams and Maunsell 1999 but see Ito and Gilbert 1999). Yet, relatively little is known about the temporal dynamics of these attention-related changes of V1 spiking. Most existing studies focused on stimulus processing during periods of sustained visual attention and used a task design where the onset of attention preceded or coincided with the onset of visual stimulation, which prohibits the study of attentional onsets in isolation. Interestingly, a handful of studies focusing on the temporal dynamics of attentional modulation in macaque areas MT, MST, and LIP revealed a biphasic profile of spiking activity, in which attentional spiking enhancement set in after brief, but widespread suppression of activity (Bisley and Goldberg 2006; Busse et al. 2008; Herrington and Assad 2009, 2010). Whether this biphasic temporal profile of attentional spiking modulation is idiosyncratic for parietal areas or can also be observed in early sensory areas, such as V1, remains an open question.

Here we investigate the temporal dynamics of attentional modulation of spiking responses within V1's laminar microcircuit in monkeys performing a covert attentional orienting task that temporally dissociated the target and distractor stimulus onset from attention-related task events. We used linear microelectrodes that spanned the depth of cortex to capitalize on the anatomical segregation of feedforward and feedback projections across V1 layers. In line with previous studies, we found that spiking responses throughout V1—but particularly in the upper and lower layers—were enhanced by attention. However, this spiking increase was preceded by a V1-wide sharp, transient decrease in spiking following onset of the attentional cue that persisted for ~100–250 ms. Importantly, this transient attention-induced suppression of V1 spiking was absent in a subject that passively observed the same stimulus sequence without ever learning the task, suggesting that the biphasic modulation of V1 spiking we observed in the task condition is due to top-down modulation. This assumption is consistent with the laminar profile of activity, which shows the strongest modulation in feedback-recipient layers during both phases of the biphasic response.

Methods

Subjects: Four adult male monkeys (*Macaca mulatta*, monkey B and E; *Macaca radiata* monkey Es and monkey Br who is also referred to as the naïve monkey) were used in this study. All procedures were in compliance with regulations by the Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC), approved by the National Institute of Mental Health's Animal Care and Use Committee or Vanderbilt University's Institutional Animal Care and Use Committee and followed National Institutes of Health guidelines.

Surgical Procedures: In a series of surgeries, each monkey was implanted with a custom-designed head holder and a plastic recording chamber. All surgeries were performed under sterile surgical conditions using isoflurane anesthesia (1.5–2.0%),

following induction with an intramuscular injection of ketamine hydrochloride (10 mg/kg). Vital signs, including blood pressure, heart rate, SpO_2, CO_2, respiratory rate, and body temperature, were monitored continuously. During surgery, a head holder and a recording chamber were attached to the skull using transcranial ceramic screws (Thomas RECORDING GmbH), Copalite varnish (Cooley & Cooley, Ltd.), and self-curing denture acrylic (Lang Dental Manufacturing Co., Inc.). Further, a craniotomy was performed over the perifoveal visual field representation of primary visual cortex (area V1; left hemisphere) concurrent with the position of the recording chamber. Each monkey was given analgesics and antibiotics for post-surgical care.

Visual Stimulation: Stimuli consisted of either red or green gratings (see Supplementary Fig. 10 for stimulus-specific analyses), and were generated and presented using one of two apparatus. Apparatus #1: Custom-written OpenGl-based software (ESS/STIM; copyright Dr D. Sheinberg, Brown University) was run on a PC (Kontron) using NVIDIA Quadro FX 3000 graphics boards in order to produce the visual stimuli. In this case, visual stimuli were presented on two separate 27 inch TFT monitors (X2Gen MV2701, 1024 × 768 resolution) positioned 80 cm in front of the subject with a custom-made stereoscope (Maier et al. 2008). Gaze position was monitored continually and recorded at 200 Hz using an infrared light sensitive camera and commercially available eye tracking software (Eye Link II, SR Research). Eye movements and behavioral events were synced to the neurophysiological data using a separate PC running a real-time operating system (QNX Software Systems). The majority of data presented here (N = 45 penetrations in B, N = 25 in E) was collected using this apparatus. Apparatus #2: The freely available open-source library "MonkeyLogic" (Asaad and Eskandar 2008; Asaad et al. 2013) for MATLAB (R2011b, The MathWorks) running on a PC (Dell, Windows 7, 4 GB RAM) with a NVIDIA GeForce 680 graphics card was used to generate visual stimuli as well as to synchronize eye movements and behavioral events. Visual stimuli were presented on a 45.7 cm CRT monitor (NEC Fe991SB-BK, 60 Hz, 1280 × 1024) positioned 32 cm or 52 cm from the subject with a custom-made stereoscope (Maier et al. 2008). Gaze position was recorded at 1 kHz (NIDAQ PCI-6229) using an infrared light sensitive camera and commercially available eye tracking software (Eye Link II, SR Research). Apparatus #2 was used exclusively for monkeys Br (N = 18 penetrations) and Es (N = 11 penetrations).

Behavioral Tasks: We used 3 different tasks in total (a main task, plus two controls). All 4 monkeys were trained to hold their gaze steady within 0.5–1 degree radius of visual angle (dva) around a small (0.2 dva) spot presented in the center of the monitor for extended periods of time (5–10 s) while circular gratings and other stimuli were presented in the perifoveal visual field. Three of the monkeys (B, E, and Es) were further trained on a cued change detection task.

The *main attention task* (monkeys B & E) consisted of the following sequence (Fig. 1a): On each trial, the subjects fixated for 1 s on a blank screen. The subjects continued to maintain fixation while the main stimulus set appeared, consisting of an array of 4 evenly spaced, perifoveal, isoeccentric gratings. After another 800 ms, a short (0.5 dva) bar-shaped cue appeared, radiating from the fixation spot and pointing toward one of the 4 grating stimuli. After a 2.5 s delay following this event, a random timer was initiated that lasted for a maximum of 2 s. For 20% of trials, no events occurred during this time and the subjects were rewarded for not responding (catch trials). In the remaining 80% of trials (non-catch trials), a stimulus change
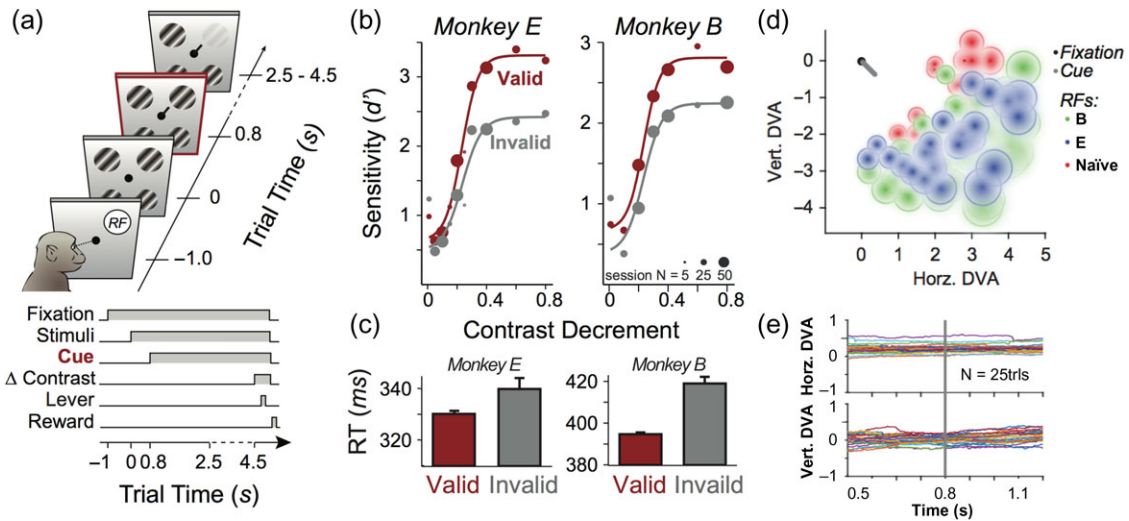
Figure 1. Experimental paradigm and psychophysical data. (a) Modified Posner spatial attention task. Animals are rewarded for correctly indicating a contrast change in one of the peripheral gratings via a lever press or correctly withholding a response in the absence of a change on catch trials. The cue presented at fixation earlier in the trial predicts the location of stimulus change with 80–90% validity. (b) Behavioral performance. Plots show contrast sensitivity in d' as a function of contrast decrements across "valid cue" (red) and "invalid cue" (gray) trials for each monkey average over behavioral testing days (point size scales with the number of sessions included in average). Curves are logistic fits. (c) Average reaction time of each monkey to the contrast decrement across valid cue (red) and invalid cue (gray) trials. Error bars represent standard error of the mean (SEM). See Supplementary Table 1 for detailed statistical analysis of behavioral data. (d) Relative locations of fixation dot (black), foveal cue (gray), and mapped receptive fields (green dots for B, blue for E, red for naïve subject). For visualization purposes, receptive field diameters (solid lines) were estimated from the mapped receptive field (RF) centers using an estimate of cortical magnification (diameter = 0.21*eccentricity (Freeman and Simoncelli 2011)), and fitted with a Gaussian gradient. (e) Stability of fixation. Horizontal (top) and vertical (bottom) gaze position relative to fixation for the ±250 ms surrounding foveal cue onset (n = 25 trials, monkey E).

could occur with equal probability at any point during this time period. The stimulus change consisted of one of the 4 gratings in the array slightly dimming in contrast (spanning each subject's perceptual threshold). On 80–90% of the non-catch trials, the stimulus change was applied to the grating indicated by the cue (valid trial). In the other trials, the stimulus change was applied to a different grating (invalid trial). When the stimulus change occurred, subjects were rewarded for indicating the stimulus change via a lever press within 800 ms of the change regardless of its location.

We employed a "modified attention" task (monkey Es) as another control (Supplementary Fig. 4). The major difference between the main task and the modified task was that the subject was cued twice, specifically, the future location of the stimulus change was first cued before stimulus onset and then a second, redundant cue was shown later in the trial. This redundant cue was similar to the main attention task. The stimulus sequence was as following: After 250 ms of fixation, a monochromatic spot appeared at a predetermined location in visual space. After 500 ms, the cue disappeared and following another 250 ms, two gratings stimuli were presented. One of the two gratings coincided with the cue location, and the other grating was shown with equal eccentricity on the opposite side of the fixation spot. At this point a random timer was initiated that lasted for a maximum of 500 ms. On 10% of trials, no events occurred and the subjects were rewarded for not responding (catch trials). On the remaining 90% of the non-catch trials (target trials), a stimulus change could occur with equal probability at any point during this time period. The stimulus change consisted of a subset of the grating taking on a slight reddish hue (spanning the subject's perceptual threshold). On 90% of the non-catch/target trials, the stimulus change was applied to the grating indicated by the cue (valid cue). In the other trials, the stimulus change was applied to the opposite grating (invalid cue). When the stimulus change occurred, subjects were

rewarded for indicating the stimulus change by saccading to the stimulus location within 450 ms.

Behavioral Analysis. Performance for both the main and the modified attention task was assessed in the form of response accuracy (change detection sensitivity) and reaction time (Fig. 1b,c, Supplementary Table 1). Detection sensitivity across *valid cue* and *invalid cue* trials was computed in units of d' for each change magnitude for each day that the subject performed the task (Monkey B, N = 62 behavioral testing days; Monkey E, N = 56, Monkey Es, N = 11) and then averaged for each monkey across testing days. d' was estimated as the difference between z-transformed (inverse cumulative of the normal distribution: $\mu = 0$, $\sigma = 1$) hit rate and false alarm rate, which were each set to a minimum value of $1/n$ and a maximum value of $(n-1)/n$ where n is the number of trials per testing day. Hit rate was calculated for each change magnitude. Catch trials, along with trials where the subjects made a response before a stimulus change, provided a measure of false alarm rate. Reaction time was measured as the time that passed between the stimulus change event and the subjects' response. Reaction times for "valid cue" versus "invalid cue" conditions were examined for correct responses (i.e., hits), both collapsed across contrast decrements (Fig. 1c) and for each contrast decrement (Supplementary Table 1).

Eye Movement Analysis: A previously published algorithm (Otero-Millan et al. 2014) was used to extract microsaccades from the horizontal and vertical gaze position signals recorded concurrently with electrophysiological data.

Neurophysiological Procedures. Broadband (0.5 Hz–12.207 kHz, sampled at 24.4141 kHz for monkeys B and E and 0.3 Hz–7.5 kHz sampled at 30 kHz for monkeys Br and Es) extracellular voltage fluctuations were recorded inside an electromagnetic radio frequency-shielded booth with an acute laminar probe. Two types of laminar probes were used, Plexon UProbes with 16–24 active microelectrodes and NeuroNexus Vector Arrays with 32

active microelectrodes (monkey Es only). In all cases, the microelectrodes of the laminar probes were linearly spaced 0.1 mm apart, with impedances ranging 0.2–0.8 MΩ at 1 kHz. Extracellular voltages were measured in reference to the shaft of the probe and were collected simultaneously from all microelectrode contacts. Extracellular voltages were amplified, filtered, and digitized using either a Tucker Davis 64-channel RZ2 recording system (monkeys E and B) or a Blackrock Microsystems 128-channel Cerebus® Neural Signal Processing System (monkeys Br and Es). Each recording session, a laminar probe was introduced into dorsal V1 through the intact *dura mater* using a chamber-mounted microdrive (either custom-made, Alpha Omega FlexMT or Narishige Micromanipulator) and adjusted until microelectrode contacts spanned the entire cortical thickness, from the subdural space to the white matter (see Laminar Alignment).

Laminar Alignment: Laminar location was verified by a combination of physiological criteria, such as the distribution of receptive fields, intra-electrode coherence as well as the pattern of visually evoked responses to a flash of light (Maier et al. 2010). For each penetration with the laminar probe, CSD analysis (see Neurophysiological Data Analysis and Statistical Testing) was used to resolve the bottom of the prominent initial current sink immediately following stimulus onset. CSD analysis of visual responses to brief visual stimulation have been shown to reliably indicate the location of the primary geniculate input to V1 (granular layer 4C, L4C) by a distinct current sink that is thought to reflect combined excitatory postsynaptic potentials of the initial retino-geniculate volley of activation (Mitzdorf and Singer 1979; Schroeder et al. 1998). CSD and MUA responses were further used to identify electrode contacts that lay outside V1, either in the subdural space or white matter below. After removing these channels, the CSD initial current sink was used to align and average data across electrode penetrations, resulting in 0.1 mm ± 0.05 mm resolution across V1's cortical depth (Maier et al. 2010) (Supplementary Fig. 1).

Receptive Field Mapping: Once satisfactory electrode placement was achieved, we mapped the receptive field of the neurons under study. Specifically, we used the audible multiunit response on each channel to determine the extent of visual space that reliably evoked spiking responses at the electrode location using either a reverse correlation procedure (see Cox et al. 2013 for details) or manual receptive field mapping. The aggregate receptive field across all electrode channels was used to determine the placement of the main stimulus set used in the attention task described above (Supplementary Table 2). In particular, we placed a single stimulus element (i.e., grating) coincident with the receptive field location, with all other stimuli taking equidistant positions around a virtual circle of iso-eccentricity surrounding the fixation spot. Once all the steps outlined above were carried out satisfactorily, we started recording the broadband laminar neuronal activity while the subjects performed the behavioral task outlined above.

Neurophysiological Data Analysis and Statistical Testing: All data analysis was performed offline using custom-written code in MATLAB. From the broadband signal, local field potentials (LFP) were extracted by low-pass filtering using a bidirectional first-order Butterworth filter with a cut off at 500 Hz. Current source density (CSD), which is a more localized estimate of synaptic activity (Mitzdorf 1985; Shmuel and Maier 2015), was estimated as the Laplacian of the laminar LFP. Specifically, an estimate of the second spatial derivative appropriate for multiple contact points was used compute the visually evoked CSD (Nicholson and Freeman 1975):

$$CSD(t, c) = -\frac{x(t, c - z) + x(t, c + z) - 2x(t, c)}{z^2}$$

where x is the extracellular voltage recorded in Volts at time t from an electrode contact at position c, and z is the electrode intercontact distance (0.1 mm). In order to yield CSD in units of current per unit volume, the resulting CSD from the formula above was multiplied by 0.4 S/m, an estimate of the conductivity of cortex (Logothetis et al. 2007).

Single units can be isolated with laminar probes in a way that is comparable to that of standard microelectrodes. However, isolating cells on all electrode contacts of the array simultaneously proves challenging in practice. For this reason, we opted to use multiunit activity (MUA) as a proxy for the activity of local neurons. We used two complementary approaches to estimate MUA. First, we computed a "discretized" MUA signal by applying a threshold of 2.5 standard deviations above the mean signal to the high-pass filtered extracellular voltage signal on each electrode contact to determine the impulses that crossed that threshold over time. This procedure is analogous to the initial step of spike sorting. Second, we computed the power of the signal within the frequency range that characterizes the spiking activity of single neurons ("analog" MUA). Specifically, we full-wave rectified the recorded high-pass filtered (at 300 Hz) data, to obtain a measure of the signal magnitude within the spiking frequency band, and smoothed this analog signal with a low-pass filter (Supèr and Roelfsema 2005). 3D representations of MUA as a function of time and cortical depth were created by interpolating MUA between adjacent electrode contacts ($n = 100$ pts).

Attentional enhancement was calculated as the mean difference between attended and unattended conditions, after trial averaging, across the 300–1600 ms following cue onset (Fig. 4). Response differences following cue onset were calculated by comparing neural activity averaged across 125–175 ms to activity averaged across −50–0 ms relative to foveal cue onset (Fig. 4). Where applicable, a measure of baseline activity was calculated by averaging across −50–0 ms relative to cue or stimulus onset. The difference in pre- versus post-cue MUA (Supplementary Fig. 2c–d) was calculated by subtracting the baseline from each of the pre- and post-period measures, dividing, subtracting 1, and multiplying by 100. The difference in CSD activity following the foveal cue onset was calculated in absolute units by subtracting the pre-cue measure from the post-cue measure of neural activity. In all cases related to pre- and post-cue comparisons, difference measures were calculated for each trial, averaged across trials for each V1 electrode, and then averaged across depth-aligned penetrations with standard error of the mean computed across penetrations. Response latency was calculated on a trial-by-trial basis within a pre-defined window of 50–300 ms relative to each stimulus and cue onset. For stimulus onset, the time to response maximum was used as a measure for latency, and for cue onset, the time to response minimum was used as a measure of latency. For Figure 3d, the median of the response latency was taken across trials, while preserving cortical depth as another dimension. The 95% confidence interval of the median was calculated across each electrode depth via Monte-Carlo bootstrapping (1000 iterations).

Analyzing data from the modified attention task (Supplementary Fig. 4) required an additional step due to the specific design of the paradigm. Specifically, the timing of cue onset within a given trial was random relative to other task events, such as stimulus onset. Thus, the cue onset was conflated with varying levels of stimulus response depending on how long after stimulus onset it appeared (i.e., was it shown during the transient or

sustained response). In order to examine cue-specific responses, rather than stimulus-driven responses, we analyzed the data in 2 ways: (1) We selected trials where the redundant cue appeared in a predetermined time window (±33 ms of $t = 0.1$ s) relative to stimulus onset and then epoched the time-varying electrophysiological signal to the cue onset, thereby jittering the relative time of stimulus onset across trials. (2) We directly compared trials where the redundant cue appeared to identical (same stimulus sequence and initial cue) trials where the redundant cue was absent. Specifically, we epoched all data relative to stimulus onset, subtracted cue absent trials from cue present trials, and then re-epoched each trial relative to cue onset.

All statistical tests were performed on either individual penetration data with trials serving as the degree of freedom or trial-averaged data with penetrations serving as the degree of freedom. Across-penetration averages and analyses excluded cortical depths for which an electrode occurred on less than 10 penetrations. Data were collapsed across a window of interest: 300–1600 ms relative to cue onset (110–2400 ms relative to stimulus onset) for analyses of attentional response gain, 100–200 ms relative to cue onset (900–1000 ms relative to stimulus onset) for analyses of attentional orienting-related activity. Two types of tests were used on neurophysiological data: Student's t-tests and analysis of variance (ANOVA) tests. Paired t-tests were employed for comparisons across conditions or laminar compartments. One-sample t-tests were used to test data against $\mu_0 = 0$. Repeat-measure ANOVAs were employed to test for a main effect of cortical depth. For three-dimensional color plots, data was normalized using z-score along the colored axis:

$$z = \frac{x - \mu}{\sigma}$$

where x is the across-penetration average, μ set to 0 in line with the assumption of a null effect, and σ is across-penetration standard deviation. For display, representations of CSD as a function of time and space were created by interpolating CSD between adjacent electrode contacts, smoothing with a 2D-Gaussian filter (sigma = 0.1 mm and 15 ms) (Pettersen et al. 2006).

Results

Two macaque monkeys (B and E) were trained on a spatial attention task designed to temporally separate attention-related task periods from other stimulus events (Fig. 1*a*). As controls, a third subject was kept uninformed about the meaning of the cue while experiencing the same stimulus sequence (monkey Br; Naïve), and a fourth subject (Es) performed a modified spatial attention task that featured a redundant cue (see Methods).

Task Design

The main attention task (performed by B and E) proceeded as follows: On each trial, the subject first fixated a small central spot (0.2 degrees of visual angle, dva) on a computer monitor (Fig. 1*a*). After 1 s continuous fixation on the gray monitor gray, the main stimulus set appeared, which consisted of an array of 4 evenly spaced, perifoveal, isoeccentric, monocular gratings (see Methods for stimulus details). The subjects were required to maintain fixation. After 800 ms of exposure to the stimulus array, a small (0.5 dva) bar was added over the fixation spot, which pointed toward one randomly selected grating of the original, unchanged stimulus array. After another 1.6 s, a random timer with a uniform probability distribution was initiated

that expired after a maximum of 2 s. Once the randomly chosen time interval expired, one of the gratings in the array slightly dimmed in contrast. For most trials, this contrast decrement was applied to the grating indicated by the cue (valid target trial). For the remaining trials, the contrast decrement was applied to a different grating (invalid target trial) or to no grating at all (catch trials). Twenty percent of all trials were catch trials, 80–90% non-catch trials were valid target trials and 10–20% were invalid trials. Whenever the contrast decrement occurred, subjects were rewarded for pulling a lever within 800 ms of the stimulus change, regardless of whether it was at the cued location or not. On trials where no contrast change occurred, subjects were rewarded for withholding a response. The magnitude of the contrast decrement was variable, spanning the subjects' perceptual threshold (see next section).

Behavioral Measures

The behavioral data collected from this task indicate that both trained subjects successfully leveraged the information provided by the foveal cue. We computed psychometric curves by estimating the sensitivity index d′ for various contrast decrements with each subject's false alarm rate determined from the catch trials without a contrast change. Comparing contrast change detection performance for the cued stimulus versus uncued stimuli, we found that the subjects' contrast sensitivity was significantly higher for cued stimuli (Fig. 1*b*; see Supplementary Table 1 for detailed statistical analysis), similar to what has been observed in humans on comparable tasks (Herrmann et al. 2010). Both subjects were also significantly faster at reporting the contrast decrement for the cued stimulus (Fig. 1*c*, Supplementary Table 1). Together, these findings demonstrate that the trained monkeys understood and utilized the predictive nature of the foveal cue to improve their behavioral performance. Importantly, from this design, we could study separately two distinct visual events—the onset of stimuli in the periphery and the subsequent onset of the behaviorally meaningful cue at the center of fixation (see Fig. 1*d* for scale). This sequence allowed us to investigate the consequences of cueing while the stimulus was continually present within neurons' receptive fields.

Biphasic Spiking Responses to Cued Attention in V1

Given the reliable behavioral performance of our trained subjects, we next examined how laminar neural activity in V1 changed while these subjects performed the task and compared their visual responses to those evoked in the naïve subject by the same stimulus sequence. Using linear multielectrode arrays with 100 micron interelectrode spacing, we measured neuronal activity across all layers of V1 cortex. Spiking activity of V1 neurons was extracted as multiunit activity (see Methods), with each electrode in the array that fell within the V1 gray matter comprising a multiunit ($n = 351$ multiunits across $N = 25$ penetrations in E, $n = 642$ multiunits across $N = 45$ penetrations in B, $n = 273$ multiunits across $N = 18$ penetrations in Naive Br). Taking advantage of the linear arrangement of microelectrodes, we reconstructed the supra- (layers 1–4B), infra- (layers 5&6) and granular (layers 4Cα&β) compartments using established neurophysiological criteria (Mitzdorf 1985; Schroeder et al. 1998; Maier et al. 2010, 2011; Spaak et al. 2012; Dougherty et al. 2015; Ninomiya et al. 2015) (see Methods and Supplementary Fig. 1). During all recordings, we placed the stimulus array so that one of the gratings covered the previously mapped receptive field of the recorded neurons (Fig. 1*d*). Animals were required to fixate

within a 0.5–1dva radius window throughout the entire trial, and we performed post-hoc analyses to ensure that the data was not systematically affected by any residual eye movements (Fig. 1e).

Upon presentation of the attention cue, neural responses in V1 exhibited two prominent response components, taking the form of a biphasic profile as shown in Figure 2a. The first component was non-specific to attentional state, suppressive, and transient (thick vertical arrow). This suppressive transient was followed by a second component of sustained spiking enhancement (thin horizontal arrow) that was exclusive for attended stimuli, matching previous reports (Motter 1993; McAdams and Maunsell 1999; Posner and Gilbert 1999; McAdams et al. 2005; Thiele et al. 2009; Chen and Seidemann 2012).

The initial response component was expressed as a transient spiking suppression, which began approximately 100 ms following the cue onset and lasted approximately 200 ms. This suppression played out despite the constant physical presence of an excitatory stimulus within neurons' receptive field and despite uninterrupted fixation on the central point. Following this transient suppression, the spiking responses eventually returned to the stimulus-driven level established prior to the presentation of the cue. The suppression took a similar form for the two subjects and was present at nearly every V1 recording site with significant suppression exhibited by 97% of multiunits ($\alpha = 0.05$; monkey B: 717 of 749 multiunits over 45 penetrations; monkey E: 429 of 434 multiunits over 25 penetrations).

The longer latency of this spiking suppression relative to the stimulus response latency is one of several pieces of evidence to suggest that the observed transient derived from internal (trained) responses to the cued meaning, rather than from the visual sensory input per se. In further support of this idea, the cue was positioned several degrees away from the stimuli, well outside the receptive field of neurons under study (Fig. 1d, Supplementary Fig. 3 & Supplementary Table 2). And notably, suppression was contingent on knowing the meaning of the cue: in a naïve monkey for which the cue had no meaning, the same stimulus sequence elicited no such suppression in the same region of V1 (Fig. 1d and Fig. 2b, blue line: $t(17) = 1.34$, $P = 0.197$). Together, these findings support the view that the observed cue-triggered suppression observed in V1 derives from trained, internal signals related to the meaning of the cue itself, rather than to either direct sensory activation or mechanisms related to reflexive or exogenous attention (see Supplementary Fig. 4 for an additional attentional control).

We further found that the observed suppression could not be ascribed to small, fixational eye movements (i.e., microsaccades). To this end, we eliminated trials where one or more microsaccades were detected in the 200 ms preceding or following cue onset and repeated this analysis. We found that, following the removal of these trials, the cue-induced suppression was still present (Fig. 2b, dark green and purple lines and Supplementary Fig. 2, ~55% of trials removed via this criteria), suggesting that the suppression occurs independent of fixational eye movements.

Following recovery of the initial suppression, a sustained attentional modulation took form beginning approximately 250 ms following the onset of the cue. This attentional modulation of V1 neurons, consistent with several previous findings using somewhat different paradigms (Motter 1993; Posner and Gilbert 1999; McAdams et al. 2005; Thiele et al. 2009), was prominent in both trained subjects (~170 trials were collected per penetration for each cue direction; see Methods). During an analysis window spanning 300–1600 ms following cue onset, multiunit responses were significantly elevated when attention
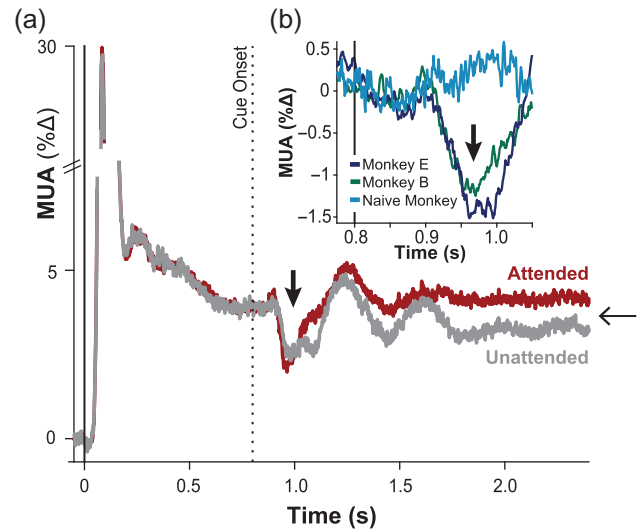

Figure 2. V1 spiking responses. (a) Time course of analog MUA for unattended (gray) and attended (red) stimulus presentations inside the receptive field, averaged across all penetrations (N = 70, monkeys E and B) and V1 multiunits (n = 993, monkeys E and B). MUA is expressed in percent change from the pre-stimulus baseline. Stimulus onset occurred at $t = 0$ s (vertical solid line), which preceded the attentional cue by 800 ms (vertical dashed line). Ordinate is truncated for display purposes. (b) Time course of MUA following cue onset. The naïve monkey (blue line: n = 273 multiunits across N = 18 penetrations) shows no decrease from baseline following cue onset. Both trained monkeys (purple line: n = 351 V1 multiunits across N = 25 penetrations in E; green: n = 642 multiunits across N = 45 penetrations in B) exhibited a decrease in MUA immediately following cue onset (vertical black arrow) even after excluding trials with microsaccades immediately before or after the cue (see Methods).

was directed to the stimulus in the receptive field (Monkey B: $t(44) = 5.95$, $P < 0.001$; Monkey E; $t(24) = 2.00$, $P = 0.028$). This sustained, attention-related enhancement is consistent with models of visual attention that increase the sensory representation of relevant stimuli, with relevance in this case indicated by the presence of a learned, endogenous cue presented after the stimulus was already in place.

Extragranular Spiking Profile of Attentional Modulation

The simultaneous recording from contacts within the linear electrode array allowed for a detailed examination of laminar differences in attention-related neural activity. As attentional modulation is likely a product of input from higher cortical centers, we reasoned that it might have laminar expression reflecting the pattern of feedback projections into V1. Cortical projections to V1 generally innervate superficial and deep layers, and almost completely avoid the middle, geniculo-recipient layer 4C (L4C) (Rockland and Pandya 1979; Lund 1988). The laminar spiking response profiles in Figure 3a–c show how neurons in different cortical layers responded to the onset of the main stimulus and subsequent attentional cue. The laminar profile of the response latencies of the respective positive- and negative-going responses are shown in Figure 3d. For the 2 trained monkeys (Fig. 3a,b), the cue onset (dashed line) is followed by a transient suppression, and the latency of this cued suppression was overall much later than the initial stimulus-driven excitation. Also, whereas the granular layers (L4C) showed the shortest latency for the stimulus onset, the same multiunits showed the longest latency for the suppression. Further, the transient neural suppression was present independent of the cue direction relative to
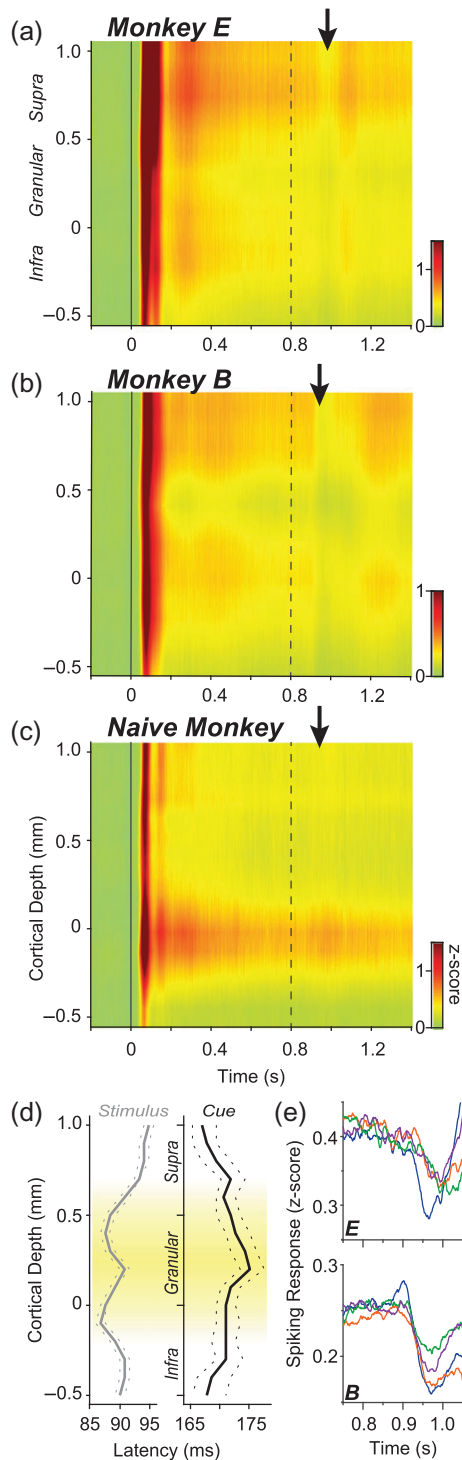
Figure 3. Laminar spiking profiles. (*a–c*) Average analog MUA across V1 layers for each monkey E, monkey B, and the naïve monkey. The stimulus array and the foveal cue where shown at each t = 0 s and t = 0.8 s respectively. Black vertical arrows mark spiking suppression exclusive to trained subjects. The naïve monkey was ignorant to the task-relevance of the cue. (*d*) Latency of responses following each the stimulus onset (gray) and the foveal cue onset (black) in trained monkeys E and B. Solid line represents the median across all penetrations and dashed lines represent 95% confidence limits on the median. Note the differences in absolute response latency between the stimulus-evoked response and the cue-evoked response as well as inverse laminar pattern between conditions. (*e*) Spatial profile of suppression. The main stimulus array consisted of 4 identical, isoeccentric gratings. The receptive fields of the

the stimulus in the receptive field (i.e., Fig. 2: "attended vs. unattended"), but the magnitude of suppression decreased slightly with retinotopic distance between the location of the attended target and the recorded neurons' receptive field (Fig. 3*e*; Table 1; Supplementary Fig. 6). In the naïve subject, the onset of the same foveal cue stimulus had no significant impact on neural firing, presumably because it did not evoke internal attention-related operations.

In further assessing the laminar distribution of the attention-related suppression found in the trained subjects, we found that suppression was significantly stronger in the upper, supragranular layers of V1 than in either the middle or lower layers (Fig. 4; significant main effect of layer, Monkey B: $F(19\,720) = 6.01$, P < 0.001; Monkey E: $F(20\,403) = 2.90$, P < 0.001, with electrode contact points in depth as degrees of freedom). The magnitude of transient suppression in the upper layers was roughly twice the magnitude of the suppression in the lower layers and more than twice as large as the attention-related enhancement.

Focusing on the sustained component of attentional response modulation, we computed attentional enhancement, defined as the mean difference in multiunit activity between attended and unattended conditions (see Methods), across all cortical layers (Fig. 4). This enhancement varied as a function of layer ($F(1\,91\,137) = 1.60$, P = 0.049), with the largest response gain located in the superficial and deep layers of the cortical column. Together, these results are consistent with previous findings showing that sustained modulation exerts modest modulation V1 spiking responses, and further show that this modulation is stronger in the feedback-recipient layers (see also Buffalo et al. 2011; Self et al. 2013; van Kerkoerle et al. 2017).

Suppression Associated with Inverted Current Source Density Profile

Next, we used current source density (CSD) analysis to further characterize the neuronal mechanism underlying the observed spiking suppression described above (see Supplementary Figs. 7 and 8 for LFP analyses). Specifically, we computed the laminar CSD profile associated with both the main stimulus onset as well as the onset of the foveal stimulus. Current sinks have been linked to the time-locked occurrence of local excitatory potentials (EPSPs), and thus are informative about synaptic events, even when their combined postsynaptic effects remain below the threshold of spiking activity (Mitzdorf 1985; Shmuel and Maier 2015).

We found that in the trained monkeys performing the task, the onset of the foveal stimulus was followed by a distinct and repeatable pattern of laminar sinks and sources (Fig. 5*a*) that differed from that evoked by the onset of the stimulus within the receptive field (Fig. 5*b*) in two major ways (see Supplementary Fig. 9 for individual monkeys): First, compared with the stimulus onset response profile, the polarity of sources and sinks distributed across the cortical layers following onset of the foveal cue onset was roughly inverted. Unlike the prominent layer 4C current sink associated with the grating stimulus onset, the foveal stimulus was associated with a layer 4C current source that was flanked by current sinks in the superficial and deep layers.

neurons under study always coincided with the grating in the lower right quadrant. On a given trial, the foveal cue could point to any of the 4 gratings in the array. For each monkey E and monkey B, average MUA following cue onset (t = 0.8 s) is plotted for each cue direction (blue = low right quadrant, purple = upper right quadrant, green = upper left quadrant, orange = lower left quadrant; see Supplementary Fig. 6 for schematic).

Table 1 Retinotopy of spiking suppression

| Location of attentional shift	Towards RF	Adjacent Quad	Diagonal Quad	N	
	Post v. Pre Cue (Mean %Δ)			Multiunits	Penetrations
Monkey E	−22.9	−12.2	−10.0	351	25
Monkey B	−24.4	−18.5	−11.7	642	45
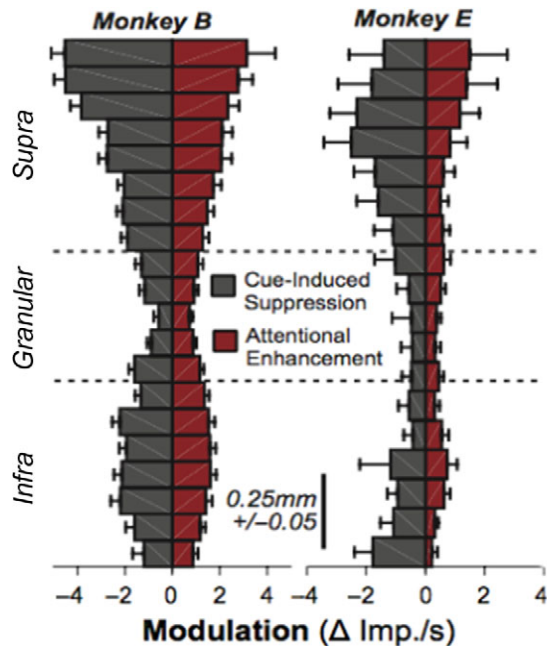
Figure 4. Laminar profile of cue-induced spiking suppression (discretized MUA) compared with attention-related response gain. For each laminar recording location (vertical axis plots cortical depth, with each bar's spacing = 100 μm ± 50 μm), in each monkey (left vs. right), metrics of each cue-induced suppression (mean decrease in spiking 100–200 ms from cue onset) and attention-related response gain (mean difference between attended and unattended condition 300–1600 ms following cue onset) are plotted. Error bars are SEM across penetrations (N = 45 for monkey B, N = 25 for monkey E).

Second, based on visual inspection of the graphs shown in Figure 5a versus b, we estimated that current sinks following the foveal stimulus took about 50 ms longer to appear than those following the stimulus presented to the receptive field. Interestingly, the shortest latency response following the foveal cue was found in the feedback-receiving layers (Rockland and Pandya 1979; Lund 1988), with the superficial layers slightly preceding the deep layers. In contrast, the shortest latency sink for visual stimulation was found in the retino-geniculate-recipient layer 4C, in line with previous literature (Schroeder et al. 1998).

When the same analysis was applied to the naïve subject, no such current sinks were found following the cue onset (Fig. 5c), although the CSD pattern of the stimulus-evoked response was comparable (Fig. 5d). These results suggest that the training on the task fundamentally altered the response to the foveal stimulus, with the first cue-related synaptic activity emerging in the deep and upper layers, possibly evoked by feedback inputs.

Behaviorally Irrelevant Cue Fails to Evoke Spiking Suppression

Attention can be evoked in both an endogenous, goal-driven manner as well as in an exogenous, stimulus-driven way.
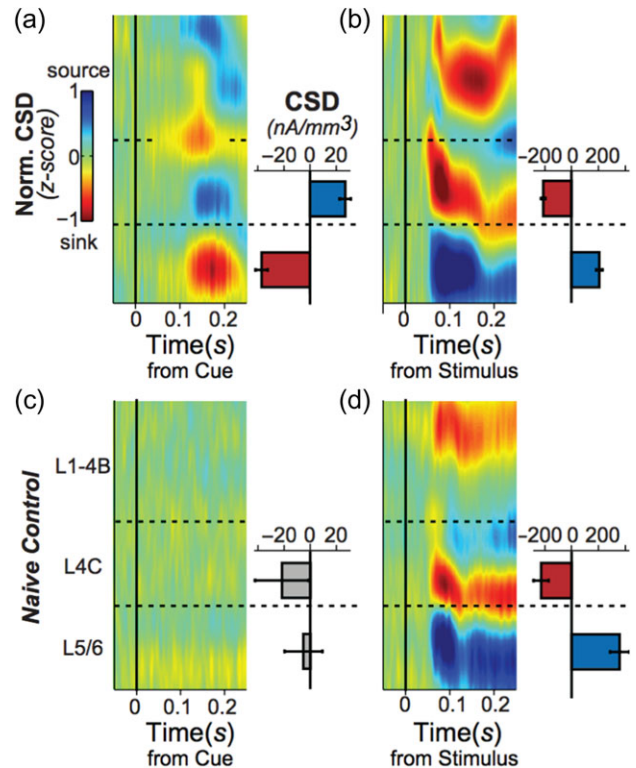
Figure 5. Current-source density (CSD) analysis. (a) Laminar time course of CSD following cue onset expressed as z-score across penetrations for the 2 trained monkeys (N = 70). Bar plot to the right shows the mean CSD amplitude across the cortical depth spanned by the bar's height in units of nanoamperes per mm³. Error bars are SEM across recording sessions. CSD activity differs significantly across the L4C/L5 border (Monkey B: $t(74) = -8.26$, $P < 0.001$; Monkey E: $t(38) = -4.46$, $P < 0.001$), with a current sink occurring in the deep layers and a current source in the middle layers. (b) Laminar time course of CSD following stimulus onset in the trained subjects (N = 70 penetrations, 2 monkeys). All conventions and units as in panel a. CSD activity differs significantly across the L4C/L5 border (Monkey B: $t(74) = 14.38$, $P < 0.001$; Monkey E: $t(38) = 8.85$, $P < 0.001$), with a current sink occurring in the middle layers and a current source in the deep layers, consistent with previous reports. The pronounced current sink marks the granular layer (L4C), the main locus of activation from LGN. (c) Laminar time course of CSD activity following cue onset in the naïve monkey (N = 15 penetrations). There was no significant difference in CSD activity the across the L4C/L5 border ($t(28) = 0.66$, $P = 0.517$). All conventions and units as in panel a. (d) Laminar time course of CSD following stimulus onset across penetrations in a naïve monkey (N = 15 penetrations). CSD activity differs significantly across the L4C/L5 border ($t(28) = 6.45$, $P < 0.001$), with a pronounced current sink in granular L4C. All conventions and units as in panel a.

Exogenous attentional responses to the orienting cue itself have been shown to modulate V1 in a manner that depended on both the predictability as well as the behavioral relevance of the cue (Wang et al. 2015). We thus performed a control experiment to isolate the exogenous attentional response to the cue and study whether it contributes to the cue-induced suppression. To do so, we employed a modified change detection task in which the

onset of the cue at fixation was both redundant and thus behaviorally irrelevant as well as less predictable in time (see Supplementary Fig. 4a–c and Methods for details). Using this redundant cueing control task, we compared responses for attended stimuli in the presence or absence of an abrupt onset of the redundant cue at fixation (Supplementary Fig. 4d). We found that, just as in the naïve subject, there was no suppression following cue onset at fixation (Supplementary Fig. 4e), suggesting that exogenous attention to the cue can be ruled out as a cause of the cue-induced suppression we observed.

Discussion

This study is the first of our knowledge to report two distinct, sequential modulatory components of attentional orienting upon V1 visual responses. Specifically, we found a profound, transient suppression that was followed by a modest response gain to targets of sustained attention. This biphasic response pattern bears similarity to responses observed during attentional shifts in MT, MST and LIP neurons under similar behavioral paradigms (Bisley and Goldberg 2006; Busse et al. 2008; Herrington and Assad 2009, 2010), though the magnitude of the suppression with respect to attentional enhancement observed here is considerably larger. Importantly, our study is first in demonstrating the laminar profile and current source density underlying this suppressive response, indicating a feedback-induced mechanism. Below we discuss possible mechanisms underlying these activity changes and speculate on how the observed suppression in V1 might relate to behavioral phenomena.

Spiking Suppression Preceding Attention-related Response Enhancement

Many cortical and subcortical visual structures exhibit enhanced responses to attended stimuli (Moran and Desimone 1985; Treue and Maunsell 1996; Sheinberg and Logothetis 1997; Corbetta et al. 1998; Roelfsema et al. 1998; Kastner et al. 1999; Kastner and Ungerleider 2000; Reynolds et al. 2000; Schroeder et al. 2001; Marcus and Van Essen 2002; Kastner and Pinsk 2004; Maunsell and Treue 2006; Armstrong and Moore 2007; Herrington and Assad 2009; Nieuwenhuis and Donner 2011; Mishra et al. 2012; Donner and Nieuwenhuis 2013; Peelen and Kastner 2014; Engel et al. 2016), including the primary visual cortex (V1). Enhanced spiking for attended stimuli is consistent with the perceptual benefits of attention, such as increased performance for stimulus detection and discrimination at the attended location, that help prioritize stimuli and plan actions. Our findings are in line with this view, as well as most studies describing rather modest attention-related gain in V1 spiking compared with that observed in other retinotopically organized areas such as V4 and MT.

We show that immediately following the covert attention cue, and "prior" to the sustained enhancement of neural responses to the attended target, there is a sharp decrease in firing rate of V1 neurons. This cue-induced suppression was transient, about 200 ms in duration. Transient depressions can be caused by a consistent phase-reset of broadband ongoing oscillatory activity, though it is challenging to analytically determine when this is the case (Fell et al. 2004; Hanslmayr et al. 2006; Mazaheri and Jensen 2006; Sauseng et al. 2007). With respect to magnitude, the cue-induced suppression was on average more than twice that of the subsequent attention-related enhancement. This profound decrease in spiking persisted regardless of the presence or absence of small, fixational

eye movements in the period around cueing. Furthermore, this transient suppression of visual responses was observed regardless of the direction of the attentional shift relative to the receptive field under study, although its magnitude systematically decreased with retinotopic distance. This observation suggests a widespread inhibition of spiking activity across V1's retinotopic representation of the visual field that may be gated spatially (Donner et al. 2013).

The decreased spiking rate in our paradigm was time-locked to the onset of a spatial cue at fixation. While this cue was small and quite remote from the receptive field, we needed to consider the possibility that the suppressed visual response was due to extra-classical receptive field effects driven by the cue onset itself. This possibility motivated the control experiments with the naïve subject. Our observation that there was no trace of cue-related activity changes in the naïve subject largely rules out the cue as a sensory driver of the suppression.

Thus, we consider that the suppression is a result of attentional processes that were triggered by cue onset. Orienting of spatial attention is thought to require several steps: disengagement from current focus, shift to new location and focusing on a new stimulus (Posner 2016). If we take the subsequent attention-related enhancement as a signature of attentional focus on the indicated target, then it is possible that suppression results from disengaging and shifting attention from its current focus to the cued target.

In the attention task, the contrast change—detection of which granted the subjects a reward—always occurred after onset of the orienting cue at fixation. Given that the subjects were not tasked to produce a behavioral report before cue onset, we cannot determine with certainty the locus of spatial attention during this phase of the task. One possibility is that before cue onset the subjects' attention followed their gaze, which was fixed on the central fixation dot. In this case, the observed suppression may be associated with the disengagement of covert attention from the locus of gaze and the shift of attention towards the cued target (see sections Comparison with Saccadic Suppression and Potential Relationship to Behavioral Phenomena). However, given the absence of behavioral report from the pre-cue portion of the task, we cannot answer with certainty whether the subjects' attention was focused at fixation.

Another possibility is that the onset of the stimulus array reflexively drew spatial attention to the periphery. In this case, the cue's onset at fixation drew the trained subjects' attention away from the peripheral targets because it was task relevant, i.e., "meaningful". According to this view, suppression of visual responses to the targets would result from transiently changing sustained attentional focus from the main peripheral stimulus to the central cue. Thus, the observed suppression would be best explained by a brief interruption in attentional enhancement, rather than suppression per se (see section Is Spiking Suppression Due to a Withdrawal of Attention?).

Population-Level Measures in V1

In this study, we use multiunit activity as a measure of neuronal spiking activity to ensure even sampling from all cortical depths simultaneously (see Methods). Multiunit activity has been estimated to carry a signature of the spiking activity of up to a few thousand cells located within 140–300 μm of an electrode contact (Super et al. 2001). As a result, a given multiunit contains a mixture of neuronal types, including excitatory and inhibitory neurons. The diversity across the individual neuronal responses that

comprise this signal is unknown, and it is unlikely that all the comprising signals follow the pattern of activity seen at the multiunit level. Thus, it is important to note that some V1 neurons exhibit a different response pattern during our task than the one reported here.

Possible Origins of Response Suppression During Attentional Orienting

Each of the attention-related response components in the present study had a distinct laminar profile, which can shed some light on mechanistic aspects related to the origin of the measured signals. We were particularly interested in the origin of the apparent suppression concurrent with the attentional shift. Here we consider 3 distinct possibilities that might account for our observed changes in spiking and CSD.

The first possibility is that the dip in spiking is due to a transient decrease in driving input from the LGN to layer 4C (Vanduffel et al. 2000). In other words, the LGN might be the primary site of attentional modulation and the decrease in spiking in V1 reported here is a consequence of reduced retinogeniculate drive. LGN neurons exhibit increased firing rates and synaptic efficacy for selectively attended stimuli (Briggs et al. 2013). V1 neurons provide organized feedback projections to the LGN, whereas other cortical visual areas lack such connections (Casagrande et al. 2005). However, the adjacent thalamic reticular nucleus (TRN) receives inputs from both thalamus and cortex and sends topographically organized inhibitory projections to the LGN (Conley and Diamond 1990). Thus, in theory, the TRN has the capacity to regulate LGN responses. Previous studies demonstrate that focused attention can decrease visual responses in the TRN, which, in turn, can increase LGN activity (McAlonan et al. 2008). Thus, it is possible that in our task, prior to attentional engagement at a target, an attentional shift could mediate an opposing cascade of activations between the TRN and LGN. Specifically, a transient increase in TRN activity momentarily silences visual responses in the LGN while attentional focus moves between spatially distinct targets (Crick 1984). While such a mechanism is theoretically possible, our CSD findings are difficult to reconcile with that view. The earliest modulation we observed during the attentional shift (both sinks and sources) was superficial to layer 4C, in the supragranular layers rather than in the retino-geniculate input layer. In other words, the attention-related activity in V1 was initiated outside the main LGN input layers, largely ruling out an account invoking bottom-up via the LGN.

The second possibility is that V1 modulation is caused by the pulvinar, the other main visual nucleus of the thalamus. The pulvinar's role for visual processing is understood to a lesser degree, but its modulatory connections to the superficial layers of V1 make it a candidate to consider for the observed modulation (Ogren and Hendrickson 1977; Kaas and Lyon 2007). Modulatory effects of the pulvinar with attention have been reported in cortical low frequency (10–25 Hz) activity (Wilke et al. 2009), which seems to match the spectral response pattern we observed during the attentional shift. While we cannot rule out entirely a contribution of the pulvinar, its V1 input is directed primarily to layers 1/2 (Ogren and Hendrickson 1977), which is somewhat removed from the observed current sink in the layers right above the main geniculocortical input layer 4C.

The third possibility is that the modulation we see is caused by long-range feedback projections. Feedback is known to reach V1 within less than 100 ms (Angelucci and Bullier 2003), which matches the time courses observed in our study. Top-down activation from higher areas can lead to a local reduction of spiking responses (Moore and Armstrong 2003; Bisley et al. 2004), either via direct inhibitory action (Reed et al. 2011) or by activation of local interneurons (Mitchell et al. 2007; Liu et al. 2013). The initial CSD response points to the layers directly superficial to layer 4C as the initial site of action, which is consistent with the fact that cortico-cortical feedback connections target these layers (Rockland and Pandya 1979; Lund 1988). Thus, the laminar distribution of spiking and current source density (CSD) during this period, taken together, is consistent with the assumption that this signal is initiated via top-down signals from higher cortical areas. Broadly speaking, these results suggest that cortico-cortical feedback to area V1 does not only enhance neuronal responses, but can also cause transient disruption of sensory activity.

Is Spiking Suppression Due to a Withdrawal of Attention?

Another consideration is that the cue's onset at fixation drew the trained subjects' attention away from the peripheral targets. According to this view, suppression of visual responses to the targets results from transiently deflecting sustained attentional focus from the main peripheral stimuli to the more central cue, resulting in reduced feedback-induced synaptic drive. In other words, the observed spiking suppression might be best explained by a brief "interruption" of prior feedback-induced enhancement that is constantly spread across the visual field, rather than active inhibition.

It is important to note that feedback activation is the root cause of the drop in spiking we observed, regardless of whether the spiking change is due to an increase or decrease of top-down modulation of peripheral V1. However, the feedback-driven mechanisms that institute the spiking drop might differ between these two scenarios. The first possibility is that feedback inputs to V1's periphery directly affect spiking, possibly by activating local interneurons and thus increasing local inhibition (see Mitchell et al. 2007). Another, intriguing possibility is that there are reciprocal connections between foveal and peripheral V1 that give rise to a push-pull competition between them. In the latter case, increased feedback activation of foveal V1 due to attentional reorienting towards the cue might tip the scale and reduce peripheral V1 spiking via V1-internal mechanisms. That is, the peripheral drop in spiking is feedback-induced, yet still primarily driven by V1-internal mechanisms such as horizontal connections.

The interpretation that the observed peripheral suppression is caused by increased feedback to foveal V1 is consistent with our findings in the naïve subject and other controls in that endogenous attentional selection of the foveal region might evoke a putative mechanism of foveal-to-peripheral competition. However, we found that the CSD evoked by an additional peripheral stimulus outside the receptive field did not resemble that evoked by the learned foveal cue. Thus, this interpretation rests on an assumption that peripheral-to-peripheral competition within V1 differs fundamentally from foveal-to-peripheral competition. Putative specialization for foveal-to-peripheral competition rests on theoretical considerations regarding the close relationship between attentional selection and eye movements that cause foveation in primates (Rizzolatti et al. 1987; Clark 1999; Ignashchenkova et al. 2004; Pinsk et al. 2004; de Haan et al. 2008; Zénon et al. 2014) (but see Thompson et al. 2005; Smith and Schenk 2012), and is partially supported by the finding of competitive interactions between foveal and

peripheral saccade-related neurons in the superior colliculus (Munoz and Guitton 1991; Munoz and Wurtz 1995a; 1995b but see Goffart et al. 2012).

Comparison with Saccadic Suppression

In contrast to covert attentional shifts, overt orienting of attention involves moving the gaze to the location of the attended target. This link between attention and action is formalized by the premotor theory of attention, which proposes that spatial attentional orienting is controlled by the same mechanism responsible for motor orientating towards specific spatial locations. In this view, our covert shift-related suppression might be similar to saccadic suppression, whereby visual sensitivity decreases around the time of an eye movement (Rizzolatti et al. 1987; Steinmetz and Moore 2014). Interestingly, recent work has demonstrated that saccades produce a biphasic modulation of V1 neuron firing rates where spiking first decreases and then increases (McFarland et al. 2015). While the polarity of this saccade-evoked activity matches the activity pattern evoked here, the latency of saccadic activity is considerably faster, reaching peak suppression approximately 50 ms following the saccade and peak enhancement approximately 120 ms following the saccade. Strikingly, the CSD pattern evoked by saccades in V1 is the inverse of the attention-evoked pattern observed here, with a prominent current source in the granular, retinogeniculate input layer. Neurons in the granular layer also exhibit the shortest response latencies and strongest saccadic response modulation. By contrast, attention cue onset maximally modulates neurons in V1's extragranular layers, which predominantly project to other cortical targets (Martinez-Millán and Holländer 1975). Therefore, while the polarity of the spiking modulation is similar between saccades and attentional modulation, the spatiotemporal profile of spiking and CSD activity with the V1 cortical column suggests different origins for the two processes.

Potential Relationship to Behavioral Phenomena

One open question is if the observed suppression in activity has behavioral relevance for deploying covert spatial attention more generally. Since we did not test the subjects' performance concurrent with the decrease in V1 spiking, we cannot answer directly what the perceptual consequences of the transient suppression might be. However, it is reasonable to assume that a decrease in V1 spiking for any given stimulus could be detrimental to perceptual sensitivity. Based on our findings, we would expect such a perceptual detriment to be spatially non-specific and brief. Psychophysically, brief and spatially non-specific detriments in perceptual sensitivity have been observed in tasks that require rapid, sequential deployments of endogenous (goal-oriented), covert spatial attention (Duncan et al. 1994). These so-called "dwell time paradigms" have typically been used to measure the time required to move attention to various target locations across the visual field. Attentional dwell times using endogenous, spatial cueing appear around 250 ms following stimulus onset (e.g., Theeuwes et al. 2004), which approximately matches the timing of the suppressive response described in our study. Furthermore, the effects of attentional dwell can be observed across the entire visual field (Duncan et al. 1994). The relatively late latency of the suppressive effects of attentional shifting has previously been interpreted as evidence against involvement of sensory cortex.

Indeed, a number of behavioral and neuroimaging studies suggest that processing limitations associated with attention occur within fronto-parietal areas (Vogel et al. 1998), though the evidence implicating these regions in attentional dwell is mixed (Duncan et al. 1994). The laminar pattern of our results suggests that the attentional cueing-related decrease in V1 activity is feedback mediated, consistent with the idea that attentional dwell could be instantiated by an additional feedback-induced component that suppresses sensory responses at the earliest stage of cortical processing.

Conclusion

Understanding the role of feedback to sensory cortex is critical for determining how sensation and cognition interact on the neuronal level. In this study, we demonstrate that V1 spiking activity first decreases and then increases following the onset of a behaviorally meaningful cue far outside the neurons' receptive field. The laminar profiles of both the initial suppression and the successive enhancement following attentional cueing suggest that both modulatory effects are mediated by feedback projections. By demonstrating that feedback to V1 can biphasically induce both negative as well as positive spike rate changes, this finding offers important implications for both our understanding of covert attentional orienting and top-down modulation of sensory cortex during visual cognition more generally.

Authors' Contributions

A.M., D.A.L., and M.A.C. conceptualized and designed the study. A.M., M.A.C., G.K.A., and E.A.R. implemented the attention task, trained subjects, and analyzed behavioral data. K.D. and M.A.C trained the naïve subject and analyzed eye movement data. A.M., M.A.C., B.S.M., J.A.W and K.D. collected neurophysiological data. A.M. and M.A.C. analyzed the neurophysiological data. A.M., D.A.L., and M.A.C. wrote the manuscript.

Supplementary Material

Supplementary data is available at *Cerebral Cortex* online.

References

Angelucci A, Bullier J. 2003. Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? J Physiol-Paris. 97(2):141–154.

Armstrong KM, Moore T. 2007. Rapid enhancement of visual cortical response discriminability by microstimulation of the frontal eye field. PNAS. 104:9499–9504.

Asaad WF, Eskandar EN. 2008. A flexible software tool for temporally-precise behavioral control in Matlab. J Neurosci Methods. 174:245–258.

Asaad WF, Santhanam N, McClellan S, Freedman DJ. 2013. High-performance execution of psychophysical tasks with complex visual stimuli in MATLAB. J Neurophysiol. 109:249–260.

Bisley JW, Goldberg ME. 2006. Neural correlates of attention and distractibility in the lateral intraparietal area. J Neurophysiol. 95:1696–1717.

Bisley JW, Zaksas D, Droll JA, Pasternak T. 2004. Activity of neurons in cortical area MT during a memory for motion task. J Neurophysiol. 91:286–300.

Bosman CA, Schoffelen J-M, Brunet N, Oostenveld R, Bastos AM, Womelsdorf T, Rubehn B, Stieglitz T, De Weerd P, Fries P. 2012. Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron. 75:875–888.

Briggs F, Mangun GR, Usrey WM. 2013. Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. Nature. 499:476–480.

Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R. 2011. Laminar differences in gamma and alpha coherence in the ventral stream. PNAS. 108:11262–11267.

Busse L, Katzner S, Treue S. 2008. Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. PNAS. 105: 16380–16385.

Casagrande VA, Royal DW, Sary G 2005. The Primate Visual System: A Comparative Approach (ed. Kremers, J.) 191–200.

Chen Y, Seidemann E. 2012. Attentional modulations related to spatial gating but not to allocation of limited resources in primate V1. Neuron. 74:557–566.

Clark JJ. 1999. Spatial attention and latencies of saccadic eye movements. Vision Res. 39:585–602.

Conley M, Diamond IT. 1990. Organization of the visual sector of the thalamic reticular nucleus in Galago. Europ J Neurosci. 2(3):211–226.

Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van Essen DC, et al. 1998. A common network of functional areas for attention and eye movements. Neuron. 21:761–773.

Cox MA, Schmid MC, Peters AJ, Saunders RC, Leopold DA, Maier A. 2013. Receptive field focus of visual area V4 neurons determines responses to illusory surfaces. Proc Natl Acad Sci USA. 110(42):17095–17100.

Crick F. 1984. Function of the thalamic reticular complex: the searchlight hypothesis. PNAS. 81:4586–4590.

Donner TH, Nieuwenhuis S. 2013. Brain-wide gain modulation: the rich get richer. Nat Neurosci. 16:989–990.

Donner TH, Sagi D, Bonneh YS, Heeger DJ. 2013. Retinotopic patterns of correlated fluctuations in visual cortex reflect the dynamics of spontaneous perceptual suppression. J Neurosci. 33(5):2188–2198.

Dougherty K, Cox MA, Ninomiya T, Leopold DA, Maier A. 2015. Ongoing alpha activity in V1 regulates visually driven spiking responses. Cereb Cortex. 27(2):1113–1124.

Duncan J, Ward R, Shapiro K. 1994. Direct measurement of attentional dwell time in human vision. Nature. 369:313–315.

Engel TA, Steinmetz NA, Gieselmann MA, Thiele A, Moore T, Boahen K. 2016. Selective modulation of cortical state during spatial attention. Science. 354:1140–1144.

Fell J, Dietl T, Grunwald T, Kurthen M, Klaver P, Trautner P, Schaller C, Elger CE, Fernández G. 2004. Neural bases of cognitive ERPs: more than phase reset. J Cog Neurosci. 16(9): 1595–1604.

Freeman J, Simoncelli EP. 2011. Metamers of the ventral stream. Nat Neurosci. 14:1195–1201.

Goffart L, Hafed ZM, Krauzlis RJ. 2012. Visual fixation as equilibrium: evidence from superior colliculus inactivation. J Neurosci. 32:10627–10636.

de Haan B, Morgan PS, Rorden C. 2008. Covert orienting of attention and overt eye movements activate identical brain regions. Brain Res. 1204:102–111.

Haenny PE, Schiller PH. 1988. State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. Exp Brain Res. 69:225–244.

Hanslmayr S, Klimesch W, Sauseng P, Gruber W, Doppelmayr M, Freunberger R, Pecherstorfer T, Birbaumer N. 2006. Alpha phase reset contributes to the generation of ERPs. Cereb Cortex. 17(1):1–8.

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

Herrington TM, Assad JA. 2009. Neural activity in the middle temporal area and lateral intraparietal area during endogenously cued shifts of attention. J Neurosci. 29:14160–14176.

Herrington TM, Assad JA. 2010. Temporal sequence of attentional modulation in the lateral intraparietal area and middle temporal area during rapid covert shifts of attention. J Neurosci. 30:3287–3296.

Herrmann K, Montaser-Kouhsari L, Carrasco M, Heeger DJ. 2010. When size matters: attention affects performance by contrast or response gain. Nat Neurosci. 13:1554–1559.

Ignashchenkova A, Dicke PW, Haarmeier T, Thier P. 2004. Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. Nat Neurosci. 7:56–64.

Ito M, Gilbert CD. 1999. Attention modulates contextual influences in the primary visual cortex of alert monkeys. Neuron. 22:593–604.

Kaas JH, Lyon DC. 2007. Pulvinar contributions to the dorsal and ventral streams of visual processing in primates. Brain Res Rev. 55:285–296.

Kastner S, Pinsk MA. 2004. Visual attention as a multilevel selection process. Cogn Affect Behav Neurosci. 4:483–500.

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG. 1999. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron. 22:751–761.

Kastner S, Ungerleider LG. 2000. Mechanisms of visual attention in the human cortex. Ann Rev Neurosci. 23:315–341.

Liu Y-J, Ehrengruber MU, Negwer M, Shao H-J, Cetin AH, Lyon DC. 2013. Tracing inputs to inhibitory or excitatory neurons of mouse and cat visual cortex with a targeted rabies virus. Curr Biol. 23:1746–1755.

Logothetis NK, Kayser C, Oeltermann A. 2007. In vivo measurement of cortical impedance spectrum in monkeys: implications for signal propagation. Neuron. 55:809–823.

Luck SJ, Chelazzi L, Hillyard SA, Desimone R. 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol. 77:24–42.

Lund JS. 1988. Anatomical organization of macaque monkey striate visual cortex. Annu Rev Neurosci. 11:253–288.

Maier A, Adams GK, Aura C, Leopold DA. 2010. Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. Front Syst Neurosci. 4:1–11.

Maier A, Aura CJ, Leopold DA. 2011. Infragranular sources of sustained local field potential responses in macaque primary visual cortex. J Neurosci. 31:1971–1980.

Maier A, Wilke M, Aura C, Zhu C, Ye FQ, Leopold DA. 2008. Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. Nat Neurosci. 11:1193–1200.

Marcus DS, Van Essen DC. 2002. Scene segmentation and attention in primate cortical areas V1 and V2. J Neurophysiol. 88 (5):2648–2658.

Martinez-Millán L, Holländer H. 1975. Cortico-cortical projections from striate cortex of the squirrel monkey (Saimiri sciureus). A radioautographic study. Brain Res. 83:405–417.

Maunsell JHR, Treue S. 2006. Feature-based attention in visual cortex. TINS. 29:317–322.

Mazaheri A, Jensen O. 2006. Posterior α activity is not phase-reset by visual stimuli. PNAS. 103(8):2948–2952.

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

McAdams CJ, Reid RC, McAdams CJ. 2005. Attention modulates the responses of simple cells in monkey primary visual cortex. J Neurosci. 25:11023–11033.

McAlonan K, Cavanaugh J, Wurtz RH. 2008. Guarding the gateway to cortex with attention in visual thalamus. Nature. 456:391–394.

McFarland JM, Bondy AG, Saunders RC, Cumming BG, Butts DA. 2015. Saccadic modulation of stimulus processing in primary visual cortex. Nat Commun. 6:8110.

Mishra J, Martínez A, Schroeder CE, Hillyard SA. 2012. Spatial attention boosts short-latency neural responses in human visual cortex. Neuroimage. 59:1968–1978.

Mitchell JF, Sundberg KA, Reynolds JH. 2007. Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron. 55:131–141.

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.

Mitzdorf U, Singer W. 1979. Excitatory synaptic ensemble properties in the visual cortex of the macaque monkey: a current source density analysis of electrically evoked potentials. J Comp Neurol. 187:71–83.

Moore T, Armstrong KM. 2003. Selective gating of visual signals by microstimulation of frontal cortex. Nature. 421:370–373.

Moran J, Desimone R. 1985. Selective attention gates visual processing in the extrastriate cortex. Science. 229:782–784.

Motter BC. 1993. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J Neurophysiol. 70:909–919.

Munoz DP, Guitton D. 1991. Control of orienting gaze shifts by the tectoreticulospinal system in the head-free cat. II. Sustained discharges during motor preparation and fixation. J Neurophysiol. 66:1624–1641.

Munoz DP, Wurtz RH. 1995a. Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. J Neurophysiol. 73:2313–2333.

Munoz DP, Wurtz RH. 1995b. Saccade-related activity in monkey superior colliculus. II. Spread of activity during saccades. J Neurophysiol. 73:2334–2348.

Nicholson C, Freeman JA. 1975. Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. J Neurophysiol. 38:356–368.

Nieuwenhuis S, Donner TH. 2011. The visual attention network untangled. Nat Neurosci. 14:542–543.

Ninomiya T, Dougherty K, Godlove DC, Schall JD, Maier A. 2015. Microcircuitry of agranular frontal cortex: contrasting laminar connectivity between occipital and frontal areas. J Neurophysiol. 113(9):3242–3255.

Ogren MP, Hendrickson AE. 1977. The distribution of pulvinar terminals in visual areas 17 and 18 of the monkey. Brain Res. 137:343–350.

Otero-Millan J, Castro JLA, Macknik SL, Martinez-Conde S. 2014. Unsupervised clustering method to detect microsaccades. JoV. 14:18–18.

Peelen MV, Kastner S. 2014. Attention in the real world: toward understanding its neural basis. TINS. 18:242–250.

Pettersen KH, Devor A, Ulbert I, Dale AM, Einevoll GT. 2006. Current-source density estimation based on inversion of electrostatic forward solution: effects of finite extent of neuronal activity and conductivity discontinuities. J Neurosci Methods. 154:116–133.

Pinsk MA, Doniger GM, Kastner S. 2004. Push-pull mechanism of selective attention in human extrastriate cortex. J Neurophysiol. 92:622–629.

Poort J, Roelfsema PR. 2009. Noise correlations have little influence on the coding of selective attention in area V1. Cereb Cortex. 19:543–553.

Posner MI. 2016. Orienting of attention: then and now. Q J Exp Psychol (Hove). 69:1864–1875.

Posner MI, Gilbert CD. 1999. Attention and primary visual cortex. PNAS. 96:2585–2587.

Reed JL, Qi H-X, Kaas JH. 2011. Spatiotemporal properties of neuron response suppression in owl monkey primary somatosensory cortex when stimuli are presented to both hands. J Neurosci. 31(10):3589–3601.

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

Rizzolatti G, Riggio L, Dascola I, Umiltá C. 1987. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. Neuropsychologia. 25:31–40.

Rockland KS, Pandya DN. 1979. Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res. 179:3–20.

Roelfsema PR, Lamme VA, Spekreijse H. 1998. Object-based attention in the primary visual cortex of the macaque monkey. Nature. 395:376–381.

Ruff DA, Cohen MR. 2016. Stimulus dependence of correlated variability across cortical areas. J Neurosci. 36:7546–7556.

Sauseng P, Klimesch W, Gruber WR, Hanslmayr S, Freunberger R, Doppelmayr M. 2007. Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. Neuroscience. 146(4):1435–1444.

Schroeder CE, Mehta AD, Foxe JJ. 2001. Determinants and mechanisms of attentional modulation of neural processing. Front Biosci. 6:D672–D684.

Schroeder CE, Mehta AD, Givre SJ, Schroeder C. 1998. A spatio-temporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb Cortex. 8:575–592.

Self MW, van Kerkoerle T, Supèr H, Roelfsema PR. 2013. Distinct roles of the cortical layers of area V1 in figure-ground segregation. Curr Biol. 23:2121–2129.

Sheinberg DL, Logothetis NK. 1997. The role of temporal cortical areas in perceptual organization. PNAS. 94:3408–3413.

Shmuel A, Maier A 2015. Locally Measured Neuronal Correlates of Functional MRI Signals. In: FMRI: From Nuclear Spins to Brain Functions. Biological Magnetic Resonance. Boston, MA: Springer US. p. 105–128.

Siegel M, Donner TH, Oostenveld R, Fries P, Engel AK. 2008. Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. Neuron. 60:709–719.

Smith DT, Schenk T. 2012. The Premotor theory of attention: time to move on? Neuropsychologia. 50:1104–1114.

Spaak E, Bonnefond M, Maier A, Leopold D, Jensen O. 2012. Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. Curr Biol. 22: 2313–2318.

Steinmetz NA, Moore T. 2014. Eye movement preparation modulates neuronal responses in area V4 when dissociated from attentional demands. Neuron. 83:496–506.

Supèr H, Roelfsema PR. 2005. Chronic multiunit recordings in behaving subjects: advantages and limitations. Prog Brain Res. 147:263–282.

Supèr H, Spekreijse H, Lamme VA. 2001. Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat Neurosci. (3):304–310.

Theeuwes J, Godijn R, Pratt J. 2004. A new estimation of the duration of attentional dwell time. Psychon Bull Rev. 11: 60–64.

Thiele A, Pooresmaeili A, Delicato LS, Herrero JL, Roelfsema PR. 2009. Additive effects of attention and stimulus contrast in primary visual cortex. Cereb Cortex. 19:2970–2981.

Thompson KG, Biscoe KL, Sato TR. 2005. Neuronal basis of covert spatial attention in the frontal eye field. J Neurosci. 25:9479–9487.

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

van Kerkoerle T, Self MW, Roelfsema PR. 2017. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. Nat Commun. 8:13804.

Vanduffel W, Tootell RB, Orban GA. 2000. Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. Cereb Cortex. 10:109–126.

Vogel EK, Luck SJ, Shapiro KL. 1998. Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. J Exp Psychol Hum Percept Perform. 24:1656–1674.

Wang F, Chen M, Yan Y, Zhaoping L, Li W. 2015. Modulation of neuronal responses by exogenous attention in macaque primary visual cortex. J Neurosci. 35:13419–13429.

Wilke M, Mueller K-M, Leopold DA. 2009. Neural activity in the visual thalamus reflects perceptual suppression. PNAS. 106: 9465–9470.

Zénon A, Corneil BD, Alamia A, Filali-Sadouk N, Olivier E. 2014. Counterproductive effect of saccadic suppression during attention shifts. PLoS ONE. 9:e86633.