Directed information exchange between cortical layers in macaque V1 and V4 and its modulation by selective attention

Demetrio Ferro[1,2], Jochem van Kempen[3], Michael Boyd[3], Stefano Panzeri[1,4] and Alexander Thiele[3,4+]

[1] Neural Computation Laboratory, Istituto Italiano di Tecnologia, Rovereto, Italy
[2] Center for Mind and Brain Sciences (CIMeC), University of Trento, Rovereto, Italy
[3] Biosciences Institute, Newcastle University, NE1 7RU, Newcastle upon Tyne, United Kingdom
[4] Senior and corresponding authors: stefano.panzeri@iit.it and alex.thiele@ncl.ac.uk
[+]Lead Contact

Keywords: Feedforward processing, feedback processing, visual cortex, Granger causality, attention, cortical laminae

Summary

Achieving behavioral goals requires integration of sensory and cognitive information, across cortical laminae and cortical regions. How this computation is performed remains unknown. Using local field potential recordings and spectrally resolved conditional Granger causality (cGC) analysis, we mapped visual information flow, and its attentional modulation, between cortical layers within and between macaque areas V1 and V4. Stimulus induced inter-laminar information flow within V1 dominated upwardly, channeling information towards supragranular cortico-cortical output layers. Within V4, information flow dominated from granular to supragranular layers, but interactions between supragranular and infragranular layers dominated downwardly. Low-frequency across-area communication was stronger from V4 to V1, with little layer specificity. Gamma-band communication was stronger in the feedforward V1 to V4 direction. Attention to the receptive field of V1 decreased communication between all V1 layers, except for granular to supragranular layers interactions. Communication within V4, and from V1 to V4, increased with attention across all frequencies. While communication from V4 to V1 was stronger in lower frequency bands (4-25 Hz), attention modulated cGCs from V4 to V1 across all investigated frequencies. Our data show that top down cognitive processes result in reduced communication within cortical areas, increased feedforward communication across all frequency bands and increased gamma band feedback communication.

Introduction

Goal-directed behavior requires the brain to integrate sensory information with cognitive variables. In neocortical areas sensory information is conveyed by feedforward connections, while feedback connections convey information about cognitive states and goals. Feedforward and feedback connections rely on separate anatomical pathways and have been proposed to map onto distinct oscillatory frequency bands. It is, however, unknown whether these signals differ across laminae, or how they are communicated between laminae within and between cortical areas.

Feedforward connections predominantly terminate in layer IV of sensory cortical areas. This information is passed on to layers II/III and further to layers V/VI, where recurrent inputs to layer II/III arise (Callaway, 1998; Callaway, 2004; Douglas *et al.*, 1989; Douglas and Martin, 2004). Cognitive variables affect sensory processing through feedback connections, which predominantly terminate in layer I and V (Rockland and Pandya, 1979), but this termination pattern varies depending on hierarchical distances between areas (Markov *et al.*, 2014). Feedforward and feedback signals have been proposed to show separate local field potential (LFP) spectral signatures. Feedforward signals have been associated with low-frequency theta (Bastos *et al.*, 2015; Spyropoulos *et al.*, 2018) and gamma band oscillatory activity, originating and dominating in supragranular layers (Bastos *et al.*, 2015; Bollimunta *et al.*, 2011; Buschman and Miller, 2007; Lakatos *et al.*, 2008; Maier *et al.*, 2010; Smith *et al.*, 2013; Spaak *et al.*, 2012; Spyropoulos *et al.*, 2018; van Kerkoerle *et al.*, 2014; Xing *et al.*, 2012). Feedback signals have been associated with lower frequency (alpha, beta band) oscillations, prominent in infragranular layers across the cortical hierarchy (Bastos *et al.*, 2015; Buffalo *et al.*, 2011; Buschman and Miller, 2007; Popov *et al.*, 2017; Smith *et al.*, 2013; van Kerkoerle *et al.*, 2014; Xing *et al.*, 2012), although attention-related feedback signals in the gamma frequency band between FEF and V4 have been reported (Gregoriou *et al.*, 2009; Gregoriou *et al.*, 2012). Alpha related feedback has been linked to active inhibition (Bagherzadeh *et al.*, 2020; Zumer *et al.*, 2014), suggesting that feedback signals, induced by attention to the receptive field, should result in reduced alpha oscillatory power. This occurs in infragranular layers in visual areas (Buffalo *et al.*, 2011), but can also be less layer specific (van Kerkoerle *et al.*, 2014). It is thus questionable whether feedback is characterizable by alpha frequencies as attention, employing feedback, shunts alpha oscillations, instead of using them. In extrastriate sensory areas, attention increases LFP power in the gamma frequency band (Bosman *et al.*, 2012; Chalk *et al.*, 2010; Fries *et al.*, 2001; Gregoriou *et al.*, 2009; Gregoriou *et al.*, 2012; Grothe *et al.*, 2012; Grothe *et al.*, 2018; Richter *et al.*, 2017; Taylor *et al.*, 2005), while in primary visual cortex attention can increase

or decrease LFP power in the gamma frequency band (Bastos *et al.*, 2015; Bosman *et al.*, 2012; Buffalo *et al.*, 2010; Buffalo *et al.*, 2011; Chalk *et al.*, 2010).

Many of the above results were obtained by methods which do not provide insight how these signals differ between laminae within an area, or between laminae across different areas. Thus, it remains unclear whether layer differences in these signals between cortical areas exist, and whether they are differently affected by cognitive goals.

To understand how information within and between areas is conveyed as a function of cognitive task, we performed simultaneous laminar recordings in areas V1 and V4 using 16-contact laminar probes while macaque monkeys performed a feature based spatial attention task. We quantified communication between laminae and areas using Granger Causality.

Results

Monkeys performed a covert, top-down, feature guided spatial attention task. On each trial attention was directed by a central colored cue to one of 3 possible locations in a pseudo-randomized manner (Figure 1A). Monkeys had to detect a stimulus change at the cued location and ignore changes at uncued locations. To investigate how spatial attention affects interactions within cortical columns and between cortical columns of areas V1 and V4, we recorded Local Field Potentials (LFP), using 16 channel laminar probes (150 µm inter-contact spacing) in 2 adult male monkeys (62 sessions in total: 34 for monkey 1, 28 for monkey 2). We inserted probes perpendicular to the cortical surface (Figure 1B). The depth of recording contacts relative to cortical layers was determined by computing the LFP current source density (CSD, Figure 1C) (Nicholson, 1973; Nicholson and Freeman, 1975) and the multi-unit response latency (Roelfsema et al., 2007). The earliest current sink of the CSD and the shortest response latency identified input layer IV (Figure 1C). Recording sites superficial to the input layer contacts were defined as supragranular layers (L I/II/III), deeper sites were defined as infragranular layers (L V/VI) (for exact assignments see Methods) (Bollimunta *et al.*, 2008; Lakatos *et al.*, 2008; Self *et al.*, 2013; Nandy *et al.*, 2017; van Kerkoerle *et al.*, 2014). Data were analyzed for sessions in which V1 and V4 receptive fields (RF) overlapped (supplemental information for details), although centre-to-centre RF positioning could be offset (Figure 1D).

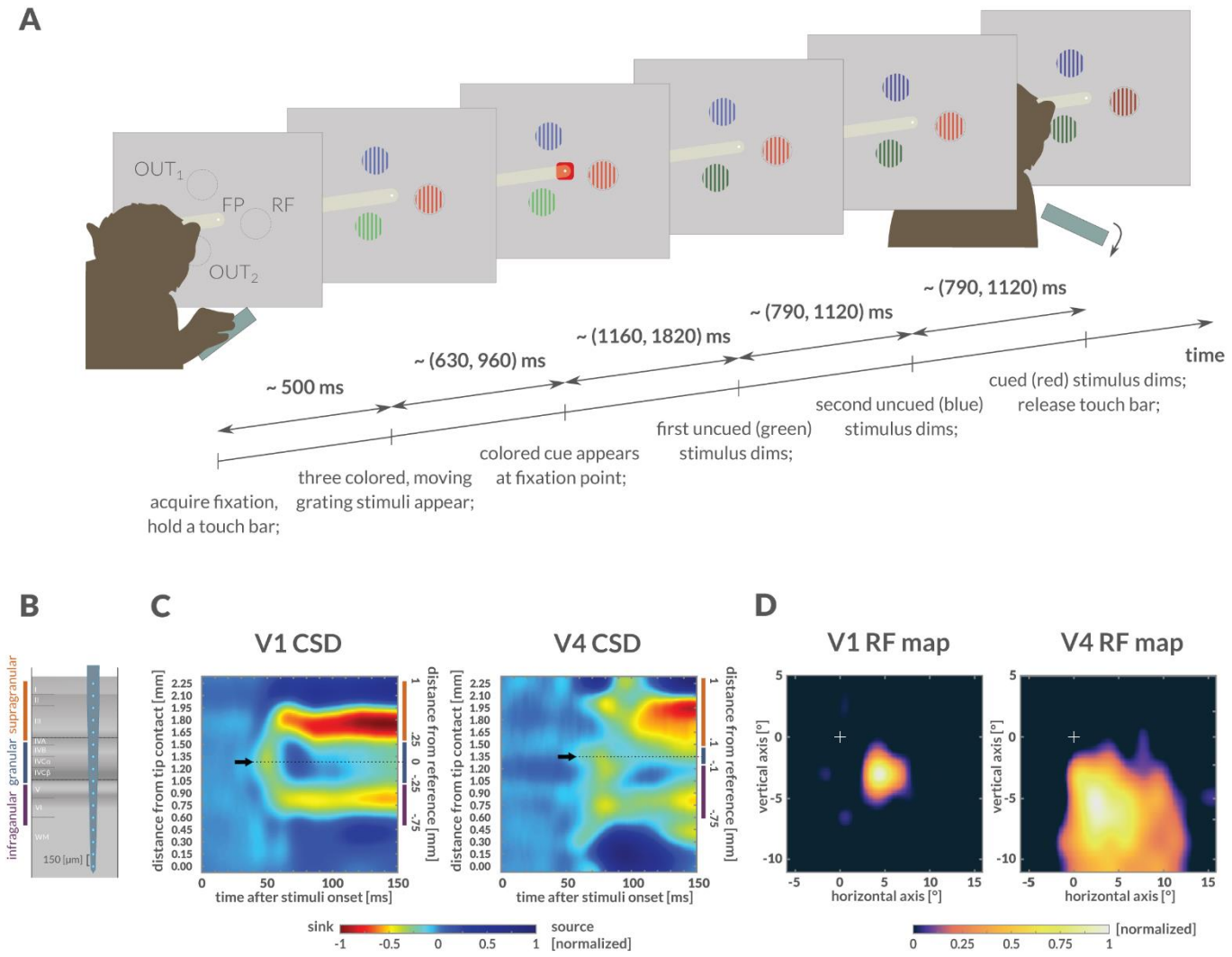


Figure 1: Behavioral task and recording setup. A) Covert, feature-guided visuospatial attention task. Monkeys fixated a fixation point (FP), and held a touch bar. Following fixation, three colored, moving grating stimuli were presented equidistant to the FP: one stimulus covered receptive field (RF) locations, the other two were located outside the RF (OUT_1 and OUT_2). With a random delay from stimuli presentation, a colored attention directing cue was presented at FP, indicating which stimulus was relevant on the current trial. Following the cue, the stimuli sequentially dimmed at unpredictable delays. When the relevant stimulus dimmed, the monkey had to release the touch bar to receive a fluid reward. Stimuli and cue colors, as well as the order of dimming of colored stimuli, were randomized across task trials. Ranges on the time line indicate the range of random event delays. **B)** Sketch to indicate laminar recording sites. Probes (16 contacts, 150 µm contact spacing) were injected normal to the cortical surface, aiming to cover all layers. **C)** Stimulus-induced CSD (example session), for both V1 (left) and V4 (right). Earliest current sinks were identified as layer IV (black arrows). Based on their distance from reference depth, recording contacts were assigned to granular, infragranular and supragranular compartments (shown on the right of CSDs). **D)** RF maps (mean across depths, same example session as C), for V1 sites (left) and V4 sites (right).

LFP data were analyzed in different time windows. Here we mostly present data from the time window preceding the first stimulus dimming (-503.25 to 0 ms, 512 time points). This corresponds to the period when attention was most focused on the relevant stimulus, and when attentional modulation of spiking activity is most profound in this task (Supplementary Figure S8, Thiele et al., 2016; Dasilva et al., 2019). We used bipolar re-referencing to improve spatial specificity of LFP signals (Methods for details).

In line with various reports, location of spectral power peaks differed between animals (Bosman et al., 2012; Bastos et al., 2015; Rohenkohl et al., 2018). Despite this, key analyses were performed within frequency ranges widely used in the literature (Buzsáki and Draguhn, 2004; Fries, 2005; Lakatos et al., 2008; Bosman et al., 2012; van Kerkoerle et al., 2014; Bastos et al., 2015; Richter et al., 2017; Rohenkohl et al., 2018; Spyropoulos et al., 2018). We focused on theta 4-8 Hz, alpha 8-13 Hz, beta 13-25 Hz, low-gamma 25-50 Hz, and high-gamma 50-80 Hz frequency. Adjusting frequency ranges to align with key features of individual monkey spectra (e.g. spectral peak locations) yielded qualitatively similar outcomes for all results described.

Spectral power and coherence across V1 and V4 layers

In V1, stimulus presentation increased spectral power relative to baseline (pre-stimulus) power, across cortical layers at beta band frequencies and above ($p<0.001$ for beta and gamma bands for monkey 1, n=224 pooled contacts; $p<0.001$ for all frequency bands for monkey 2, n=257; two-sided Wilcoxon signed rank test; Figure 2A shows data pooled across layers, supplementary Figures S1, S3 show layer resolved results). Attending to the RF increased low gamma frequency peak power in monkey 1 across all layers when compared to attend out conditions (Figure 2A, supplementary Figure S1). An increase in low gamma frequency peak power was not seen in monkey 2 (Figure 2A, supplementary Figure S1). However, in both monkeys attending to the RF stimulus resulted in 3-4 Hz higher low gamma power peak location compared to attend away conditions (changes were 32.82 ± 0.30 (S.E.M) Hz to 35.58 ± 0.26 (S.E.M) Hz in monkey 1, 46.83 ± 0.15 (S.E.M) Hz to 50.63 ± 0.15 (S.E.M) Hz in monkey 2; $p<0.001$ both monkeys; n=224 for monkey 1, n=257 for monkey 2; two-sided Wilcoxon signed rank test; Figure 2A). This phenomenon has been described as a shift towards higher frequencies with attention (Bosman *et al.*, 2012), but it is better described as a drop in frequencies when attention is directed away from the receptive field, as stimulus presentation results in a gamma peak slightly higher than that induced by attention (Figure 2A, dashed lines). Due to the differences in peak location, attention to the RF resulted in significantly higher spectral power at frequencies above the average of attend RF and attend out peak frequency location ($p<0.001$ for monkey 1, n=224; $p<0.001$ for monkey 2, n=257; two-sided Wilcoxon signed rank tests) and significantly lower power below the average frequency ($p<0.001$ in beta band for monkey 1, n=224; $p<0.001$ in low gamma band for monkey 2, n=257; two-sided Wilcoxon signed rank tests; Figures 2A). Additionally, decreases in V1 LFP spectral power with attention were found at lower frequencies ($p<0.001$ for theta and alpha bands in monkey 1, n=224; $p<0.05$ in alpha band

in monkey 2, n=257; two-sided Wilcoxon signed rank test; Figure 2A). These attentional effects were similar across cortical layers in V1 (supplementary Figures S1, S3).

Stimulus onset reduced low frequency spectral power in V4 in monkey 1 but increased it in monkey 2 (<13 Hz, p<0.001 in theta and alpha bands relative to pre-stimulus power; two-sided Wilcoxon signed rank tests; supplementary Figures S2 and S3). However, in both monkeys it increased spectral power for higher frequencies (>13 Hz, beta and gamma bands; p<0.01 in monkey 1 beta band, p<0.001 in monkey 1 low and high gamma band, n=306; p<0.001 in beta and gamma bands for monkey 2, n=225; two-sided Wilcoxon signed rank tests). Attention to the RF stimulus resulted in significant increases in LFP spectral power in intermediate and high frequencies (from beta to gamma band; p<0.001 in both monkeys, n=306 contacts in monkey 1, n=225 contacts in monkey 2; two-sided Wilcoxon signed rank tests), and significant decreases at low frequencies (p<0.001 in theta and alpha bands for monkey 1, n=306; p<0.001 theta band for monkey 2, n=225; two-sided Wilcoxon signed rank tests; Figures 2B, supplementary Figures S2, S3). In V4, effects of attention on spectral power were largely similar across cortical layers in both monkeys (supplementary Figures S2, S3).

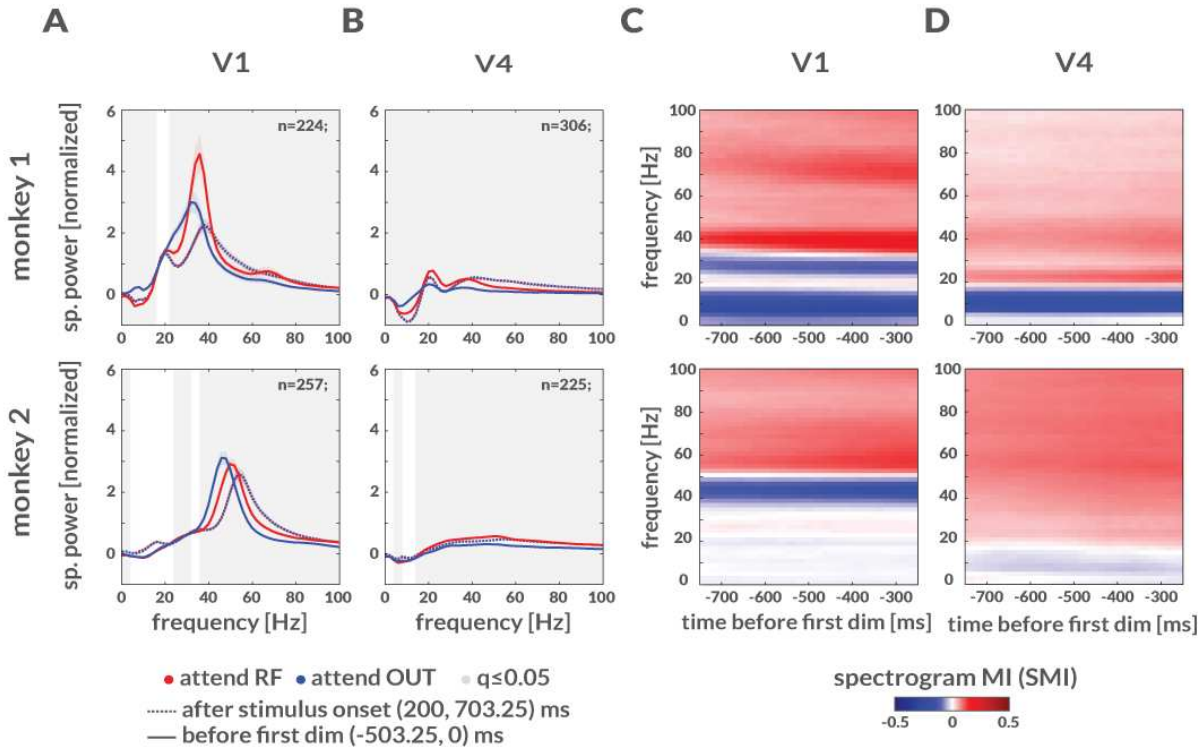


Figure 2: Attention decreases spectral power at lower frequencies and increases power at higher frequencies. A) Spectral power (mean ± S.E.M across sessions and depths) of bipolar LFP signals in ≈500 ms time windows. Dashed lines show spectral power after stimulus onset (200, 703.25) ms; solid lines show spectral power at times (-503.25, 0) ms before first dimming; shaded areas show S.E.M. Frequencies with significant difference between attentional conditions are shown by gray background (two-sided Wilcoxon signed-rank tests, FDR corrected q≤0.05). **B)** Same as in A, but for V4 LFPs. **C)** LFP attention spectral power modulation index (SMI, mean across sessions and depths) for LFPs from monkey 1 (top) and monkey 2 (bottom). Spectral analysis was applied to 503.25 ms time windows sliding in 20 ms steps, at times ≈(-1000, 0) ms before the first dimming. **D)** Same as in C, but for V4.

To assess attentional modulation of spectral power relative to the time of cue onset and to the time of the first dimming we calculated spectrogram modulation indices (SMIs) using a sliding window of 512 time points (503.25 ms length, Methods). Attentional modulation of spectral power (either positive or negative) increased after cue onset and persisted until the time of first dimming (supplementary Figures S1-S3). In V1, SMIs were positive for higher gamma frequencies, showed negative SMI for a narrow frequency just below the average gamma peak, followed by positive SMIs in the beta band and negative SMIs in low frequency ranges (alpha and theta band, Figure 2C). In V4, SMIs were negative for low frequency spectral power, i.e. attention reduced low frequency power in V4, while they were positive for frequencies >15-20 Hz, i.e. attention increased spectral power for mid and high frequencies (Figure 2D).

Attentional modulation of intra-area LFP spectral coherence largely followed the pattern described for spectral power (Figures 3A-B). This indicates that the local (bi-polar referenced) LFP power at specific frequencies is tightly coupled between layers. Attention to the RF resulted in significantly ≈1-2 Hz higher spectral coherence peak locations in the gamma-band in V1 (from 35.53 ± 0.13 (S.E.M) Hz to 36.50 ± 0.12 (S.E.M) Hz in monkey 1, from 47.53 ± 0.06 (S.E.M) Hz to 49.61 ± 0.06 (S.E.M) Hz in monkey 2; $p<0.001$ in both monkeys, n=1100 contact pairs for monkey 1, n=1512 for monkey 2; two-sided Wilcoxon signed rank tests; Figure 3A), it increased spectral coherence at higher frequencies ($p<0.001$ in low and high gamma in monkey 1, n=1100 contact pairs; $p<0.001$ high gamma in monkey 2, n=1512; two-sided Wilcoxon signed rank tests; Figure 3A) and decreased coherence at lower frequencies ($p<0.001$ in theta and alpha bands, $p<0.05$ in beta band for monkey 1, n=1100; $p<0.001$ in beta and low gamma bands in monkey 2, n=1512; two-sided Wilcoxon signed rank tests; Figure 3A). Slight increases were also found in lower bands ($p<0.05$ in lower beta band within ≈16-18 Hz in monkey 1, n=1100; $p<0.001$ in theta band for monkey 2, n=1512; Figure 3A). In V4 spectral coherence was increased by attention at higher frequencies (beta and gamma bands, $p<0.001$ in both monkeys; n=1949 contact pairs in monkey 1, n=1295 in monkey 2; two-sided Wilcoxon signed rank tests; Figure 3B), and decreased at lower frequencies ($p<0.001$ in theta and alpha bands in monkey 1, n=1949; $p<0.01$ in theta band in monkey 2, n=1295; two-sided Wilcoxon signed rank tests; Figure 3B). Inter-areal spectral coherence showed three main peaks (Figure 3C). One peak occurred at low frequencies (theta/alpha band), where attentional modulation differed between monkeys for the theta, but not for the alpha band (coherence was decreased in theta band for monkey 1, $p<0.001$, n=1940; increased in alpha band for monkey 1, $p<0.001$, n=1940; increased in theta band $p<0.05$, and alpha band $p<0.001$ for monkey 2, n=1802;

two-sided Wilcoxon signed rank tests). A second peak occurred in the beta band, with increased coherence for attend RF conditions ($p<0.001$ in both monkeys; n=1940 in monkey 1, n=1802 in monkey 2; two-sided Wilcoxon signed rank tests). A third peak occurred in the gamma band which increased for attend RF conditions ($p<0.001$ in low gamma for both monkeys, $p<0.001$, n=1940 in high gamma in monkey 1; $p<0.001$, n=1802 in monkey 2; two-sided Wilcoxon signed rank tests; Figure 3C). The effects of attention on spectral coherence were largely similar between layer pairs within areas, as well as across layer pairs between areas (supplementary Figure S4).

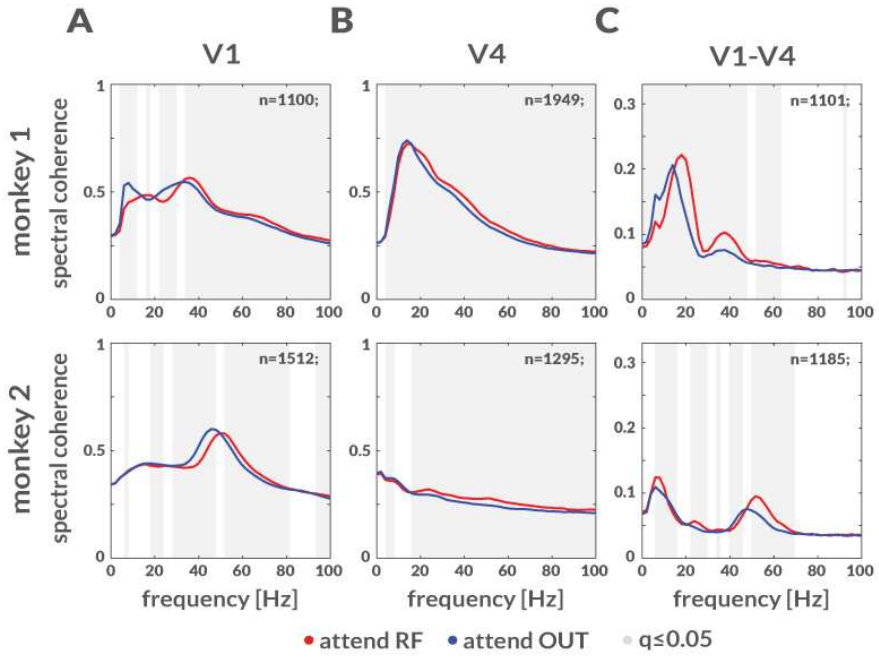


Figure 3: Effect of attention on LFP spectral coherence. A) LFP spectral coherence across V1 depths (mean ± S.E.M across sessions and depth pairs) at times (-503.25, 0) ms before the first dimming for monkey 1 (top) and monkey 2 (bottom). Gray background shows frequencies with significant difference between attentional conditions (two-sided Wilcoxon signed-rank tests, FDR corrected q≤0.05). **B)** Same as in A, but for V4. **C)** Same as in A, but for V1 -V4 coherence.

Causal communication between cortical layers and between cortical areas

To determine the flow of information within and between layers within and between areas we calculated conditional Granger Causality (cGC). cGC (Geweke, 1984) is a multivariate directed measure that allows to quantify the degree of causal relationship (or communication) between two nodes. For any directed contact pair (X,Y), cGC yields a conditional estimate of the causal flow from Y to X (and from X to Y), with the aim to discount the indirect influence of time-lagged interactions with contacts not covering the same laminar compartments as X and Y (Methods).

We first describe dominant interactions between layers and areas, irrespective of the effects of attention. This provides insight which frequency bands predominantly carry feedforward

and which frequency bands predominantly carry feedback information, independent of changing cognitive variables. Spectrally resolved intra-areal and inter-areal cGCs averaged across contact pairs are shown in Figure 4. All cGCs were significant (significance threshold is shown by dashed line in Figures 4A-D, computed as 95th percentile of cGCs with trials randomly shuffled; Methods).

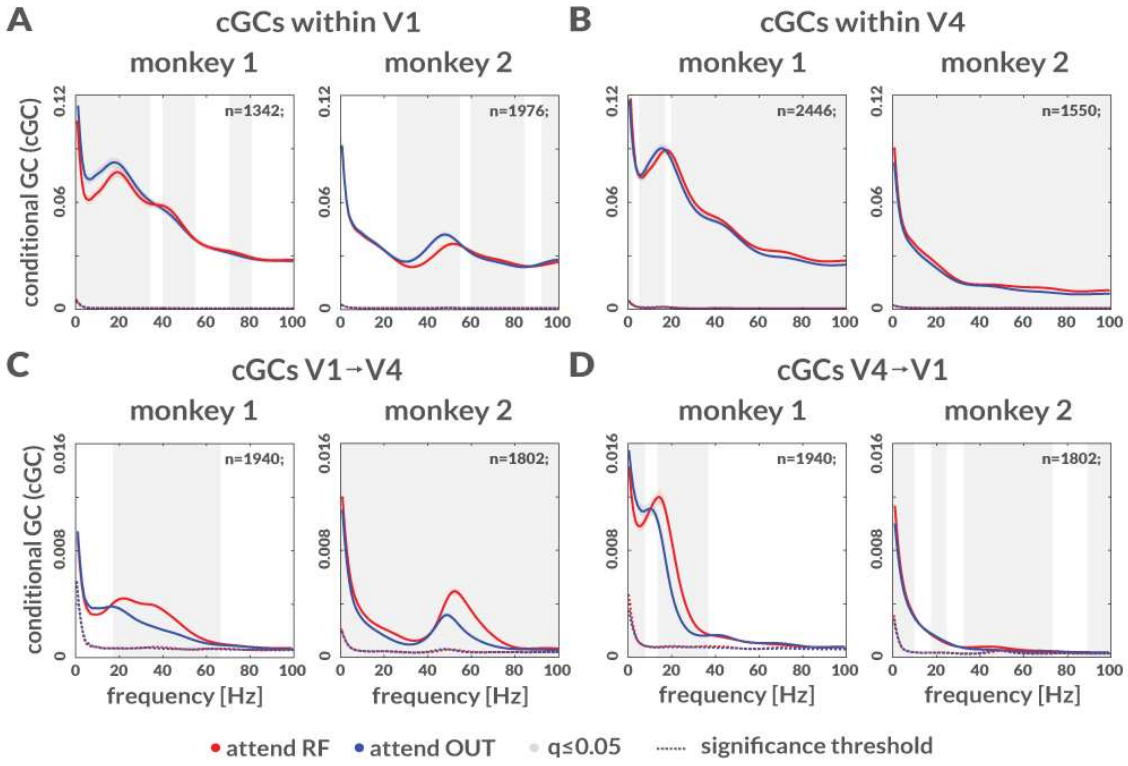


Figure 4: Conditional Granger Causality (cGC) for LFP signals. A) Conditional cGC pooled across depths in V1 (mean ± S.E.M across sessions and directed depth pairs) at times (-503.25 to 0) ms before first dimming for monkey 1 (left) and monkey 2 (right). Gray background shows frequencies with significant difference between attentional conditions (Wilcoxon signed-rank tests, FDR corrected q≤0.05,). **B)** Same as in A, but for V4. **C)** Same as in A, but for cGC from V1 to V4. **D)** Same as in A, but for cGC from V4 to V1.

To plot cGC results, we normalized each cGC to the maximum cGC across the 5 frequency bands (separately for within area and between areas cGCs after averaging across all sessions) for each monkey. To assess the dominant directionality of communication, for each contact pair (X,Y) we determined whether cGC was stronger from X to Y, or whether it was stronger from Y to X, and whether the directional difference was significant for a given frequency range ($q<0.05$, two-sided Wilcoxon signed rank tests, FDR corrected within frequency bands). We only present contact pairs where the directional cGC difference was significant. Significant differences are reported with color code indicating the dominant directions. For example, if a granular to supragranular cGC was larger than vice versa, it will be displayed in green in the cGC matrices, while the inverse direction will be displayed in magenta (Figure 5; for all contact cGCs differences, including non-significant ones see

supplementary Figure S6). Color intensity shows the relative strength of the interactions.

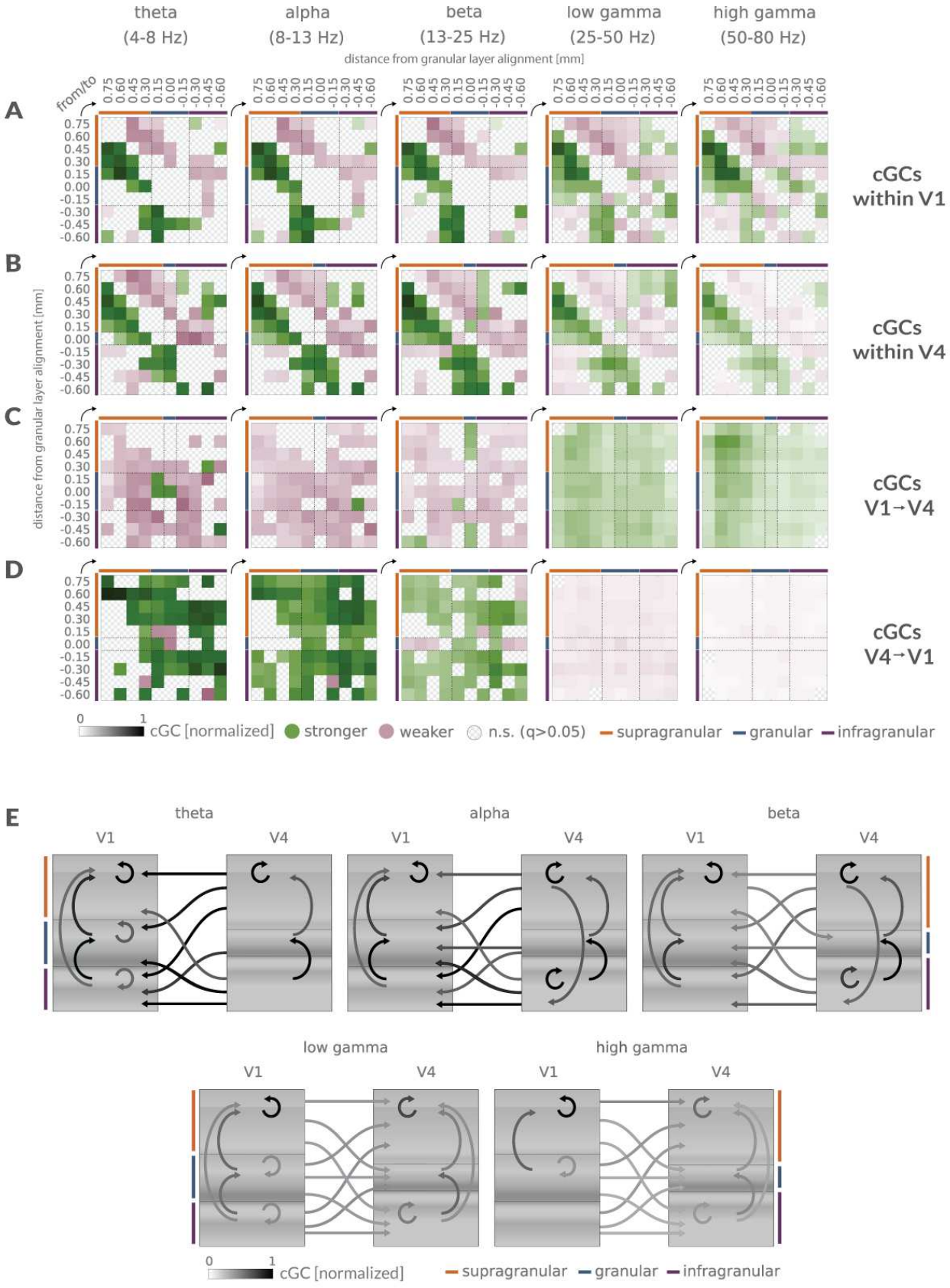


Figure 5: Directed connection matrices and influencer diagram of dominant cGC interactions before the first dimming. A) Normalized cGCs matrices (mean across sessions, pooled for the two monkeys) within V1 columns, for different frequency bands. Connection matrices are color coded to show significant dominant directions (green) and the opposite weaker directionality (magenta). Color intensity shows the relative cGC strength. Significance of cGCs difference

was assessed by two-sided Wilcoxon signed rank tests, FDR corrected ($q \leq 0.05$) within frequency bands. cGCs were first normalized for each monkey to the peak magnitude across frequencies, then pooled. **B)** Same as in A, but cGCs for V4. **C)** Same as in A, but for V1 to V4 pairs. **D)** Same as in A, but for V4 to V1 pairs. **E)** Influencer diagram of significant dominant cGC interactions, summarizing results in A-D. Arrows show dominant cGC interactions pooled for the two monkeys, averaged for the three laminar compartments (supragranular, granular, infragranular). Gray scale intensity of arrows indicates relative strength of cGCs (independently normalized for directions within V1, within V4, and between V1 and V4).

In V1, cGCs dominate in an upwards direction within supragranular layers for all frequencies (Figure 5A), they dominate in an upwards direction for all frequencies from granular to supragranular contacts, and they dominate in an upwards direction from infragranular to granular and supragranular contacts in the theta, alpha and beta frequency range, with smaller directional differences in the gamma frequency ranges. This pattern suggests that dominant interactions converge onto feedforward cortico-cortical output (supragranular) layers.

In V4 (Figure 5B), dominant interactions occurred in an upward direction within supragranular layers, across all frequency bands. Additionally, dominant cGCs were present in an upward direction from granular to supragranular layers, and from infragranular to granular layers. However, unlike in V1, cGCs dominated in a downward direction from supragranular to infragranular layers for most contacts and frequencies. Thus, within V4, a bidirectional dominance was found, whereby directly neighboring compartments communicated more strongly in an upward direction, while more distant compartments communicated more strongly in a downward direction.

Interactions between V1 and V4 were dominated in the feedback direction in lower (theta to beta) frequency bands (magenta color dominates for these frequency bands in Figures 5C-D) and in the feedforward directions in the gamma frequency ranges (green color dominates for these frequency bands in figure 5C-D). These V1-V4 interactions had little layer specificity with respect to origin or destination.

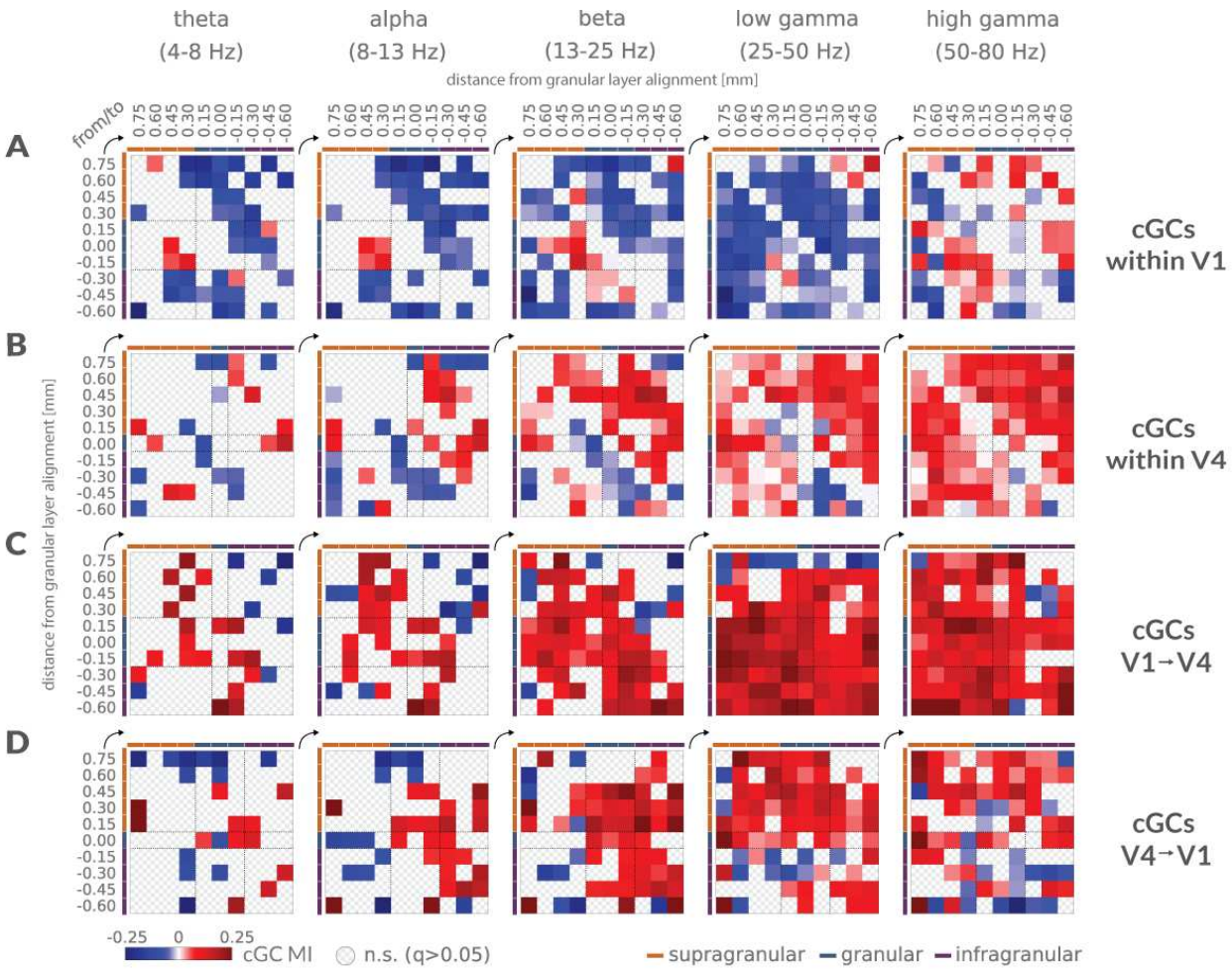
cGC interactions from V4 to V1 were strongest in theta to beta frequency bands (Figure 5D). In the theta and alpha band, they were most pronounced from V4 supragranular to all V1 layers. Strong interactions also occurred from V4 infragranular to V1 infragranular layer (Figure 5D). In comparison, V4 to V1 cGCs in the gamma frequency ranges were small (even though they were significant). Thus, the feedback cGC interactions predominantly occurred in lower frequency bands, they originated in V4 supra- and infragranular layers and affected V1 supra- and infragranular layers.

These intra- and inter-areal cGC interactions are summarized in an ‘influencer’ diagram (Figure 5E). It shows that in V1 dominant communication across almost all frequencies occurs in an upwards direction towards the supragranular cortico-cortical output layer. In

area V4 dominant communication occurs in a circular manner for lower frequencies (theta to beta), upwards within compartments and between neighboring compartments, but downwards from supragranular layers onto infragranular layer. In the gamma frequency range, dominant V4 communications were directed upwards towards the supragranular cortico-cortical output layer, mirroring the effects seen in V1. In the theta to beta frequency range, almost all interactions between V1 and V4 dominated in the feedback direction, while feedforward cGCs significantly dominated in the gamma frequency range.

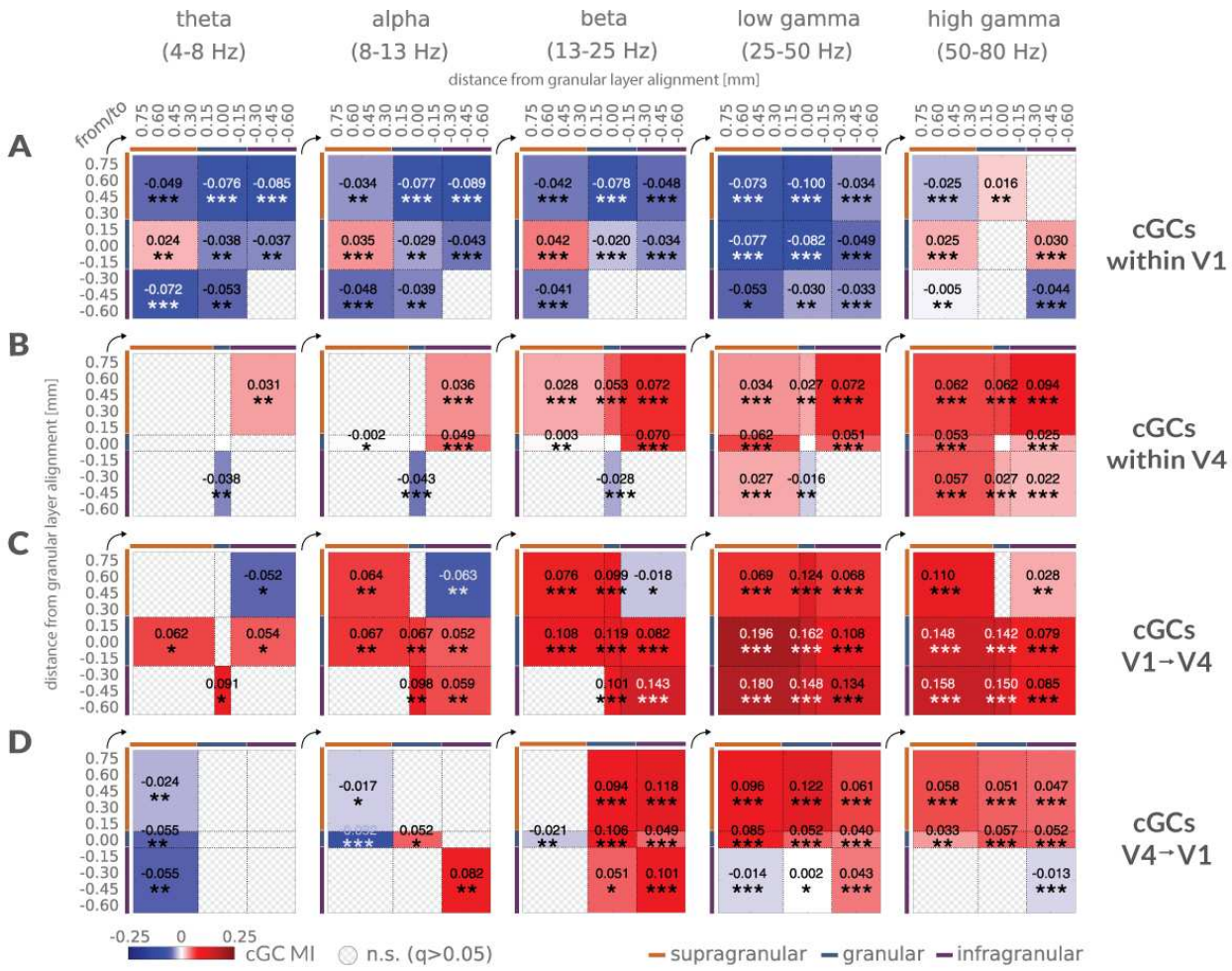
Attentional modulations of cGC interactions.

To assess attentional modulation of cGCs, we calculated modulation indices (MIs) for each recording and determined whether MIs of cGCs between layer compartments were significant ($q<0.05$, two-sided Wilcoxon signed rank tests, FDR corrected within frequency bands, Methods). Figure 6 shows significant intra- and inter-areal contact pairwise cGC attentional MIs pooled for the two monkeys. Figure 7 shows significant cGC attentional MI averaged by laminar compartments (supragranular, granular, infragranular; for all contact cGCs MIs differences, including non-significant ones see supplementary Figure S7). Surprisingly, within V1 cGC MIs were mostly negative, indicating that attention reduced cGCs in an upward and a downward direction across frequency bands (Figures 6A, 7A). except from granular to supragranular contacts. Additionally, high gamma band cGCs increased with attention from granular to supragranular layers, from supragranular to granular contacts, and from granular to infragranular contacts (Figures 6A, 7A). The predominant reduction in cGC with attention within V1 was surprising given that attention increased neuronal firing rates across layers within V1 (supplementary Figure S8).



315

Figure 6: Attentional modulation of cGCs. A) Significant attentional modulation index of cGC (cGC MI) among depth pairs within V1 (mean across sessions, pooled for the two monkeys), at different frequency bands (significance assessed via two-sided Wilcoxon signed rank tests, FDR corrected ($q \leq 0.05$) within frequency bands). **B)** Same as in A, but for V4. **C)** Same as in A, but for cGCs MIs from V1 to V4. **D)** Same as in A, but for cGCs MIs from V4 to V1.



316

Figure 7: Attentional modulation of compartment-wise cGCs. A) Significant attentional modulation index of cGC (cGC MI) among compartment pairs within V1, (mean across sessions, pooled for the two monkeys), at different frequency bands (significance assessed via two-sided Wilcoxon signed rank tests, FDR corrected within frequency bands; * $q \leq 0.05$, ** $q \leq 0.01$, *** $q \leq 0.001$). **B)** Same as in A, but for V4. **C)** Same as in A, but for cGCs MIs from V1 to V4. **D)** Same as in A, but for cGCs MIs from V4 to V1.

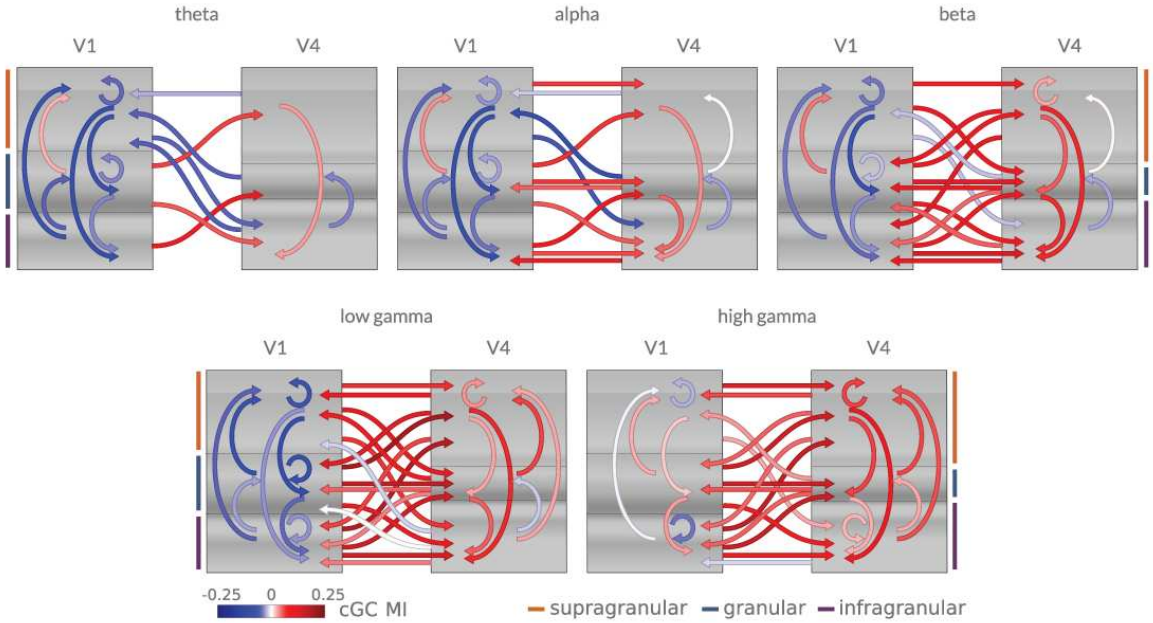


Figure 8: Main effects of attention on directed communication in different frequency bands. Arrows show significant attentional cGC modulation indices (cGC MI) (mean across sessions, pooled for the two monkeys), for the three laminar compartments (supragranular, granular, infragranular). Color indicates whether attention increases directed communication (red), or whether it decreases directed communication (blue), within and between areas.

Attentional modulation of cGCs in V4 was very different to the pattern seen in V1 (Figure 6B, 7B). Across all frequency ranges it increased from supragranular to infragranular layers but decreased from infragranular to granular layers for theta to low gamma frequencies. This could enable feedback information to flow prominently to lower areas (supragranular V4 to infragranular V4 and onwards to e.g. V2, V1), while at the same time limiting potentially inhibitory interactions (assuming infragranular layers communicate inhibitory prediction signals) on stimulus related processing (V4 infragranular to granular layers). In addition, downward communication (supragranular to infragranular) was increased by attention from theta to low gamma frequencies (Figures 6B, 7B). With an increasing oscillatory frequency, attentional modulation of cGCs between compartments increased, such that for high gamma frequencies, communication increases occurred across almost all compartments.

Despite the overall reduction of cGCs by attention within V1, its influence on V4 increased across frequency bands for most compartment comparisons (Figure 6C, 7C). In lower frequency bands, attention increased cGCs from V1 granular to all V4 layers (except for theta V1-V4 granular-granular interactions). However, in the theta to beta band V1 supragranular to V4 infragranular interactions were decreased. In the gamma frequency bands almost all V1 to V4 interactions were increased by attention.

cGCs from V4 to V1 were decreased by attention in the theta band from all V4 layers to V1 supragranular layers. In the alpha band significant decreases occurred from granular and supragranular V4 to supragranular V1, (Figure 6D, 7D). In the beta and low gamma band, attention increased V4 to V1 cGCs in a downwards direction (V4 supragranular to V1 granular and infragranular layer; from V4 granular to V1 granular and infragranular layer, and from V4 infragranular to V1 infragranular layer, Figures 6D, 7D). In the high gamma range attention increased cGCs from V4 supragranular and from V4 granular layers to all V1 layers, but decreased cGCs from V4 infragranular to V1 supragranular and infragranular layers.

These patterns of attentional modulations are summarized in a frequency dependent ‘influencer’ diagram in Figure 8. It shows the attention dependent reduction in cGCs across cortical layers and frequencies within V1, which nevertheless resulted in an increase in cGCs from area V1 to area V4. Feedback interactions were reduced by attention in the theta band, but mostly increased in the beta and gamma band. Within V4 cGCs were mostly increased in the beta and gamma bands. Some of these interactions are predicted by established theories of the frequency specificity interactions of feedforward and feedback connections, but many were in violation of established theory.

Discussion

Granger causal communication and its modulation by attention within and between areas V1 and V4 partly confirms, but also challenges current models of cortical processing. Within V1 dominant communication streams are directed towards supragranular cortico-cortical feedforward outputs. Conversely, in V4, dominant communication was bi-directional, with one stream of supragranular cortico-cortical feedforward flow, and a separate stream of supra- to infragranular feedback flow. Stimulus driven feedforward communication from V1 to V4 dominated in theta and gamma frequency ranges, with little layer specificity. Stimulus driven feedback communication from V4 to V1 dominated in the low frequency range. Surprisingly, attention to the receptive field generally reduced communication between cortical layers in area V1, with a notable exception for granular to supragranular communication. Within area V4, attention predominantly increased communication in beta and gamma frequency ranges. Despite the attention induced decrease of intra-areal V1 communication, attention increased feedforward communication from V1 to V4 across frequency bands. Attentional effects on feedback communication (V4 to V1) differed between frequency ranges. Theta and alpha communication decreased, while beta and gamma communication increased. Thus, feedforward interactions within and between cortical areas are neither limited to, nor dominant, in the gamma frequency range. Moreover, attention does not selectively increase gamma feedforward communication. Finally, feedback interactions between cortical areas, while dominant in the lower frequency range, are generally decreased by attention at low frequencies, but increased by attention in the gamma band.

For V1 LFPs, spectral power peak locations in the gamma range differed between attend RF and attend away conditions. The peak location for attend RF conditions resided at higher frequencies than for attend away conditions (~3-4 Hz). An equivalent result using ECoG surface recordings has been interpreted as a shift towards higher gamma peak frequencies induced by attention (Bosman *et al.,* 2012). However, our comparison with steady state post stimulus gamma peak locations shows that attention keeps the peak gamma frequency closer to the stimulus induced gamma frequency, i.e. attention stops it from dropping. This difference in interpretation is important, as it speaks to the role of attention, and potential mechanisms involved. Attention affects normalization circuits, causing a concomitant increase in excitatory and inhibitory drive of the attended object/location (Carandini and Heeger, 2013; Lee and Maunsell, 2009; Ray *et al.*, 2013; Reynolds and Heeger, 2009; Sanayei *et al.*, 2015). This increases the power and the frequency of gamma oscillations (Gieselmann and Thiele, 2008; Ray and Maunsell, 2010; Ray *et al.*, 2013). Attention thus ensures that stimulus representations remain sensory input driven, and sustained responses remain elevated (Pooresmaeili *et al.*, 2010; Reynolds *et al.*, 1999; Reynolds and Heeger, 2009; Treue and Maunsell, 1999; Williford and Maunsell, 2006). Within a predictive coding framework (PC) (Bastos *et al.*, 2012; Kanai *et al.*, 2015; Rao and Ballard, 1999), this could be interpreted in two ways. First, if attention reduced prediction generation for attended locations/features/objects, then prediction error coding populations would respond more strongly to sensory stimuli, as these are less predicted. This in turn increases feedforward communication, which has been associated with gamma frequency oscillations (Bosman *et al.*, 2012; van Kerkoerle *et al.*, 2014; Von Stein *et al.*, 2000). Second, according to an extension of predictive coding that allows attentional signatures to arise naturally within the model (Feldman and Friston, 2010; Spratling, 2008), attention increases the precision of predictions, making neurons respond more strongly to hidden causes (sensory input). Gamma oscillations, as a signal of prediction errors (Bastos *et al.*, 2012; Bastos *et al.*, 2015) in superficial layers would thus be increased. Which of the two interpretations is correct remains to be determined.

We did not find consistent increases in gamma power with attention in V1 (only consistent differences in peak location were found). However, V4 gamma power, and peak location were increased in both monkeys, in line with previous reports (Bosman *et al.*, 2012; Buffalo *et al.*, 2011; Fries *et al.*, 2001; Gregoriou *et al.*, 2009; Vinck *et al.*, 2013). Prominent gamma oscillations and its modulation by attention have been argued to be largely confined to supragranular layers (Buffalo *et al.*, 2011; Maier *et al.*, 2010; Xing *et al.*, 2012). We did not

find major differences in absolute gamma power, gamma power peak location, or attentional modulation of gamma power across supra-, granular, or infragranular layers in either V1 or V4. Using local bipolar referencing for all our analyses ensured that this was not an artefact of volume conduction. It shows that gamma frequency oscillations are not restricted to superficial layers, and they are thus unlikely a unique signature of feedforward interactions. Attention reduced oscillatory activity in theta and alpha bands in area V1 and V4, consistent with previous work (Bastos *et al.*, 2015; Bollimunta *et al.*, 2011; Brunet *et al.*, 2015; Buffalo *et al.*, 2011; Chalk *et al.*, 2010; Spyropoulos *et al.*, 2018; van Kerkoerle *et al.*, 2014). However, just as for gamma oscillations, these changes were not restricted to infragranular layers, but occurred across laminar compartments. These results equally question a strict separation between layer specific oscillatory frequency bands (Maier *et al.*, 2010; Spaak *et al.*, 2012; Xing *et al.*, 2012), and their potential association with feedforward and feedback signaling. They are more in line with recent reports about alpha sources across different modalities in primary sensory cortex (Haegens *et al.*, 2015).

Communication across layers within and between areas

Interareal cGCs support the proposal that gamma and theta frequency interactions dominate in the feedforward direction (V1 to V4. Bastos *et al.*, 2015; Bosman *et al.*, 2012; Spyropoulos *et al.*, 2018; van Kerkoerle *et al.*, 2014), while alpha and beta frequency interactions dominate in the feedback direction (V4 to V1. Bastos *et al.*, 2015; Bosman *et al.*, 2012; Spyropoulos *et al.*, 2018; van Kerkoerle *et al.*, 2014). However, cGCs within areas deviated from this scheme in important aspects. While local feedback interactions from infragranular to granular layers and to supragranular layers were most prominent at low frequencies (Spaak *et al.*, 2012; van Kerkoerle *et al.*, 2014), strong and dominant low frequency cGCs from granular to supragranular layers occurred. Moreover, dominant gamma cGC intra-areal feedback direction occurred (from infragranular to granular and to supragranular) were present, which have been argued to label feedforward circuits (van Kerkoerle *et al.*, 2014). Thus, all cGCs in V1 dominate in a direction that targets the cortico-cortical output (supragranular) layers. This was the case for all frequencies, irrespective of the assumed role of oscillations in different frequency bands (Babapoor-Farrokhran *et al.*, 2017; Bastos *et al.*, 2015; Bonnefond and Jensen, 2013; Buschman *et al.*, 2012; Fries, 2015; Gregoriou *et al.*, 2009; Spaak *et al.*, 2012; Womelsdorf *et al.*, 2010). It suggests that V1 plays a key role as a distributor of feedforward information, with relatively less responsibility of feedback processing (as a consequence, it may have little effect in the generation of predictions. Bastos *et al.*, 2012; Feldman and Friston, 2010). The pattern changes slightly in V4, but it

equally violates some key predictions about feedforward and feedback interactions. Namely, low frequency cGCs dominated in the feedforward direction (supra- to infragranular layers), while they dominated in the feedback direction in the gamma band (infra- to supragranular layers).

Attention to the RF reduced almost all cGCs within area V1, except for low frequency interactions from granular to supragranular layers. In the low gamma frequency band even those interactions were reduced, while most interactions were increased in the high gamma frequency band. The increase of cGCs from granular to supragranular layers is likely to boost feedforward output to other cortical areas, an expected effect given the increased efficacy demonstrated for feedforward spiking interactions and thalamocortical interactions with attention (Briggs *et al.*, 2013; Hembrook-Short *et al.*, 2019). If most intra-columnar feedback interactions served to compute context, while spatial attention boosts elementary processing (at the expense of contextual processing), then these cGC reductions are expected. Low frequency bands may predominantly play inhibitory roles (Bonnefond and Jensen, 2013; Haegens *et al.*, 2011; Spaak *et al.*, 2012). If these were reduced by attention, the increased firing rate seen in V1 in our and other studies (Hembrook-Short *et al.*, 2019; Herrero *et al.*, 2013; McAdams and Maunsell, 1999; Roelfsema *et al.*, 1998; Sanayei *et al.*, 2015; Wannig *et al.*, 2011) would be a natural consequence. Within the PC framework, it could be postulated that attention reduces the relative weight of predictions (although this is contrary to the proposal put forth by Feldman and Friston, 2010). Intuitively, attending to stimuli from the external world could mean re-shifting the balance from inferential to actuality processing, i.e. reducing the weight of internal priors. This would be achieved through reduction of feedback (local and interareal) and increase of feedforward processing. Such a re-shifting has been shown to be mediated by acetylcholine (Hasselmo and Bower, 1992; Roberts *et al.*, 2005), which plays an important role in attention (Dasilva *et al.*, 2019; Deco and Thiele, 2011; Herrero *et al.*, 2008; Roberts *et al.*, 2005).

The attentional modulation of cGCs in V4 differed radically from that in V1. Attention increased theta to beta band cGCs from supra- to infragranular layers and reduced theta to beta band cGCs from infra- to granular layers. In gamma bands almost all cGCs were increased. V4 is a major recipient of feedback from attentional signals originating in FEF (Gregoriou *et al.*, 2012; Gregoriou *et al.*, 2009; Moore and Armstrong, 2003; Moore *et al.*, 2003). The feedback is excitatory and predominantly targets excitatory cells in layer 2/3 (Anderson *et al.*, 2011). It could explain why cGCs originating from V4 supragranular layers show the most pronounced increases with attention. However, it does not explain why it occurs across all frequencies, if low frequency interactions label inhibitory interactions. Our

data suggest that this association with inhibitory roles is debatable for the case of FEF-V4 interactions, as we do not expect attention mediated feedback to increase inhibition, The strong increases of cGCs between all layer compartments across frequency bands in V4 suggest that feedback and feedforward intracolumnar interactions within V4 do not strongly differentiate between frequencies.

Interactions from V1 to V4 were mostly increased by attention across frequency bands. Attentional increase was most profound in the gamma band, in line with the notion that gamma oscillations mediate feedforward communication (Bastos *et al.*, 2015; Bosman *et al.*, 2012; van Kerkoerle *et al.*, 2014). However, low frequency interactions were also increased, which questions the generality of imputing feedforward communication exclusively to the gamma band.

A structure involved in coordinating large scale network interactions is the pulvinar, which regulates cortical synchrony in an attention dependent manner, particularly in the low frequency range (Saalmann *et al.*, 2012). However, pulvinar also affects oscillatory activity in the gamma frequency range in V4 (Zhou *et al.*, 2016). The changes seen for V1 to V4 cGCs in the low frequency range could be mediated through cortico-pulvinar-cortical interactions (Sherman *et al.*, 2002; Shipp, 2003). This might also explain the relative absence of layer specificity in cGC interactions between V1 and V4, irrespective of their direction.

Attention reduced communication from V4 to V1 in the theta band, and most strongly increased cGC in the beta band. However, strong increases also occurred in the gamma band, demonstrating that feedback interactions also operate strongly in the gamma band. V4 to V1 cGCs equally did not show strong laminar specificity. While this could be a consequence of subcortical routing (Sherman *et al.*, 2002; Shipp, 2003), it could also be a consequence of a termination pattern of V4 feedback that predominantly targets layer 1 dendritic spines through excitatory synapses (Anderson and Martin, 2006). These terminations can thereby influence pyramidal cells across supra- and infragranular layers. The predominance of excitatory connections on pyramidal cell dendrites is not consistent with the proposal that predictions generated at higher cortical levels act through di-synaptic inhibition for messages passing to lower areas (Bastos *et al.*, 2012).

A recent theory of 'predictive routing' (Bastos *et al.*, 2020) proposed that low frequency feedback prepares feedforward pathways, by inhibiting gamma and spiking activity associated with predicted inputs. A reduction in feedback (prediction) signals would thus cause disinhibition. Our results align, but also argue for an extension of this predictive

routing scheme. We argue for different hierarchies of prediction generation, some are automatic (e.g. surround suppression, basic contour integration, contrast normalization), while others are associated with higher cognitive functions (e.g. working memory, feature search, spatial attention, value estimation). We also propose that these to some extent employ different feedback networks. Automatic prediction generation mostly works within connections that affect non-classical receptive field interactions. This would explain why cooling of higher level areas results in reduced surround suppression (Hupé *et al.*, 1998), i.e. upon cooling, higher areas cannot pass predictions to lower areas. Inhibition is thus reduced and prediction error (or to use different words, sensory coding) signaling will be large. On the other hand, interactions between neurons sharing classical receptive field (cRF) locations counterbalance the prediction coding, i.e. they are predominantly excitatory. This explains why cooling of higher cortical areas results in reduced cRF responses (Hupé *et al.*, 1998). It is these cRF routes that might be exploited by higher cognitive functions, which through a separate form of feedback generate biased competition, and simultaneously serve to suppress automatic Bayesian inference (PC). Our data of attention induced increased feedforward, but decreased feedback communication within V1, increased feedforward and feedback cGCs within V4, and increased bidirectionally communication between V1 and V4 (with overlapping cRFs) across most frequency ranges support such a proposal.

Acknowledgments

Funded by Wellcome Trust 093104 (JvK, MB, AT), MRC MR/P013031/1 (JvK, AT), NIH Brain Initiative R01 NS108410 and U19 NS107464U19 (SP) and Simons Foundation SFARI Explorer grant 602849 (SP).

Author contributions

Demetrio Ferro: Data analysis and analysis methods, data curation, manuscript writing, visualization

Michael Boyd: data acquisition,

Jochem van Kempen: data acquisition, data analysis, manuscript review,

Stefano Panzeri: Data analysis and analysis methods, manuscript writing and review, supervision, funding acquisition,

Alexander Thiele: Conceptualization, data acquisition, resources, manuscript writing and

review, supervision, funding acquisition.

Competing interests

There are no competing interests.

Methods

EXPERIMENTAL PROCEDURES

Animals and procedures

We simultaneously recorded from visual areas V1 and V4 of two adult male rhesus macaque monkeys (Macaca mulatta, 10-11 years of age), while they performed a sustained top-down, feature-guided, visuospatial attention task. Experimental procedures were in line with the Directive 2010/63/EU of the European Parliament and of the Council of the European Union, the Guidelines for Care and Use of Animals for Experimental Procedures from the National Institute of Health, the Policies on the Use of Animals and Humans in Neuroscience Research from the Society for Neuroscience, and the UK Animals Scientific Procedures Act. Animals were motivated to engage in behavioural tasks through fluid control at levels that do not affect animal physiology and have minimal impact on psychological wellbeing (Gray *et al.*, 2016).

Surgical preparation

Animals were implanted with a head post and recording chambers over area V1 and V4 under sterile conditions and general anaesthesia. Surgical procedures and postoperative care conditions have been described in detail previously (Thiele *et al.*, 2006).

Behavioral paradigm

Monkeys were trained to comfortably sit in a primate chair while being head stabilized by the cranial head holder. Stimuli were presented on a cathode ray monitor (22" CRT, 120Hz, 1280x1024 pixel resolution) placed at 54 cm distance to the monkey's eyes. Eye position was calibrated and monitored by an eye tracking system operating at a sampling rate of 220Hz. Stimulus presentation and behavioral control was handled by Remote Cortex 5.95 (Laboratory of Neuropsychology, National Institute for Mental Health, Bethesda, MD).

Attention Behavioral Task

Monkeys had to touch a lever for the appearance of a centrally placed fixation spot. Thereafter they had to direct their gaze at a fixation point (FP) positioned at the center of the CRT screen, with a fixation window of ± 0.7°-1.5° of visual angle (DVA) throughout the trial duration.

500 ms after fixation onset monkeys were presented with three colored, moving grating

stimuli positioned equidistant from the FP. One stimulus was centered on the receptive field (RF) of recorded cells in V1, the other two were positioned outside (at locations OUT_1 and OUT_2). The RFs of recorded cells were mapped at the beginning of each experimental session (see below).

630 - 960 ms after stimulus onset (random delay, uniformly distributed, 1 ms steps) a colored cue was presented at FP. The color of the cue instructed the monkey to monitor the stimulus of matching color (e.g. a red cue instructed the animal to monitor the red visual stimulus) for a change in luminance contrast and ignore changes at the other stimulus locations.

After a random delay, the three stimuli started to sequentially dim in a pseudo-random order. Delays for subsequent dimmings ranged between 1160 – 1820 ms (the first dimming could occur 1160 - 1820 ms after cue onset, the second dimming could occur 790 - 1120 ms after the first dimming, etc.).

During the entire trial period monkeys had to keep fixating the FP. Upon cued stimulus dimming, monkeys had to release the touch bar within 600 ms to receive a fluid reward. Figure 1A graphically shows the time course of a sample trial of the main behavioral task. The grating stimuli had a diameter between 2 to 4 DVA, adjusted in accordance with the size and eccentricity of the recorded RFs. Their spatial frequency was 1.5 cycles/DVA, with a temporal frequency of 1 cycle/s (in sessions where they moved) and an orientation of 30°. The stimulus color at a given location was fixed (red, green or blue) for trials of the same session but randomized across sessions to cover all the 6 possible color configurations. In the same way, the cue color (red, green or blue), the order of dimming of the three stimuli (6 possible dimming orders), as well as the direction of movement of the grating stimuli (2 possible opposite directions, where applicable) were pseudo-randomized across trials to cover all possible task configurations.

Thus, there were 36 conditions total, which comprised a so-called cycle. In each cycle all 36 conditions would occur at least once, selected on a random basis. If the monkey performed the trial correctly, the condition was removed from the cycle pool. If the trial was not completed correctly, the condition was reinserted into the cycle pool, and would be reselected on a random basis, until all conditions had had been performed correctly.

Electrophysiological recordings

Electrophysiological recordings were performed using passive laminar probes with 16 recording contacts, inter-contact spacing of 150 µm (ATLAS Neuroengineering, Belgium). The laminar probes were inserted perpendicularly to the cortical surface with the support of a hydraulic micromanipulator (NARISHIGE MO-97A, Japan). All contacts were initially

referenced to a wire positioned either in the V1 chamber, or in the V4 chamber.
Data from the two chambers were simultaneously recorded using a digital acquisition and control system (Digital Lynx, Neuralynx, USA) with a sampling frequency of 32556 Hz (~32 kHz), at 24 bits.
The data were collected over 62 sessions (34 for monkey 1, 28 for monkey 2), yielding a total of 35744 correct trials (15892 in monkey 1, 19852 in monkey 2). These were out of 36912 total trials (16698 in monkey 1, 20214 in monkey 2), where monkeys kept fixation, yielding a behavioral performance of 95.17% correct for monkey 1, and 98.21% correct for monkey 2.

Receptive Field Mapping

Prior to starting the attention paradigm, the location and size of the RF was measured by a reverse correlation method (Gieselmann and Thiele, 2008).
From this, RF maps were initially estimated online to determine the stimulus locations in the attention paradigm. Offline RF analysis was done based on spike sorted single units, on thresholded multi-unit activity, as well as based on local population activity (envelope multi-unit activity, MUA_E, Supèr and Roelfsema, 2005) using a time window from 50-130 ms after RF mapping stimulus onset.

OFFLINE DATA ANALYSIS

Electrophysiological data analysis

Signals were extracted in time windows relative to task-related events: after stimulus onset (0 to 503.25 ms, 512 data points see below for LFP sampling frequency), after cue onset (0 to 503.25 ms, 512 data points see below for LFP sampling frequency) and before dimming times (503.25 ms before each of the three subsequent dimmings, 512 data points see below for LFP sampling frequency). Baseline activity time window started 200 ms before stimulus onset and covered up to 30 ms after stimuli onset.
Data were replayed offline, sampled with 16-bit, band-pass filtered at 0.5-300 Hz and down sampled by a factor of 32 to a sampling frequency F_s =1017.375 Hz to obtain local field potential (LFP) data. Spiking Activity was accessed by band-pass filtering between 600 and 9000 Hz, then further analyzed both at the level of multi-unit activity by extracting the Multi-Unit Activity Envelope (MUA_E), and by sorting single-unit spiking waveforms for RF mapping. Spikes were sorted manually using SpikeSort3D (Neuralynx).

Multi-Unit Activity Envelope and Signal to Noise Ratio

MUA_E was computed as described by (Supèr and Roelfsema, 2005). The higher frequency signal component (600-9000 Hz) was down sampled by a factor of 4, full-wave rectified, low-pass filtered at 500 Hz (Butterworth, zero-phase digital filter of order 5), and further down sampled by a factor of 8 to a frequency of 1017.375 Hz.

Signal to Noise Ratio (SNR) computation was performed on MUA_E signals in $n = 8$ sliding time windows of length 50 ms shifted every 10 ms from 30 to 150 ms after stimuli onset. The SNR is computed as the maximum average magnitude of baseline corrected MUA_E signal across time windows, i.e. $SNR = \max_n\{(\langle s_n(t)\rangle - \langle b(t)\rangle)/\sigma_b\}$, where $\langle s_n(t)\rangle$ is MUA_E average within time window n, and $\langle b(t)\rangle$ and σ_b are respectively the baseline mean and standard deviation.

Laminar alignment

Laminar signals from different experimental sessions were aligned to layer IV of both V1 and V4. Layer IV was identified for each session as the earliest current sink across laminae using current source density (CSD) of LFPs, and by analyzing the shortest latency of the stimulus evoked MUA_E response. Based on their distance from reference coordinate, signals from the corresponding recording channels were assigned to three main laminar compartments: supragranular, granular and infragranular. For V1, channels above the reference channel at distances of 0.25 - 1 mm were labelled as supragranular, channels above or below the reference channel within 0.25 mm were labelled as granular, and channels below reference at distance range 0.25 - 0.75 mm were labelled as infragranular. For V4, channels above the reference channel in the range 0.1 - 1 mm were labelled as supragranular, channels within 0.1 mm above or below the reference channel were labelled as granular, channels below the reference channel at 0.1 - 0.75 mm were labelled as infragranular.

Current Source Density analysis

The current source density (CSD) signal was obtained by applying the spline inverse CSD (iCSD) method (Pettersen *et al.*, 2006). Starting from the direct equation for the field potential $\mathbf{\Phi}$ generated by a point source $\mathbf{C}$ positioned at the origin of an isotropic medium $\mathbf{\Phi} = \mathbf{F} \cdot \mathbf{C}$, the iCSD was estimated by inversion of the conduction matrix $\mathbf{F}$ as $\hat{\mathbf{C}} = \mathbf{F}^{-1} \cdot \mathbf{\Phi}$. The coefficients of $\mathbf{F}$ were computed by electrostatic field equations for point sources by assuming that they are evenly distributed within isotropic cylindrical discs of finite radius R, and by assuming smooth CSD variation along depth dimension. CSD variation along depths

was approximated by cubic splines interpolation. In our computations we assumed a disc radius R =500 μm (Mountcastle, 1957), and we used conductance $\sigma = 0.4$ S/m (Logothetis et al., 2007). The conductance term could affect the magnitude of iCSDs but not their spatial profile. The iCSD was filtered by a Gaussian filter with standard deviation 200 μm along the depth dimension.

Response Latency analysis

The method used to compute stimulus response latency followed the formulation by Roelfsema *et al.* (2007), i.e. by assuming that stimulus responses arises at random times with a Gaussian distribution, and that a fraction of the response function dissipates exponentially after reaching a peak magnitude. On our data, the response function was estimated by fitting the 150 ms baseline-corrected post-stimulus MUA_E signals to a distribution $f(t)$ consisting in the sum of ex-Gaussian and cumulative Gaussian functions:

$$f(t) = d \cdot e^{\mu\alpha+(\sigma^2\alpha^2/2)-\alpha t} \cdot G(t, \mu + \sigma^2\alpha, \sigma) + c \cdot G(t, \mu, \sigma).$$

The parameters μ and σ respectively match the mean and standard deviation of the response function onset time when considering the response as non-dissipating. The parameter α is the dissipation rate, and the parameters c and d act as weighting factors for the response magnitude and dissipation terms. The functions $G(t, \mu', \sigma')$ are cumulative density functions of a generic normally distributed variable with mean μ' and standard deviation σ'.

Response latency is computed as the smallest time delay allowing to achieve 33% of the peak in the response magnitude $\hat{f}(t)$ estimated by least-square error minimization. In symbols, we computed latency as: $lat_{33} = \min\left\{t \in T: \hat{f}(t) = 0.33 \cdot \max_{u \in T} \hat{f}(u)\right\}$. To reduce computational cost of the least-square fit procedure, the empirical MUA_E response was smoothed by a moving average filter with length 5 samples, covering ≈5 ms.

Trials and Channels inclusion criteria

All our analyses only included trials with behaviorally successful outcome. To correct for eventual artifacts, which could be due to transient drifts of the probe, possibly caused by slight movements of the animal, we set a signal thresholding rule for trials selection. Trials were discarded if the baseline normalized signal energy was higher than the energy of a signal with magnitude 20 times bigger than baseline, i.e. if the signal $\bar{x}(t) = (x(t) - \langle b(t)\rangle)/\sigma_b$ had energy $\xi_{\bar{x}} > 20^2$, where $\xi_{\bar{x}} = \frac{1}{T}\int_0^T |\bar{x}(t)|^2 dt$, $x(t)$ is the LFP/MUA_E signal in any of the task-relevant time windows, $\langle b(t)\rangle$ and σ_b are the signal mean and

729 standard deviation at baseline.
730 Applying this thresholding rule led to the rejection of 2.1% of the trials, hence to the selection
731 of 34992 out of 35744 behaviorally correct trials (15468 were from monkey 1, 19524 for
732 monkey 2).
733 In all analyses we ensured to use equal amount of trials per attentional condition (RF, OUT_1,
734 OUT_2) by random sub-selecting trials in each session so that the amount of trials per
735 condition was equal to the minimum amount of trials available in the three conditions.
736 To prevent signal contamination due to common grounding or strong remote signal sources,
737 the signals for each electrode contact from the two cortical areas were locally referenced via
738 bipolar differentiation. The signal from depth z_i was replaced by the difference between
739 signals at depths z_{i+1} and z_{i-1}, as if it was recorded by a virtual electrode located at
740 intermediate depth between its two neighboring contacts. This procedure could not allow us
741 to consider the two most outer channels (as they could not be re-referenced with respect to
742 their neighbor channels), but this was often not problematic as the channels located at outer
743 positions were usually outside the grey matter of the targeted cortical areas. In addition, the
744 quality of signals recorded from any of the channels was determined by the computation of
745 SNR, and we only included channels with SNR ≥ 3. This resulted in data included from 481
746 channels for V1 (224 in monkey 1, 257 in monkey 2) and 531 channels for V4 (306 in monkey
747 1, 225 in monkey 2).
748

749 **Spectral Power**

750 The estimation of LFP signal power across frequencies was performed using a multi-
751 tapering approach (Thomas, 1982). We used the Chronux toolbox developed by Mitra and
752 Bokil (2008). We set the tapering to $K = 3$ Slepian waveforms with time-bandwidth product
753 $TW = 2$ ($T = N/F_s \approx 500$ ms, $W \approx 4$ Hz). The LFP spectral power $S_i(\lambda)$ was normalized for
754 each $\lambda \in [0, F_s/2]$ to baseline spectral power (minus trial-averaged baseline power, divided
755 by the standard deviation of baseline power).
756

757 **Spectral Coherence**

758 The relationship between the spectral components of LFP signals recorded from multiple
759 channels was quantified in terms of spectral coherence. This measure is computed by
760 means of the cross-spectrum power density $S_{ij}(\lambda) = X_i(\lambda) \cdot X_j^*(\lambda)$, involving the spectral
761 representations $X_i(\lambda)$ and $X_j(\lambda)$ of signals in channels i and j. The spectral coherence is

defined as $C_{ij}(\lambda) = |S_{ij}(\lambda)|^2 / |S_i(\lambda) \cdot S_j(\lambda)|, \lambda \in [0, F_s/2]$. The values assumed by $C_{ij}(\lambda)$ are in the range [0,1], where 0 means that the frequency components of the two signals are completely unrelated, and 1 means the two signals have perfectly linear relationship at given frequency component. The terms $S_i(\lambda)$, $S_j(\lambda)$, and $S_{ij}(\lambda)$ were computed with the use of the Chronux toolbox via multi-taper estimation (using $K = 3$ Slepian sequences, $TW = 2$).

Time-frequency spectral modulation

The spectral characterization was also performed in the time/frequency domain. LFP spectral power and coherence were both computed by using sliding time windows of duration 503.25 ms (N =512 time points), shifted in time every 20 ms to cover 1000 ms before the time of first stimulus dimming. The spectral resolution was $\Delta f = F_s/N \approx$ 2Hz and temporal resolution was Δt =20 ms.

Granger Causality Analysis

We measured directed causal communication between LFPs recorded at different contacts by using Granger causality (GC). We analyzed GC in a 503.25 ms time window (512 time points at 1017.375 Hz sampling rate) preceding the first dimming time. To reduce computational time, the signals were down sampled to 128 time points at a sampling frequency of 254.34 Hz.

In its original formulation (Granger, 1969; Geweke, 1982), the GC between two times series $Y(t)$ and $X(t)$ is computed by fitting a multivariate vector autoregressive model (MVAR) with finite memory p. The fit consists in estimating the linear interaction coefficients $\mathbf{A}_{k,k=1\ldots p}$ by least squares regression, yielding residual fit error of mean zero and covariance $\mathbf{\Sigma}$.
Spectral GC is characterized at each frequency $\lambda \in [0, F_s/2]$ via the cross-spectral density matrix $\mathbf{S}(\lambda)$ and the MVAR transfer function matrix $\mathbf{H}(\lambda)$ yielding factorization $\mathbf{S}(\lambda) = \mathbf{H}(\lambda) \cdot \mathbf{\Sigma} \cdot \mathbf{H}(\lambda)^*$ (Geweke, 1982). Spectral GC is then defined as:

$$f_{Y \to X}(\lambda) = \ln\left(\frac{|S_{xx}(\lambda)|}{\left|S_{xx}(\lambda) - H_{xy}(\lambda) \cdot \left(\Sigma_{yy} - \Sigma_{xy} \cdot \Sigma_{xx}^{-1} \cdot \Sigma_{xy}^*\right) \cdot H_{xy}^*(\lambda)\right|}\right), \quad \lambda \in [0, F_S/2].$$

To provide a more refined measure of the communication between time series of LFPs in channels Y and X, in our analysis we computed the GC between Y and X conditioned on $\mathbf{Z}$ (called here Conditional GC, with acronym cGC).
This more refined measure discounts the possible confounding effect of time-lagged

interactions mediated by activity of other recorded channels $\mathbf{Z} = \{Z_1, \ldots, Z_m\}$ rather than direct communication between the two considered nodes Y and X (see below for details of our infomax partial conditioning choice of the m channels $\mathbf{Z}$).

Following the derivation by Geweke (1984), cGC $f_{Y \to X|\mathbf{Z}}$ was computed by first applying a reduced least-square autoregression to the time series $X, \mathbf{Z}$ only, yielding residual error time series $X^{\dagger}$, $\mathbf{Z}^{\dagger}$. Then, cGC was defined via the identity: $f_{Y \to X|\mathbf{Z}}(\lambda) = f_{(Y \oplus \mathbf{Z}^{\dagger}) \to X^{\dagger}}(\lambda)$, $\lambda \in [0, F_s/2]$, allowing to express cGC with the original definition as unconditional GC between the variables $(Y \oplus \mathbf{Z}^{\dagger}) = \begin{pmatrix} Y \\ \mathbf{Z}^{\dagger} \end{pmatrix}$ and $X^{\dagger}$.

In our analysis, cGC was computed by the 'MVGC' method based on the computation of MVAR autocovariance sequences via Yule-Walker equations using the 'MVGC toolbox' by Barnett and Seth (2014). The magnitude of spectral GCs did not qualitatively vary when using alternative methods such as matrix partitioning (Chen et al., 2006), nonparametric spectral factorization (Dhamala *et al.*, 2008), or time reversed GC (Vinck et al., 2015).

The stationarity of LFPs, an important check for the application of Granger analyses, was assessed by ensuring that the MVAR characteristic polynomial $\varphi_{\mathbf{A}_k}(z) = \left|\mathbf{I} - \sum_k \mathbf{A}_k z^k\right|$ was invertible within unit disc, i.e. that $\max\{1/|z|, z \in \mathbb{C} : \varphi_{\mathbf{A}_k}(z) = 0\}$ was always < 1 (Barnett and Seth, 2014).

An important aspect of the analysis is the choice of which and how many (m parameter) the channels $\mathbf{Z}$ are chosen for conditioning in cGC. Conditioning on all available channels complementary to Y and X (m = ALL) ('full conditioning') might suffer from lack of sufficient data to estimate all autoregressive models needed for this calculation. In addition, full conditioning would make cGC unevenly scaled (since the number of available channels could vary across sessions), and regressing out channels $\mathbf{Z}$ within the same laminar compartments of Y or X would likely discount genuine interactions, because of the stronger correlations between channels within the same compartment due to volume conduction or other effects. Following Marinazzo et al. (2012), we thus applied the infomax partial conditioning strategy. We considered for conditioning only channels outside the laminar compartments of channels X and Y.

We then chose $\mathbf{Z}$ to be the m channels with the highest mutual information with Y and X (cGCs for different m are shown in Supplementary Figure S5). Mutual information (Shannon, 1948) between channel pairs was computed on the Hilbert envelope of LFP time series demodulated in 2 Hz frequency bins, then integrated in time and frequency. We used the 'Information toolbox' (Magri *et al.*, 2009) for mutual information estimation, and the method

by Panzeri and Treves (1996) for subtracting the limited sampling bias.

We chose the free parameters of the conditioning (the number and identity of conditioning channels) as the ones with the best Akaike Information Criterion (AIC) and the autoregression coefficient of determination R^2 (Supplementary Figures S5A-D). The two indices were adjusted to the size of the data sample to AICc (Hurvich and Tsai, 1989) and R_c^2 (Theil, 1961) in order to prevent from data overfitting, though the correction did not affect the results much. The optimization of AICc and R_c^2 led us to the selection of an MVAR model with memory p=10 (≈40 ms), and conditioning variable $\mathbf{Z}$ made by $m = 2$ channels outside the compartments of X and Y, achieving average R^2=0.8, s.e.m.= 0.0006 (p=10 and m=2 (out) in Figure S5).

To estimate the statistical significance of the empirical cGCs and to exclude any possible residual limited sampling biases, we recomputed cGCs after randomly shuffling the data across trials (we used 100 different shuffles for each directed channel pair). The significance of empirical cGC magnitudes was then assessed by setting a significance threshold equal to the 95th percentile of shuffled cGCs (Chen et al., 2006).

Attentional Modulation Index

To investigate the effects of attention, we compared results for the trials where attention was directed towards RF visual location against the ones where it was directed at outside locations OUT_1, OUT_2. Since the LFP spectral characterization for these two latter cases did not show prominent differences, we combined them in a single attend OUT condition by random subsampling an equal number of trials with condition OUT_1 and OUT_2. The modulation index (MI) for the measure F (spectral power or cGC) was defined as $F_{MI} = (F_{RF} - F_{OUT})/(F_{RF} + F_{OUT})$.

Statistical tests and significance

In all our analyses, the significance of the difference in spectral power, coherence, or cGCs (e.g. between time windows [before stimuli onset and after stimuli onset], attentional conditions [attend RF vs attend OUT], directionality of cGCs ($f_{X\to Y|Z}$ vs $f_{Y\to X|Z}$)), as well as the significance of attentional modulation indices (F_{MI}), were tested across experimental sessions by two-sided Wilcoxon signed rank tests (Wilcoxon, 1945). The p-values were corrected for False Discovery Rate (FDR) at q = 0.05 (Benjamini and Hochberg, 1995).

References

Anderson, J.C., Kennedy, H., 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. *The Journal of Neuroscience*. **31**(30), 10872–10881.

Anderson, J.C., Martin, K.A.C. (2006) Synaptic Connection from Cortical Area V4 to V2 in Macaque Monkey. *Journal of Comparative Neurology*. **721**, 709–721.

Babapoor-Farrokhran, S., Vinck, M., Womelsdorf, T., Everling, S. (2017) Theta and beta synchrony coordinate frontal eye fields and anterior cingulate cortex during sensorimotor mapping. *Nature Communications*. **8**(1).

Bagherzadeh, Y., Baldauf, D., Pantazis, D., Desimone, R. (2020) Alpha Synchrony and the Neurofeedback Control of Spatial Attention. *Neuron*. **105**(3), 577-587.e5.

Barnett, L., Seth, A.K. (2014) The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*. **223**, 50–68.

Bastos, A.A.M., Lundqvist, M., Waite, A.S., Kopell, N., Earl, K. (2020) Layer and rhythm specificity for predictive routing Affiliations : *bioRxiv*. (20200127921783).

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J. (2012) Canonical Microcircuits for Predictive Coding. *Neuron*. **76**(4), 695–711.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.M., Oostenveld, R., Dowdall, J.R., DeWeerd, P., Kennedy, H., Fries, P. (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*. **85**(2), 390–401.

Benjamini, Y., Hochberg, Y. (1995) Benjamini-1995.pdf. *Journal of the Royal Statistical Society B*. **57**(1), 289–300.

Bollimunta, A., Chen, Y., Schroeder, C.E., Ding, M. (2008) Neuronal Mechanisms of Cortical Alpha Oscillations in Awake-Behaving Macaques. *Journal of Neuroscience*. **28**(40), 9976–9988.

Bollimunta, A., Mo, J., Schroeder, C.E., Ding, M. (2011) Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations. *Journal of Neuroscience*. **31**(13), 4935–4943.

Bonnefond, M., Jensen, O. (2013) The role of gamma and alpha oscillations for blocking out distraction. *Communicative and Integrative Biology*. **6**(1), 20–22.

Bosman, C.A., Schoffelen, J.M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., Fries, P. (2012) Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas. *Neuron*. **75**(5), 875–888.

Briggs, F., Mangun, G.R., Usrey, W.M. (2013) Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. *Nature*. **499**(7459), 476–480.

Brunet, N., Bosman, C.A., Roberts, M., Oostenveld, R., Womelsdorf, T., De Weerd, P., Fries, P. (2015) Visual cortical gamma-band activity during free viewing of natural images. *Cerebral Cortex*. **25**(4), 918–926.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., Desimone, R. (2011) Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences*. **108**(27), 11262–11267.

Buffalo, E.A., Fries, P., Landman, R., Liang, H., Desimone, R. (2010) A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences*. **107**(1), 361–365.

Buschman, T.J., Denovellis, E.L., Diogo, C., Bullock, D., Miller, E.K. (2012) Synchronous Oscillatory Neural Ensembles for Rules in the Prefrontal Cortex. *Neuron*. **76**(4), 838–846.

Buschman, T.J., Miller, E.K. (2007) Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science*. **315**(Ci), 1860–1862.

Buzsáki, G., Draguhn, A. (2004) Neuronal Oscillations in Cortical Networks. *Science*. **304**(June), 1926–1929.

Callaway, E.M. (2004) Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Networks*. **17**(5–6), 625–632.

Callaway, E.M. (1998) Local Circuits in Primary Visual Cortex of the Macaque Monkey. *Annual Review of Neuroscience*. **21**(1), 47–74.

Carandini, M., Heeger, D.J. (2013) Normalization as a canonical neural computation. *Nature Reviews Neuroscience*. **13**(1), 51–62.

Chalk, M., Herrero, J.L., Gieselmann, M.A., Delicato, L.S., Gotthardt, S., Thiele, A. (2010) Attention

Reduces Stimulus-Driven Gamma Frequency Oscillations and Spike Field Coherence in V1. *Neuron*. **66**(1), 114–125.

Chen, Y., Bressler, S.L., Ding, M. (2006) Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *Journal of Neuroscience Methods*. **150**(2), 228–237.

Dasilva, M., Brandt, C., Gotthardt, S., Gieselmann, M.A., Distler, C., Thiele, A. (2019) Cell class-specific modulation of attentional signals by acetylcholine in macaque frontal eye field. *Proceedings of the National Academy of Sciences of the United States of America*. **116**(40), 20180–20189.

Deco, G., Thiele, A. (2011) Cholinergic control of cortical network interactions enables feedback-mediated attentional modulation. . **34**(January), 146–157.

Dhamala, M., Rangarajan, G., Ding, M. (2008) Analyzing information flow in brain networks with nonparametric Granger causality. *NeuroImage*. **41**(2), 354–362.

Douglas, R.J., Martin, K.A.C. (2004) Neuronal Circuits of the Neocortex. *Annual Review of Neuroscience*. **27**(1), 419–451.

Douglas, R.J., Martin, K.A.C., Whitteridge, D. (1989) A Canonical Microcircuit for Neocortex. *Neural Computation*. **1**(4), 480–488.

Feldman, H., Friston, K.J. (2010) Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*. **4**(December), 1–23.

Fries, P. (2005) A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*. **9**(10), 474–480.

Fries, P. (2015) Rhythms for Cognition: Communication through Coherence. *Neuron*. **88**(1), 220–235.

Fries, P., Reynolds, J.H., Rorie, A.E., Desimone, R. (2001) Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention. *Science*. **291**(5508), 1560–1563.

Geweke, J.F. (1982) Measurement of Linear Dependence and Feedback Between Multiple Time Series. *Journal of the American Statistical Association*. **77**(378), 304–313.

Geweke, J.F. (1984) Measures of Conditional Linear Dependence and Feedback between Time Series. *Journal of the American Statistical Association*. **79**(388), 907–915.

Gieselmann, M.A., Thiele, A. (2008) Comparison of spatial integration and surround suppression characteristics in spiking activity and the local field potential in macaque V1. *European Journal of Neuroscience*. **28**(3), 447–459.

Gieselmann, M.A., Thiele, A. (2016) Stimulus dependent laminar microcircuit interactions in primate V1. *Society for Neuroscience, abstracts 241.07*.

Granger, C.J.W. (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods Authors (s): C . W . J . Granger Published by : The Econometric Society Stable URL : http://www.jstor.org/stable/1912791 Accessed : 25-03-2016 19 : 26 UTC Your use of the JS. *Econometrica*. **37**(3), 424–438.

Gray, H., Bertrand, H., Mindus, C., Flecknell, P., Rowe, C., Thiele, A. (2016) Physiological, behavioral, and scientific impact of different fluid control protocols in the rhesus macaque (Macaca mulatta). *eNeuro*. **3**(4), 1–15.

Gregoriou, G.G., Gotts, S.J., Desimone, R. (2012) Cell-type-specific synchronization of neural activity in FEF with V4 during attention. *Neuron*. **73**(3), 581–594.

Gregoriou, G.G., Gotts, S.J., Zhou, H., Desimone, R. (2009) High-Frequency, Long-Range Coupling Between Prefrontal and Visual Cortex During Attention. *Science*. **324**(5931), 1207–1210.

Grothe, I., Neitzel, S.D., Mandon, S., Kreiter, A.K. (2012) Switching Neuronal Inputs by Differential Modulations of Gamma-Band Phase-Coherence. *Journal of Neuroscience*. **32**(46), 16172–16180.

Grothe, I., Rotermund, D., Neitzel, S.D., Mandon, S., Alexander, U., Kreiter, A.K., Pawelzik, K.R. (2018) Attention selectively gates afferent signal transmission to area V4. *Journal of Neuroscience*. **38**(14), 3441–3452.

Haegens, S., Barczak, A., Musacchia, G., Lipton, M.L., Mehta, A.D., Lakatos, P., Schroeder, C.E. (2015) Laminar profile and physiology of the α rhythm in primary visual, auditory, and somatosensory regions of neocortex. *Journal of Neuroscience*. **35**(42), 14341–14352.

Haegens, S., Nácher, V., Luna, R., Romo, R., Jensen, O. (2011) α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of

neuronal spiking. *Proceedings of the National Academy of Sciences of the United States of America*. **108**(48), 19377–19382.
Hasselmo, M.E., Bower, J.M. (1992) Cholinergic suppression specific to intrinsic not afferent fiber synapses in rat piriform (olfactory) cortex. *Journal of Neurophysiology*. **67**(5), 1222–1229.
Hembrook-Short, J.R., Mock, V.L., Martin Usrey, W., Briggs, F. (2019) Attention enhances the efficacy of communication in V1 local circuits. *Journal of Neuroscience*. **39**(6), 1066–1076.
Herrero, J.L., Gieselmann, M.A., Sanayei, M., Thiele, A. (2013) Attention-induced variance and noise correlation reduction in macaque v1 is mediated by NMDA receptors. *Neuron*. **78**(4), 729–739.
Herrero, J.L., Roberts, M.J., Delicato, L.S., Gieselmann, M.A., Dayan, P., Thiele, A. (2008) Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature*. **454**(7208), 1110–1114.
Hudson, A.E., Schiff, N.D., Victor, J.D., Purpura, K.P. (2009) Attentional modulation of adaptation in V4. *European Journal of Neuroscience*. **30**(1), 151–171.
Hupé, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., Bullier, J. (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*. **394**(August), 784–787.
Hurvich, C.M., Tsai, C.L. (1989) Regression and time series model selection in small samples. *Biometrika*. **76**(2), 297–307.
Kanai, R., Komura, Y., Shipp, S., Friston, K. (2015) Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **370**(1668).
van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., Roelfsema, P.R. (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences*. **111**(40), 14332–14341.
Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E. (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*. **320**(5872), 110–113.
Lee, J., Maunsell, J.H.R. (2009) A Normalization Model of Attentional Modulation of Single Unit Responses. *PLoS ONE*. **4**(2), 1–13.
Logothetis, N.K., Kayser, C., Oeltermann, A. (2007) In Vivo Measurement of Cortical Impedance Spectrum in Monkeys: Implications for Signal Propagation. *Neuron*. **55**(5), 809–823.
Magri, C., Whittingstall, K., Singh, V., Logothetis, N.K., Panzeri, S. (2009) A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. *BMC Neuroscience*. **10**.
Maier, A., Adams, G.K., Aura, C., Leopold, D.A. (2010) Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. *Frontiers in Systems Neuroscience*. **4**(August), 1–11.
Marinazzo, D., Pellicoro, M., Stramaglia, S. (2012) Causal Information Approach to Partial Conditioning in Multivariate Data Sets. *Computational and Mathematical Methods in Medicine*. **2012**, 1–8.
Markov, N.T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K., Kennedy, H. (2014) Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*. **522**(1), 225–259.
McAdams, C.J., Maunsell, J.H.R. (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*. **19**(1), 431–441.
Mitra, P.P., Bokil, H. (2008) *Observed Brain Dynamics*. Oxford University Press.
Moore, T., Armstrong, K.M. (2003) Selective gating of visual signals by microstimulation of frontal cortex. *Nature*. **421**(January), 370–373.
Moore, T., Armstrong, K.M., Fallah, M. (2003) Visuomotor Origins of Covert Spatial Attention. *Neuron*. **40**, 671–683.
Mountcastle, V.B. (1957) Modality and Topographic Properties of Single Neurons of Cat'S Somatic Sensory Cortex. *Journal of Neurophysiology*. **20**(4), 408–434.
Nandy, A.S., Nassi, J.J., Reynolds, J.H. (2017) Laminar Organization of Attentional Modulation in Macaque Visual Area V4. *Neuron*. **93**(1), 235–246.
Nicholson, C. (1973) Theoretical Analysis of Field Potentials in Anisotropic Ensembles of Neuronal

Elements. *IEEE Transactions on Biomedical Engineering*. **BME-20**(4), 278–288.
Nicholson, C., Freeman, J.A. (1975) Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. *Journal of neurophysiology*. **38**(2), 356–68.
Panzeri, S., Treves, A. (1996) Analytical estimates of limited sampling biases in different information measures. *Network: Computation in Neural Systems*. **7**(1), 87–107.
Peter, A., Uran, C., Klon-Lipok, J., Roese, R., Van Stijn, S., Barnes, W., Dowdall, J.R., Singer, W., Fries, P., Vinck, M. (2019) Surface color and predictability determine contextual modulation of V1 firing and gamma oscillations. *eLife*. **8**, 1–38.
Pettersen, K.H., Devor, A., Ulbert, I., Dale, A.M., Einevoll, G.T. (2006) Current-source density estimation based on inversion of electrostatic forward solution : Effects of finite extent of neuronal activity and conductivity discontinuities. . **154**, 116–133.
Pooresmaeili, A., Poort, J., Thiele, A., Roelfsema, P.R. (2010) Separable codes for attention and luminance contrast in the primary visual cortex. *Journal of Neuroscience*. **30**(38), 12701–12711.
Popov, T., Kastner, S., Jensen, O. (2017) FEF-controlled alpha delay activity precedes stimulus-induced gamma-band activity in visual cortex. *Journal of Neuroscience*. **37**(15), 4117–4127.
Rao, R.P.N., Ballard, D.H. (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*. **2**(1), 79–87.
Ray, S., Maunsell, J.H.R. (2010) Differences in Gamma Frequencies across Visual Cortex Restrict Their Possible Use in Computation. *Neuron*. **67**(5), 885–896.
Ray, S., Ni, A.M., Maunsell, J.H.R. (2013) Strength of Gamma Rhythm Depends on Normalization. *PLoS Biology*. **11**(2), 1–12.
Reynolds, J.H., Chelazzi, L., Desimone, R. (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*. **19**(5), 1736–1753.
Reynolds, J.H., Heeger, D.J. (2009) The Normalization Model of Attention. *Neuron*. **61**(2), 168–185.
Richter, C.G., Thompson, W.H., Bosman, C.A., Fries, P. (2017) Top-Down Beta Enhances Bottom-Up Gamma. *The Journal of Neuroscience*. **37**(28), 6698–6711.
Roberts, M.J., Zinke, W., Guo, K., Robertson, R., Mcdonald, J.S., Thiele, A. (2005) Acetylcholine Dynamically Controls Spatial Integration in Marmoset Primary Visual Cortex. *Journal of Neurophysiology*. **93**, 2062–2072.
Rockland, K.S., Pandya, D.N. (1979) Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*. **179**(1), 3–20.
Roelfsema, P.R., Lamme, V.A., Spekreijse, H. (1998) Object-based attention in the primary visual cortex of the macaque monkey. *Nature*. **395**(6700), 376–81.
Roelfsema, P.R., Tolboom, M., Khayat, P.S. (2007) Different Processing Phases for Features, Figures, and Selective Attention in the Primary Visual Cortex. *Neuron*. **56**(5), 785–792.
Rohenkohl, G., Bosman, C.A., Fries, P. (2018) Gamma Synchronization between V1 and V4 Improves Behavioral Performance. *Neuron*. **100**(4), 953-963.e3.
Saalmann, Y.B., Pinsk, M.A., Wang, L., Li, X., Kastner, S. (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. *Science*. **337**(6095), 753–756.
Sanayei, M., Herrero, J.L., Distler, C., Thiele, A. (2015) Attention and normalization circuits in macaque V1. *The European journal of neuroscience*. **41**(7), 949–964.
Self, M.W., van Kerkoerle, T., Supèr, H., Roelfsema, P.R. (2013) Distinct Roles of the Cortical Layers of Area V1 in Figure-Ground Segregation. *Current Biology*, 2121–2129.
Shannon, C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*. **27**, 379-423 623-656.
Sherman, S.M., Guillery, R.W., Feig, S.L., Van Lieshout, D.P. (2002) The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*. (November), 1695–1708.
Shipp, S. (2003) The functional logic of cortico-pulvinar connections. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **358**(1438), 1605–1624.
Smith, M.A., Jia, X., Zandvakili, A., Kohn, A. (2013) Laminar dependence of neuronal correlations in visual cortex. *Journal of Neurophysiology*. **109**(4), 940–947.
Spaak, E., Bonnefond, M., Maier, A., Leopold, D.A., Jensen, O. (2012) Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. *Current Biology*.

22(24), 2313–2318.
Spratling, M.W. (2008) Predictive coding as a model of biased competition in visual attention. *Vision Research*. **48**(12), 1391–1408.
Spyropoulos, G., Bosman, C.A., Fries, P. (2018) A theta rhythm in macaque visual cortex and its attentional modulation. *Proceedings of the National Academy of Sciences of the United States of America*. **115**(24), E5614–E5623.
Von Stein, A., Chiang, C., König, P. (2000) Top-down processing mediated by interareal synchronization. *Proceedings of the National Academy of Sciences of the United States of America*. **97**(26), 14748–14753.
Supèr, H., Roelfsema, P.R. (2005) Chronic multiunit recordings in behaving animals: Advantages and limitations. *Progress in Brain Research*. **147**(SPEC. ISS.), 263–282.
Taylor, K., Mandon, S., Freiwald, W.A., Kreiter, A.K. (2005) Coherent oscillatory activity in monkey area v4 predicts successful allocation of attention. *Cerebral Cortex*. **15**(9), 1424–1437.
Theil, H. (1961) *Economic Forecasts and Policy*. North Holland Publishing Co.
Thiele, A., Brandt, C., Dasilva, M., Gotthardt, S., Chicharro, D., Panzeri, S., Distler, C. (2016) Attention Induced Gain Stabilization in Broad and Narrow-Spiking Cells in the Frontal Eye-Field of Macaque Monkeys. *Journal of Neuroscience*. **36**(29), 7601–7612.
Thiele, A., Delicato, L.S., Roberts, M.J., Gieselmann, M.A. (2006) A novel electrode-pipette design for simultaneous recording of extracellular spikes and iontophoretic drug application in awake behaving monkeys. *Journal of Neuroscience Methods*. **158**(2), 207–211.
Thomas, D. (1982) Spectrum Estimation and Harmonic Analysis. *Proceedings of the IEEE*. **70**(9), 1055–1096.
Treue, S., Maunsell, J.H.R. (1999) Effects of Attention on the Processing of Motion in Macaque Middle Temporal and Medial Superior Temporal Visual Cortical Areas. *Journal of neurophysiology*. **19**(17), 7591–7602.
Vinck, M., Huurdeman, L., Bosman, C.A., Fries, P., Battaglia, F.P., Pennartz, C.M.A., Tiesinga, P.H. (2015) How to detect the Granger-causal flow direction in the presence of additive noise? *NeuroImage*. **108**, 301–318.
Vinck, M., Womelsdorf, T., Buffalo, E.A., Desimone, R., Fries, P. (2013) Attentional Modulation of Cell-Class-Specific Gamma-Band Synchronization in Awake Monkey Area V4. *Neuron*.
Wannig, A., Stanisor, L., Roelfsema, P.R. (2011) Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience*. **14**(10), 1243–1244.
Wilcoxon, F. (1945) Individual comparisons of grouped data by ranking methods. *Biometrics Bulletin*. **1**(6), 80–83.
Williford, T., Maunsell, J.H.R. (2006) Effects of Spatial Attention on Contrast Response Functions in Macaque Area V4. *J Neurophysiology*, 40–54.
Womelsdorf, T., Johnston, K., Vinck, M., Everling, S. (2010) Theta-activity in anterior cingulate cortex predicts task rules and their adjustments following errors. *Proceedings of the National Academy of Sciences of the United States of America*. **107**(11), 5248–5253.
Xing, D., Yeh, C.I., Burns, S., Shapley, R.M. (2012) Laminar analysis of visually evoked activity in the primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(34), 13871–13876.
Zhou, H., Schafer, R.J., Desimone, R. (2016) Pulvinar-Cortex Interactions in Vision and Attention. *Neuron*. **89**(1), 209–220.
Zumer, J.M., Scheeringa, R., Schoffelen, J.M., Norris, D.G., Jensen, O. (2014) Occipital Alpha Activity during Stimulus Processing Gates the Information Flow to Object-Selective Cortex. *PLoS Biology*. **12**(10).