Selective V1-to-V4 communication of attended stimuli mediated by attentional effects in V1

Christini Katsanevaki[1,2,*], André Moraes Bastos[1,3], Hayriye Cagnan[4,5], Conrado Arturo Bosman[6,7], Karl John Friston[4], Pascal Fries[1,2,6,8,*]

[1]Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, 60528 Frankfurt, Germany
[2]International Max Planck Research School for Neural Circuits, 60438 Frankfurt, Germany
[3]Department of Psychology and Vanderbilt Brain Institute, Vanderbilt University, TN 37240 Nashville, USA
[4]The Wellcome Trust Centre for Neuroimaging, University College London, London, WC1N 3BG, UK
[5]Medical Research Council Brain Network Dynamics Unit, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, OX1 3TH, UK
[6]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, 6525 EN Nijmegen, Netherlands
[7]Swammerdam Institute for Life Sciences, Center for Neuroscience, University of Amsterdam, 1098 XH Amsterdam, Netherlands
[8]Lead Contact
*Correspondence: christini.katsanevaki@esi-frankfurt.de and pascal.fries@esi-frankfurt.de

SUMMARY

Selective attention implements preferential routing of attended stimuli, likely through increasing the influence of the respective synaptic inputs on higher-area neurons. As the inputs of competing stimuli converge onto postsynaptic neurons, presynaptic circuits might offer the best target for attentional top-down influences. If those influences enabled presynaptic circuits to selectively entrain postsynaptic neurons, this might lead to selective routing. Indeed, when two visual stimuli induce two gamma rhythms in V1, only the gamma induced by the attended stimulus entrains gamma in V4. Here, we modeled this selective entrainment with a Dynamic Causal Model for Cross-Spectral Densities and found that it can be explained by attentional modulation of intrinsic V1 connections. Specifically, local inhibition was decreased in the granular input layer and increased in the supragranular output layer of the V1 circuit that processed the attended stimulus. Thus, presynaptic attentional influences and ensuing entrainment were sufficient to mediate selective routing.

INTRODUCTION

The visual system has evolved to represent the environment in a hierarchy of brain areas. Lower-area neurons have small receptive fields (RFs) and send converging projections to higher-area neurons. This convergence creates complex stimulus selectivity and large RFs. These large RFs usually cover, under natural viewing conditions, multiple stimuli, whereas those stimuli typically fall into RFs of separate neurons in primary visual cortex, V1. When one of those stimuli is attended, firing rates of V1 neurons are hardly affected, whereas firing rates of V4 neurons are similar to those observed when the attended stimulus is shown in isolation (Luck et al., 1997; Moran and Desimone, 1985; Reynolds et al., 1999). Thus, firing rates of neurons in area V4 and inferotemporal cortex primarily reflect the attended stimulus, whereas unattended stimuli appear to be filtered out.

This pattern of V1 and V4 firing rates can be modeled by assuming that attention increases the influence of ascending inputs to V4, specifically, the inputs from V1 neurons representing the attended stimulus - see “biased competition model” in (Desimone and Duncan, 1995; Reynolds et al., 1999) and “normalization model” in (Reynolds and Heeger, 2009). In these mathematical models, attention needs to selectively address the inputs representing the attended stimulus, and to implement this, separate variables are used for the inputs reporting distinct stimuli. However, those converging inputs correspond to synapses onto the dendritic fields of V4 neurons, namely, synapses that are specific to the locations and features of the visual stimuli. The complete set of all synaptic inputs of a given V4 neuron can be partitioned by the multiple stimuli in its RF in essentially infinite ways. Moreover, the features of the visual stimuli, and therefore the partitioning of synaptic inputs, might change dynamically as the stimuli change or the eyes move. Enhancing the gain of the attended stimulus at the synaptic level in V4 would therefore require top-down projections

that address exclusively and dynamically all (or at least a large number of) the synaptic inputs that correspond to the particular stimulus at any given moment. The number of those projections would need to be essentially infinite to cover all possible stimuli and attentional sets. Note that this problem arises for each of a very large number of V4 neurons activated by the attended stimulus; thereby, the challenge multiplies at the V4 population level.

By contrast, at the level of V1, distinct visual stimuli are represented by distinct neuronal populations. The population activated by a single stimulus, which encodes the location and features of the stimulus, forms an assembly that synchronizes in a gamma rhythm (Eckhorn et al., 1988; Gray et al., 1989; Lowet et al., 2017). Such assemblies have the potential to be addressed as a whole: top-down input coarsely matching the attended stimulus might well affect the whole neuronal assembly by means of its internal synchronization dynamics. Such top-down input could be implemented by a saliency map in a higher cortical area with projections onto V1 following a simple topographical mapping.

Therefore, a parsimonious account of attentional addressing relies on neuronal synchronization and is referred to as the Communication-Through-Coherence (CTC) hypothesis (Fries, 2005, 2015). CTC offers a flexible and biologically plausible mechanism—to enhance the impact of attended stimuli on higher areas—based on the coherence between higher and lower areas, e.g. macaque visual areas V1 and V4. Indeed, when two distinct visual stimuli induce two local gamma rhythms in V1, only the gamma induced by the attended stimulus entrains gamma in V4 and establishes coherence (Bosman et al., 2012; Grothe et al., 2012). Selective coherence between the attended V1 and V4 allows the attended V1 input to arrive on high-gain moments of the V4 gamma cycle (Fries, 2015; Ni et al., 2016; Rohenkohl et al., 2018). This likely results in attended inputs exerting a greater influence on V4 neurons, without the need to address those inputs at the synaptic or postsynaptic level. Instead, the top-down attentional signal should target the attended V1 population, affording it an advantage over the unattended V1 population—in their competition to entrain V4 in a coherent gamma rhythm. One piece of evidence supporting this hypothesis is the observation that the attended V1 gamma has a slightly higher frequency than the unattended gamma (Bosman et al., 2012; Ferro et al., 2021), potentially providing the attended V1 gamma a competitive edge in entraining V4 (Cannon et al., 2014; Fries, 2015).

Based on this CTC account, we hypothesized that both effects of attention, on V1 gamma frequency and V1-V4 gamma coherence, could be explained by attentional modulation acting solely within V1. To test this, we used a Dynamic Causal Model (DCM) of the canonical cortical microcircuit (Bastos et al., 2012) containing excitatory and inhibitory neuronal populations. This DCM models intrinsic (within an area) and extrinsic (between V1 and V4) connectivity in good correspondence with anatomical and functional observations in the macaque. Within the V1 microcircuitry, inhibitory populations have been proposed as targets for top-down signals from higher areas (Zhang et al., 2014) or as mediators of neuronal competition (Börgers and Kopell, 2008). Therefore, we used DCM to ask which inhibitory subpopulations are most likely to explain the observed effects. To evaluate the evidence for alternative models of attentional selection—by fitting alternative DCMs to the empirical data—we combined Variational Laplace (Friston et al., 2007) with a novel multi-start approach.

RESULTS

The experiment and recordings are described in detail in a previous publication (Bosman et al., 2012), and a brief summary is provided here. Two macaque monkeys were implanted with an electrocorticography (ECoG) grid that had multiple electrodes over areas V1 and V4. ECoG electrodes were used to acquire Local Field Potential (LFP) recordings. The LFP signal of two neighboring electrodes was subtracted to remove the recording reference, and each resulting bipolar derivation is referred to as a (recording) site (Bosman et al., 2012). V1 and V4 sites covered a wide range of RF locations. The monkeys were recorded while performing an attentional task (**Figure 1A**). Two stimuli were positioned to fall into one V4 RF, yet separate V1 RFs. That is, the two stimuli activated separate V1 sites, called "V1a" and "V1b", yet the same V4 site, called "V4". Therefore, the stimulus activating e.g. V1a will be referred to as "V1a stimulus"; a set of simultaneously recorded V1a, V1b and V4 sites will be referred to as a "V1a-V1b-V4 triplet" (or just "triplet"). In each trial, one of the two stimuli was behaviorally relevant and referred to as attended stimulus,

and the other stimulus was behaviorally irrelevant and referred to as unattended stimulus. Correspondingly, we will refer to the V1 site with the attended (resp., unattended) stimulus in its RF as the attended (resp., unattended) V1 site. Therefore, the trials were separated into two conditions: "attend V1a", where the V1a stimulus was the target, and the V1b stimulus was the distractor, and "attend V1b", where the reverse was true (see Methods section "Stimuli, task and data collection").

The contrast between these two conditions revealed the effects of selective attention on V1 and on the communication between V1 and V4 (**Figure 1B**) (Bosman et al., 2012). The key result is that the attended V1 site synchronized with the V4 site in the gamma band much more strongly than the unattended V1. Crucially, gamma power was not systematically different between the attended and the unattended V1 site, yet gamma peak frequency at the attended V1 site was higher than at the unattended V1 site, by ~3 Hertz.

Here, we investigate the mechanisms by which the attentional effects on gamma frequency and on interareal coherence could arise. As outlined in the introduction, we asked whether both effects can be explained by an influence of top-down attention solely on V1 microcircuits. Attentional top-down influences are conceived here as originating from an attentional control area, separate from V4; i.e., attentional top-down influences are considered distinct from V4-V1 feedback influences. V4-V1 feedback projections are included in the model, yet their influences are modulated indirectly, through changes in the intrinsic excitability of their targets. Therefore, we do not consider that attention modulates the V4-to-V1 feedback projections directly, and they are not the projections addressed here as the top-down influences carrying information about the attentional set. We identify the most likely V1 microcircuit changes that underwrite differences in attentional set between the two attentional conditions. Finally, we asked whether the empirical results can be explained in the absence of V4-V1 feedback projections.

To address these questions, we use a DCM of V1a, V1b and V4 activity to model the effects of attention. While the data we use here are observed on the LFP scale, DCM allows inferences at the level of specific V1 neuronal populations; in the sense of layer-specific excitatory or inhibitory neuronal populations and the connections between them.

Our model comprises three sources or nodes to account for the activity recorded from separate sites, corresponding to V1a, V1b and V4 (**Figure 1C**). The three nodes have the same internal structure, each consisting of four distinct neuronal populations: a supragranular and a deep layer pyramidal population ($\boldsymbol{s}$ and $\boldsymbol{d}$ respectively), an excitatory population in the granular layer ($\boldsymbol{g}$) and a population of inhibitory interneurons ($\boldsymbol{i}$) (**Figures 1C and 2**). The connections between the populations within each node are referred to as intrinsic connections. Extrinsic connections between V1 and V4 include excitatory V1-to-V4 feedforward projections and inhibitory V4-to-V1 feedback projections. This choice is compliant with the subtractive role of predictions conveyed by feedback projections in the framework of predictive coding; for a more extensive discussion see (Bastos et al., 2012). More specifically, each V1 microcircuit sends excitatory projections from the $\boldsymbol{s}$ population onto the V4 $\boldsymbol{g}$ and $\boldsymbol{d}$ populations, and the V4 microcircuit sends inhibitory projections from the $\boldsymbol{d}$ population onto the $\boldsymbol{s}$ and $\boldsymbol{i}$ populations of both V1 nodes. We will use the term "model architecture" or simply "architecture" to refer to a graph of intrinsic and extrinsic connections; as e.g., illustrated in **Figure 1C**; and we will use the term "model" to refer to a set of estimated model parameters or "posteriors" for a particular model architecture. We will first consider the model architecture shown in **Figure 1C** and then consider a modified architecture.

This DCM was fitted to the average V1a-V1b-V4 cross-spectral densities (CSD) of each of 20 triplets (10 from each monkey), using a multi-start approach. This approach involves restarting the fitting procedure multiple times, each time with the addition of random noise on the parameter estimates, to preclude the effects of local minima. We used this multi-start approach to obtain ~9000 models for each of the triplets (see Methods section "Multi-start Variational Laplace"). The models were compared on the basis of their free energy (a.k.a., evidence lower bound or ELBO), and the model with the greatest evidence (a.k.a. marginal likelihood) was considered the best or winning model for each triplet, and is reported in the following sections (Penny et al., 2004). Free energy quantifies how well the model fits the data (the probability of the data given the model structure and parameters), and incorporates a

complexity penalty for moving away from the priors (divergence between prior and posterior distributions; see **Table 1** for prior mean and variance values).

Intrinsic V1 modulation is sufficient to explain the effects of attention

Attention is modelled as a modulation of the strength of intrinsic V1a and V1b connections that is needed to switch between the "attend V1a" and "attend V1b" conditions to explain the LFP data features measured experimentally (**Figure 1C**). These data features include condition-specific changes in V1-V4 gamma coherence and V1 gamma power peak frequency, in the relative absence of differences in V1 gamma strength. All these features are reproduced on the CSD averaged over DCMs fitted to the 10 triplets of each monkey (**Figures 3B, S1;** see Methods section "Notes on the observation model"). As only intrinsic V1 connections are allowed to change between the two conditions, this demonstrates that intrinsic V1 modulation is sufficient to switch attention between the two targets.

DCM also allows to quantify the Directed Transfer Functions (DTF) that can be generated from the neural mass model that best explains the data. In the average over the 10 triplets of monkey P, the V1a-to-V4 DTF shows an increase in magnitude in the gamma band with attention, and the V1b-to-V4 DTF shows a slight increase in gamma peak frequency with attention, similar to the peak-frequency shift in V1 gamma power **(Figure 3C**). In the same analysis for monkey K, the V1a-to-V4 DTF did not show any appreciable attention effect (see Discussion section "Technical considerations"), and the V1b-to-V4 DTF showed a clear increase in magnitude in the gamma band with attention (**Figure S2**), consistent with similar changes in V1-to-V4 Granger causality (Bosman et al., 2012).

To identify the attentional modulations that are conserved across the two V1 sites and across all 20 triplets, we used Parametric Empirical Bayes (PEB) (Friston et al., 2016). This standard hierarchal Bayesian modelling approach assumes that attentional modulation is sampled from a normal distribution with some mean and variance, reported as the PEB posterior estimate for the attention modulation. The attentional modulation (β) of the intrinsic V1 connection strength (γ) was significantly positive for three out of the 12 possible (see explanation at end of paragraph) connections, namely for 1) the inhibitory self-connection of population $\boldsymbol{s}$ ($\beta_{s\to s} = 0.3043 \pm 0.0104$), 2) the inhibitory self-connection of population $\boldsymbol{d}$ ($\beta_{d\to d} = 0.1762 \pm 0.0107$), 3) the excitatory connection from population $\boldsymbol{s}$ to $\boldsymbol{i}$ ($\beta_{s\to i} = 0.2018 \pm 0.0077$). The attentional modulation was significantly negative for two connections, namely for 1) the excitatory connection from population $\boldsymbol{g}$ to $\boldsymbol{s}$ ($\beta_{g\to s} = -0.1418 \pm 0.0062$), and 2) the inhibitory self-connection of population $\boldsymbol{g}$ ($\beta_{g\to g} = -0.4422 \pm 0.0119$) (**Figure 3D**, also shown schematically in **Figure 3E**). Note that we refer to changes in 12 connections instead of 24 because the PEB models attentional effects in the two V1 sites in the same way: i.e., attentional modulation of the 12 V1a and 12 V1b connections are treated as equivalent (see Methods section "Quantifying attentional effects across triplets").

The role of inhibition

Bayesian Model Reduction (BMR) (Friston et al., 2016) and family-wise model comparison (Penny et al., 2010) used to test hypotheses concerning the role of inhibition in V1 microcircuitry in mediating attentional set. With BMR one can update the parameter estimates and evidence for a reduced model directly from a fitted full model in a way that precludes local minima. In this study, the full model is the one presented above, in the sense that all possible intrinsic connections within V1 were allowed to be modulated by attention. Here, a reduced model is any model where only some V1 connections are modulated by attention. Note that the attention-modulated parameters of the full model are confined within V1 and are therefore a subset of all possible parameters that attention could modulate, which would include e.g., extrinsic connectivity or intrinsic V4 connections. But since these are never considered in the present paper, the above-mentioned model is *full*, relative to the models considered in the following.

If we hypothesize that the attentional modulation, β, can be "turned on" or "off" for any number and combination of the 12 intrinsic V1 connection strengths, γ, the total number of possible reduced models with regard to these connections is 2^{12} = 4096, including the full model. In the following, the attentional modulation parameters of full models are referred to as β^F, and those of reduced models as β^R. For each model, its evidence was calculated using BMR. Then, each model was allocated to a family of models that is consistent with a hypothesis class to be tested. For example, if one would like to test whether the connection strength $\gamma_{s\to s}$ (see **Figure 2**) is modulated by attention,

the set of all reduced models would be partitioned into two non-overlapping subsets or families, corresponding to the two hypotheses: 1) one family with all models in which $\beta_{s\to s}$ is "on" (i.e. it assumes the same value in the reduced as in the full model: $\beta^R_{s\to s} = \beta^F_{s\to s}$), and 2) one family, in which $\beta_{s\to s}$ is "off" (i.e. $\beta^R_{s\to s} = 0$). Importantly, each family contains all models, in which the requisite condition is met, irrespective of all other parameters. Thus, the two families together contain all possible models in a given hypothesis. Subsequently, a posterior probability is calculated for the two families to determine which hypothesis is more likely to be true. This way evidence dilution — across all models that support a hypothesis (i.e., $\gamma_{s\to s}$ is modulated by attention) — is avoided. Note that we illustrate this for the case of one connection, which leads to two hypotheses; i.e., attentional modulation or not. For two connections, this will lead to 2^2 combinations, i.e., 4 hypotheses. Generally, for N connections, this will lead to 2^N hypotheses.

In order to assess the role of inhibition in mediating attentional set, our first test addressed the attentional modulation of the self-inhibitory connections of pyramidal populations in the DCM. We considered both self-inhibitory connections of the supragranular and of the deep pyramidal population and therefore split the reduced model into four families (**Figure 4A**): 1) family $f_1^{s,d}$, containing models where the self-inhibitory connections of both, the supragranular, $\boldsymbol{s}$, and the deep, $\boldsymbol{d}$, pyramidal populations are modulated with attention ($\beta^R_{s\to s} = \beta^F_{s\to s}$ and $\beta^R_{d\to d} = \beta^F_{d\to d}$); 2) family $f_2^{s,d}$, where *only* supragranular pyramidal population $\boldsymbol{s}$ self-inhibition is modulated with attention ($\beta^R_{s\to s} = \beta^F_{s\to s}$ and $\beta^R_{d\to d} = 0$); 3) family$f_3^{s,d}$, where *only* deep pyramidal population $\boldsymbol{d}$ self-inhibition is modulated with attention ($\beta^R_{s\to s} = 0$ and $\beta^R_{d\to d} = \beta^F_{d\to d}$); and 4) family $f_4^{s,d}$, where neither the self-inhibition of $\boldsymbol{s}$ nor $\boldsymbol{d}$ is modulated with attention ($\beta^R_{s\to s} = 0$ and $\beta^R_{d\to d} = 0$). Of these four families, the one where only supragranular pyramidal self-inhibition is modulated by attention had the greatest posterior probability ($p(f_2^{s,d}) = 0.5995$). The family with the next highest posterior probability allowed both $\boldsymbol{s}$ and $\boldsymbol{d}$ self-inhibitory connections to be modulated by attention ($p(f_1^{s,d}) = 0.2628$) (**Figure 4B**).

The second hypothesis tested whether the inhibitory population $\boldsymbol{i}$ in the V1 microcircuit plays a substantive role in attentional modulation; in the sense that attention would modulate 1) the inhibition it receives through its self-inhibitory connection ($\beta^R_{i\to i} = \beta^F_{i\to i}$), 2) the excitation it receives from other populations ($\beta^R_{g\to i} = \beta^F_{g\to i}$, $\beta^R_{s\to i} = \beta^F_{s\to i}$, $\beta^R_{d\to i} = \beta^F_{d\to i}$), and/or 3) the strength of its inhibitory projections to excitatory populations in the microcircuit ($\beta^R_{i\to g} = \beta^F_{i\to g}$, $\beta^R_{i\to s} = \beta^F_{i\to s}$, $\beta^R_{i\to d} = \beta^F_{i\to d}$). To answer this question, the reduced models were split into families that correspond to all combinations of the 3 alternatives, forming 2^3 = 8 families (shown schematically in **Figure 4D**). The family with the highest posterior probability was the family where none of the above connections are modulated by attention ($p = 0.8772$; **Figure 4E**). Note that factors 2) and 3) were defined as having each of three connections modulated by attention or none of them; if we relax this by defining them as having at least one of the respective three connections modulated, the results remained qualitatively the same.

The connectivity parameters of models that belong to a winning family can be averaged through Bayesian Model Averaging (BMA) (Litvak et al., 2015; Penny et al., 2010) to reveal the attentional modulation average over all models in the family. The BMA for the winning family of each of two comparisons reported above (**Figures 4C and 4F**), revealed that there were two common elements to attentional modulation: $\beta_{s\to s}$ was significantly increased with attention, and $\beta_{g\to g}$ was significantly decreased with attention. Following up on this result, we tested whether including the self-inhibitory connection of both the granular and the supragranular pyramidal population ($\beta^R_{s\to s} = \beta^F_{s\to s}$, $\beta^R_{g\to g} = \beta^F_{g\to g}$) would accumulate more evidence than either modulation alone (or no modulation at all) by splitting the reduced models into four families (**Figure 4G**). The family with the highest posterior probability was indeed the one where both modulations were present in the reduced models ($p = 0.8593$; **Figure 4H**). The BMA of the attentional modulation calculated across the models of this family was $\beta^R_{s\to s} = 0.3019 \pm 0.0104$ and $\beta^R_{g\to g} = -0.4428 \pm 0.0119$ (**Figures 4I and 4J**).

Attentional effects have a feed-forward character

Finally, we ask whether the presence of feedback projections from V4 to V1 is necessary for the intrinsic V1 modulation to explain the attentional effects. To this end, V4-to-V1 feedback was eliminated from the model, and

the model was fitted to the 20 V1a-V1b-V4 triplets. V1-to-V4 feedforward projections and all intrinsic connections were retained, and attentional modulation was only allowed for intrinsic V1 connections, as before (**Figure 5A**). Note that with this architecture, the attentional influence on V1 is still considered to be exerted by top-down projections onto V1. Remarkably, this modified DCM reproduced the broad features of the spectra and more specifically the attentional effects (**Figures 5B and S3**). Moreover, the DTFs obtained from the DCMs showed prominent effects in the gamma band: V1-to-V4 DTF gamma magnitude was increased selectively for the attended condition (**Figure 5C**).

We tested whether the attentional modulations found in the model architecture without feedback were similar to those obtained with the architecture with feedback. Indeed, among others, the key connections identified above, $\gamma_{s\multimap s}$ and $\gamma_{g\multimap g}$, show significant attentional modulations in the same direction as in the model with feedback ($\beta_{s\multimap s} = 0.2924 \pm 0.0121$ and $\beta_{g\multimap g} = -0.3330 \pm 0.0144$ respectively; **Figure 5D**).

We proceeded to calculate a family posterior probability for those two connections both being modulated with attention, as in the final family-wise comparison of the original architecture (see **Figure 5A**) and found that it had the highest probability among the competing hypotheses ($p = 0.5995$). Notably, with the no-feedback architecture, another intrinsic connection showed a significant attentional modulation effect in the BMA over the reduced models of the winning family: the excitatory connection from the granular population to the inhibitory population $\gamma_{g\to i}$. As a final check, we asked whether all three connection modulations are necessary in the reduced no-feedback model ($\beta^R_{s\multimap s} = \beta^F_{s\multimap s}$, $\beta^R_{g\multimap g} = \beta^F_{g\multimap g}$ and $\beta^R_{g\to i} = \beta^F_{g\to i}$) as opposed to any subset of the three connections or none. Among the eight families that were formed, the one with the highest posterior probability was indeed the one where all three connections were present ($p = 0.4745$). The BMA of the attentional modulation over the reduced models of this family was $\beta_{s\multimap s} = 0.2917 \pm 0.0115$, $\beta_{g\multimap g} = -0.3303 \pm 0.0144$ and $\beta_{g\to i} = -0.3478 \pm 0.0082$ (respective connection strength ratios shown in **Figure 5E**).

DISCUSSION

We used DCM to test whether attentional effects on inter-areal communication previously reported in the macaque visual cortex (Bosman et al., 2012) can be explained by modulating connectivity exclusively within the sender area. We used data recorded with an ECoG during an attentional task, where two visual stimuli activated separate V1 sites, V1a and V1b, but the same V4 site, thus competing for representation at the level of V4. The previous study reported that the V1 sites activated by the attended, as compared to the unattended, stimulus showed a higher gamma-peak frequency and stronger gamma coherence to V4 (Bosman et al., 2012). We used these data and fitted their spectral features with a microcircuit DCM of V1a-V1b-V4 activity. In our model, only the modulation of intrinsic V1 connection weights could account for the attentional effects. We show that intrinsic V1 modulation is sufficient to reproduce both attentional effects, the increases in V1 gamma frequency and V1-V4 gamma coherence. We found that inhibition plays a central role in attentional modulation, and we therefore tested specific inhibition-related hypotheses. The ensuing model with the highest evidence suggested that the strength of the self-inhibitory connections of both the supragranular pyramidal population and the excitatory granular population are modulated. The BMA estimate of the modulation showed that the self-inhibition is increased in the supragranular pyramidal population and decreased in the granular population. Finally, we show that intrinsic V1 modulation reproduced the attentional effects even in the absence of feedback connections from V4 to V1, suggesting that the selective entrainment of V4 can be achieved through a purely feed-forward mechanism.

Technical considerations

In this section, we offer a critical reflection on technical aspects of the model, on our assumptions and on the limits of the dataset used in the present study. First, as with any model, our DCM architecture and the accompanying priors are a simplification of the underlying cortical circuits. The description of inhibition in our model tries to capture the core inhibitory dynamics of cortical tissue. However, one could consider more expressive models, inspired by work on different interneuron populations, their different properties and distribution in the cortical column and their detailed connectivity. This would increase the complexity of the model and would require that the model be further

constrained by more informative data. Furthermore, the model could be enriched by information on the distribution of the signal power in different layers, as provided by laminar recordings.

Secondly, every model comes with a set of assumptions. For example, we assumed that the simplified circuit is sufficiently detailed to describe the dynamics that matter for selective attention. When using DCM to model attentional effects, DCM will find the simplest explanation in the sense of minimizing complexity (implicit in maximizing model evidence). This means DCM provides an accurate account of attentional effects that is as close as possible to prior assumptions. Although this ensures generalization and predictive validity, it does not guarantee the underlying attentional effects conform to prior assumptions.

The nature of the question demanded forming triplets of simultaneously recorded V1a, V1b and V4 sites, which meant that some sites participated in multiple triplets (see sub-panel labels in **Figures S1**). In the case of monkey K, one of the two V1 locations consisted only of two bipolar sites, one of which showed an attentional shift of gamma peak frequency opposite to (Bosman et al., 2012) and therefore was excluded (see exclusion criteria in Methods section "Triplet formation"), and the other showed a minimal effect in gamma band coherence. This second site was used in all triplets of monkey K (**Figures S1**). The small effect size of attention on coherence resulted in relatively high variance in the estimate of attentional effects β on connection strengths for this location (V1a, monkey K), meaning that their contribution to the overall estimate of β was very weak. Note that this limitation does not apply to our earlier study (Bosman et al., 2012), because that study uses triplets merely for illustration, yet bases the main analysis on V1-V4 pairs.

The DCM has been fitted to the entire spectrum of frequencies between 0 Hz and 100 Hz, rather than to the gamma range only. Given that the attentional effects in the gamma frequency range are the most prominent feature of the data—and that triplets were chosen to have the strongest attentional effect in gamma band coherence—we believe that the primary effect of the attentional effects β is to shift gamma peak frequency in V1 and to modulate gamma coherence between V1 and V4.

One study has suggested that LFP-LFP coherence between a sending and a receiving population can be due solely to synaptic inputs to the receiver, as those inputs contribute to the receiver LFP (Schneider et al., 2021). For our current dataset, this mechanism would predict V4 LFP to be essentially equal to the sum of V1 activities, which would in turn predict V1-V4 coherence to be dictated by the relative power between V1a and V1b. As attention did not significantly change V1 gamma power, this would predict an absence of changes in V1-V4 coherence, in contrast to the observed strong changes. Furthermore, V4 spiking is known to be entrained to the V4 gamma rhythm (Fries et al., 2001), which is entrained by the V1 gamma rhythm (Bosman et al., 2012), such that V4 spiking is coherent with V1 gamma (Grothe et al., 2012); this pattern of results strongly suggests that the V4 gamma is not merely reflecting synaptic input from V1, but genuine V4 entrainment.

Attention increases neuronal communication

Visual attention is of particular importance when two or more stimuli are simultaneously present (Desimone and Duncan, 1995; DeWeerd et al., 1999). These stimuli compete for representation in higher visual areas, and firing rates in those areas during attention can be modeled by assuming that attention selectively increases the relative influence of presynaptic inputs; namely, input-gain (Reynolds et al., 1999). However, how this input-gain modulation is physiologically implemented has remained elusive; particularly, because of the flexible nature of attention. Attention can be dynamically deployed to any visual stimulus, which should lead to the corresponding input-gain changes. Yet, the input space of a receiving neuron, corresponding to its receptive field, can be tiled by two or multiple stimuli in an essentially infinite number of ways. Thus, if attention would modulate specific subsets of synaptic inputs on the postsynaptic neuron by reaching them through hard-wired connections, an infinite number of such connections would be necessary. Rather, we argue that the efficacy of the attended subset of inputs from a lower area can be increased by enhancing their coherence with the postsynaptic neuron; this is the Communication-Through-Coherence hypothesis. (Fries, 2005, 2015). The postsynaptic neurons in higher visual areas, for whom the

inputs compete, can flexibly enter into coherence with either one of their competing inputs, and so CTC offers a mechanism by which selective communication is dynamically achieved through selective coherence.

Predictions of the CTC hypothesis have been tested in numerous previous studies. Visually induced gamma in awake macaque V4 rhythmically modulates the gain of spike responses and behavioral reaction times, suggesting gamma-rhythmic input gain modulation, a core prerequisite for CTC (Ni et al., 2016). A computational study modeled two connected neuronal populations and showed that spontaneous fluctuations in their gamma phase relation affect their transfer entropy (Buehlmann and Deco, 2010). This is supported by empirical studies showing that gamma phase relations in macaque and cat visual cortex affect effective connectivity, assessed as power correlation (Womelsdorf et al., 2007), and the gamma-phase relation between them affects their transfer entropy (Besserve et al., 2015). Several computational studies have demonstrated that this mechanism can be used in the case of two competing inputs, to selectively route forward the attended input. If a receiving population, in particular the respective inhibitory interneurons, is selectively entrained by one of two competing input gamma rhythms, the corresponding selected input is dominating the spike responses of the receiver (Börgers and Kopell, 2008). Similar selective routing effects have been found when several convergent pathways compete, and one of them is oscillatory, as long as the receiver is implemented at least as a simple spiking network with a single feed-forward interneuron layer (Akam and Kullmann, 2010), whereas filtering is much less effective when the receiver is less biologically realistic (Akam and Kullmann, 2012). In awake macaque visual cortex, as mentioned before, area V4 indeed shows selective coherence with the gamma rhythm of the attended V1 input (Bosman et al., 2012; Grothe et al., 2012). Furthermore, the attended V1 input entrains V4 on average at the gamma phase relation that is optimal for stimulus transmission, as assessed by behavioral reaction times to stimulus changes: any momentary deviation from this average gamma phase relation leads to lengthened reaction times (Rohenkohl et al., 2018). Such deviations occur frequently, due to the stochastic variability in gamma frequency (Spyropoulos et al., 2020). Intriguingly, when such stochastic variability is included in a computational model, the selective coherence can still robustly support selective communication (Palmigiano et al., 2017), probably by means of feedforward entrainment leading to frequency matching (Roberts et al., 2013); if in this scheme two inputs compete, the phase-leading input has the dominant influence on the receiver (Palmigiano et al., 2017).

Thus, numerous studies support the proposal that selective communication can be implemented by selective coherence. Here, we have investigated how the selective coherence can emerge for the attended input. Based on the considerations laid out above (in the Introduction and Discussion), we argue that selective V1-V4 coherence would emerge most parsimoniously from an attentional top-down modulation that acts exclusively at the attended V1 site. Our analysis revealed that this is indeed possible, and DCM endorsed this by quantitatively explaining empirically observed power and coherence spectra. The model furthermore allowed us to investigate the likely roles of distinct local neuronal sub-populations and suggested a decreased granular and increased supragranular self-inhibition. Note that within V1, the granular population drives the supragranular pyramidal population. Thereby, the attentional disinhibition in the granular population likely provides enhanced drive to the supragranular population. When the enhanced drive interacts with the increased inhibition in the supragranular population, this might produce the gamma frequency enhancement in the absence of gamma power changes. The enhanced gamma frequency can produce the selective coherence between the selected supragranular V1 output population and the target in V4 (Cannon et al., 2014).

Intriguingly, V1 gamma frequency shifts of comparable magnitude have previously been shown to be induced by increasing stimulus contrast (Lowet et al., 2015; Ray and Maunsell, 2010; Roberts et al., 2013), by stimulus onset, and by foveal position, all aspects enhancing stimulus salience (Fries, 2015) and by stimulus repetitions (Peter et al., 2021; Stauch et al., 2021). In a feedforward network with two competing inputs, gamma-frequency differences between the inputs might interact with theta-rhythmic gamma-phase resets to bring about selective entrainment by the faster gamma rhythm (Burwick and Bouras, 2017; Fries, 2015). Our current DCM analysis suggests that such resets are not necessary, yet this does not question their potential physiological relevance.

Targets and sources of the attentional signal

Our results provide formal evidence that attentional effects can be implemented by strengthening the self-inhibition of pyramidal neurons in the supragranular layers and weakening the self-inhibition of excitatory neurons in the granular layer. These self-inhibitory connections correspond functionally to the local, tightly connected recurrent network between pyramidal neurons and basket cells (or fast-spiking interneurons) (Buzsáki, 2006; Buzsáki and Wang, 2012). Therefore, strengthening or weakening of the self-inhibition can correspond to increasing or decreasing basket cell activity or the efficacy of the recurrent connections between them and pyramidal neurons. The recurrent interaction between excitatory and local fast-spiking inhibitory neurons is a key mechanism in the generation of gamma, as shown in theory (Buzsáki and Wang, 2012; Wang, 2010) and experiments (Cardin et al., 2009; Chen et al., 2017; Sohal et al., 2009). This gamma-generating circuit is under the inhibitory control of SOM+ (Somatostatin positive) interneurons, which in turn are inhibited by VIP+ (Vasoactive Intestinal Peptide positive) interneurons (Pfeffer et al., 2013). Thus, SOM+ interneurons might act like a break on the gamma-generation mechanism, and VIP+ interneurons might release that break. VIP+ interneurons are major targets of long-range top-down input (Lee et al., 2013; Ma et al., 2021; Zhang et al., 2014), which might correspond to attentional top-down input. Importantly, our experiments revealed an attentional increase in V1 gamma frequency and increase in V1-V4 gamma coherence, in the absence of significant changes in gamma power. Our dynamic causal modeling of those empirical data suggested differential effects on self-inhibition in different cortical layers. How this specific pattern of effects could be generated by the top-down modulation of different interneuron subclasses remains to be elucidated.

Furthermore, other studies have shown that a 30 Hz rhythm, sharing many characteristics with primate gamma, is reduced by SOM+ interneuron silencing and enhanced by VIP+ interneuron silencing (Veit et al., 2017; Veit et al., 2021); but see (Chen et al., 2017). As all those empirical data on the roles of defined interneuron subclasses were obtained in mice, the link between those rhythms and effects in mice to their putative homologues in non-human and human primates will be an important topic for future research.

Where does the attentional signal come from? Modulatory signals can have either a cortical or subcortical origin. The frontal-eye fields (Gregoriou et al., 2009; Thompson and Bichot, 2005; Wardak et al., 2006)) and the ventral pre-arcuate area (Bichot et al., 2015) in the prefrontal cortex—and the lateral intraparietal area in the posterior parietal cortex (Chen et al., 2016; Constantinidis and Steinmetz, 2005; Herrington et al., 2009; Ipata et al., 2009; Sapountzis et al., 2018)—are thought to exert attentional control, in either a spatial or a feature-based manner. They appear to construct and maintain salience or priority maps of the visual field, where locations with higher activity determine which part of the visual field is to be attended covertly or overtly. These regions interact with each other, but also project to—and modulate the activity of—V4 and early visual cortex (Anderson et al., 2011; Felleman and Van Essen, 1991; Gregoriou et al., 2009; Gregoriou et al., 2014; Richter et al., 2017).

Alternatively, attentional control may be exerted via sub-cortical nuclei: several sub-cortical nuclei provide neuromodulatory input to the cortex. There is evidence that cholinergic neuromodulation can mediate attentional effects on firing rates in the visual cortex (Herrero et al., 2008; Schmitz and Duncan, 2018; Thiele and Bellgrove, 2018); but also see (Veith et al., 2021). It can also modulate gamma band oscillatory activity (Howe et al., 2017; Kim et al., 2015), e.g., through direct projections from the basal forebrain onto PV+ (Parvalbumin positive) basket cells (Kim et al., 2015). Interestingly, administration of a cholinergic agonist in humans enhances performance in a spatial attention task without affecting gamma band power (Bauer et al., 2012), in agreement with the lack of consistent gamma power change in (Bosman et al., 2012).

Irrespective of the exact source of the attentional signal, there is a strong prerequisite, established by experiments that evince spatial attentional control in a specific subset of locations represented within a cortical area: The attentional signal needs to have high retinotopic specificity at the level of different populations within an area. The highest retinotopic resolution is available in area V1, and the present analysis shows that core attentional effects (Luck et al., 1997; Moran and Desimone, 1985; Reynolds et al., 1999) can be explained by modulations at this earliest stage of cortical visual processing and ensuing interareal synchronization phenomena.

FIGURE LEGENDS

Figure 1. Modelling the effects of selective attention on gamma-band neuronal synchronization in the visual cortex.

(A) Top: Selective attention task (Bosman et al., 2012). The monkey had to maintain fixation on the fixation spot. The stimuli consisted of two iso-luminant gratings, one in the V1a RF, the other in the V1b RF, and both contained in the V4 RF. RFs are illustrated with dashed circles, not visible to the monkey. The color of the fixation spot (cue) indicated which of the two gratings the monkey had to attend.

Bottom: Attentional cueing scheme, and resulting stimuli representation in higher area. Left panel: "Attend V1a" condition. The cue has the same color as the stimulus in the V1a RF, and primarily this attended stimulus is communicated to and represented in the higher area V4, as illustrated in the red-lined inset. The full attentional cueing scheme involves counterbalancing of cue color, stimulus color and stimulus orientation. The attentional filtering out of the unattended stimulus is illustrated by a gray masking of the V1b stimulus. Right panel: Same for the "Attend V1b" condition.

(B) Power and coherence of example V1 and V4 sites from monkey P, while the monkey was instructed to attend the stimulus in the V1a (red lines) or V1b (blue lines) RF. (Top) V4 power relative to baseline; (Bottom) V1 power relative to baseline; (Middle) V1-V4 coherence spectrum. Modified from (Bosman et al., 2012).

(C) The V1a-V1b-V4 microcircuit that was used to model the CSD measured from the data. Each area consisted of 4 subpopulations modelled as neural masses: an excitatory population in the granular layer $\boldsymbol{g}$, a pyramidal population in the supragranular layers $\boldsymbol{s}$, a pyramidal population in the deep layers $\boldsymbol{d}$ and a population of inhibitory interneurons $\boldsymbol{i}$. The pattern of intrinsic (within area) and extrinsic (between areas) connectivity is shown. Note that inhibitory (magenta) connections that appear to stem from excitatory (green) populations can be conceived as being mediated by local inhibitory interneurons. Extrinsic connections are depicted as bifurcating to simplify the illustration, but the strength of each of the 8 extrinsic connections was fitted independently. Dashed lines represent the connections whose strength could vary between the two attentional conditions.

Figure 2. The canonical microcircuit.

Equations of motion for the 4 neural masses within a canonical microcircuit node. v_x, depolarization of population $\boldsymbol{x}$; κ_x, inverse of time constant of population $\boldsymbol{x}$; $a^F_{x\to y}$, extrinsic feedforward excitatory connection strength from population $\boldsymbol{x}$ to population $\boldsymbol{y}$; $a^B_{x\to y}$, extrinsic feedback inhibitory connection strength from population $\boldsymbol{x}$ to population $\boldsymbol{y}$; $\gamma_{x\to y}$, intrinsic excitatory connection strength from population $\boldsymbol{x}$ to population $\boldsymbol{y}$; $\gamma_{x\multimap y}$, intrinsic inhibitory connection strength from population $\boldsymbol{x}$ to population $\boldsymbol{y}$; S_x, sigmoid function transforming the depolarization of population $\boldsymbol{x}$ to a firing rate; U, external input firing rate; $\dot{v}$, first temporal derivative of depolarization; $\ddot{v}$, second temporal derivative of depolarization.

Figure 3. DCM of the attentional effects.

(A) Schematic representation of the observation model used in the present study: The LFP of a node reflects the depolarization of the respective supragranular pyramidal population, weighted by the parameters of the observation model G.

(B) Power spectral density and coherence spectra averaged across the 10 V1a-V1b-V4 triplets of monkey P. Red (orange): attend V1a in the data (model), blue (green): attend V1b in the data (model). Top: V4 power spectrum, bottom: V1a and V1b power spectra, middle: V1a-V4 and V1b-V4 coherence spectrum.

(C) Absolute DTF averaged over the 10 V1a-V1b-V4 DCMs of monkey P.

(D) PEB posterior estimate of the attentional modulation β, combining data from both monkeys. This parameterizes the log ratio of connection strength in the "attend in" over the "attend out" condition for each intrinsic V1 connection. Error bars indicate 95% Bayesian confidence intervals (credible intervals), computed from the leading diagonal of the covariance matrix. Generally speaking, the effects whose credible intervals do not cross the zero line are "significant" in the sense that Bayesian model comparison provides "very strong" evidence for their presence (Kass and Raftery, 1995).

(E) Schematic representation of the "attend in" over the "attend out" connection strength ratio for connections significantly modulated with attention. The ratio is equivalent to twice the β value and is presented as the width of the pink- and blue-colored connections, relative to the width of grey-colored connections for the connections that are modulated by attention. Note that the strengths (γ) of the remaining connections, which are shown as grey lines, have also been fitted to best explain the data features but have the same value for the two attentional conditions (i.e. "attend in" over the "attend out" ratio is equivalent to 1).
See also supplementary Figures S1, S2.

Figure 4. The role of inhibition in attentional modulation.
(A-C) Hypotheses about the role of pyramidal population self-inhibition in the supragranular and deep layers.
(A) Schematic representation of the four hypotheses shown on an exemplar V1 microcircuit. Connections shown in continuous black lines are modulated by attention, connections shown in dashed black lines are not significantly modulated, whereas connections shown in grey may or may or not be modulated.
(B) Posterior probability of each family corresponding to a hypothesis shown in (A). The index of the family with the highest posterior probability is highlighted with a green number in (A).
(C) BMA of the attentional modulation β^R across all reduced models that belong to the winning family.
(D-F) Same as (A-C), but for hypotheses about the modulation of input or response gain of the inhibitory population.
(G-I) Same as (A-C), but for hypotheses about the role of self-inhibition in the supragranular pyramidal and in the granular population.
(J) Schematic representation of the results shown in (G-I), specifically the "attend in" over the "attend out" connection strength ratio for connections significantly modulated with attention. Conventions as in **Figure 3E**. Here, the significance of attentional modulation for each connection is considered relative to the BMA of β^R.

Figure 5. Attentional modulation in the no-feedback architecture.
(A) Schematic representation of the V1a-V1b-V4 microcircuit without V4-to-V1 feedback.
(B) Power spectral density and coherence spectra averaged across the 10 V1a-V1b-V4 triplets of monkey P.
(C) Absolute DTFs averaged across the 10 V1a-V1b-V4 DCMs of monkey P.
(D, E) PEB posterior estimate combining data from both monkeys, and schematic representation of the attentional modulation β. Conventions as in **Figures 3D and 3E**.
See also supplementary Figure S3.

Table 1. Prior mean and variance of DCM parameters.

Supp. Figure 1. Individual triplet DCMs of the attentional effects. Related to Figure 3.
Power spectral density of each node and coherence spectra between nodes for the 10 V1a-V1b-V4 triplets of monkey P (A) and monkey K (B). Each column represents one triplet.

Supp. Figure 2. DTF of monkey K. Related to Figure 3.
Absolute DTF averaged across the 10 V1a-V1b-V4 DCMs of monkey K.

Supp. Figure 3. DCM spectra and DTF of monkey K in the no-feedback architecture. Related to Figure 5.
(A, B) Results for DCM without V4-to-V1 feedback connections, for monkey K.
(A) Power spectral density and coherence spectra averaged across the 10 V1a-V1b-V4 triplets.
(B) Absolute DTF averaged across the 10 V1a-V1b-V4 DCMs.

STAR ★ METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Pascal Fries (pascal.fries@esi-frankfurt.de)

Materials Availability

The study did not generate new unique reagents.

Data and Code Availability

CSD data have been deposited at [repository name here upon publication] and are publicly available as of the date of publication. DOIs are listed in the key resources table.

SPM12 code is freely available. All original code and SPM12 functions that were modified for this study have been deposited at [repository name here upon publication] and are publicly available as of the date of publication. DOIs are listed in the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All procedures were approved by the animal ethics committee of Radboud University (Nijmegen, the Netherlands). Data from two adult male Rhesus monkeys (*Macaca mulatta*) were used in this study.

METHOD DETAILS

Stimuli, task and data collection

The data was collected using ECoG grids with 252 electrode contacts, implanted to cover a large portion of the left hemispheres of two adult male macaque monkeys. The monkeys were trained to perform a covert selective attention task (**Figure 1A**). After the monkeys touched a lever and attained fixation on the center of a screen, two square-wave gratings appeared at equal eccentricities from the fixation spot, one tinted blue, the other yellow, both drifting inside a static circular aperture. At a variable time during the trial, the fixation spot assumed the color of one of the gratings, indicating this grating to be the target, and the other to be the distracter. At a variable time thereafter, there was a change in the curvature of one of the two gratings, with equal probability for target and distracter changes. Changes in the target had to be reported by releasing a lever to obtain fluid rewards. For a detailed description of the task, the recordings, and the RF mapping, see (Bosman et al., 2012). The present study is based on the same dataset as (Bosman et al., 2012).

Preprocessing and MVAR model

The recorded data was low-passed filtered at 250 Hz to obtain the LFP. Bipolar derivations, referred to as (recording) sites, were calculated by subtracting the signal of two neighboring electrodes in the time domain to remove the common recording reference.

As in (Bosman et al., 2012), we segmented the data into non-overlapping epochs of 500 ms. The epochs begin at least 500 ms after the cue onset (i.e., the color change of the fixation spot) or 300 ms after a potential change in the distractor, and they end when the target changes. Epochs with a variance exceeding three times the variance across epochs were rejected; this applied to 2 epochs of monkey K (0.15% of all epochs of monkey K). The epochs were demeaned and normalized to a standard deviation of 1. The number of epochs was equalized across attention

conditions by randomly selecting epochs from the condition with the higher number of epochs, in order to avoid a sample-size dependent bias in the calculation of coherence (Cohen, 2014). This resulted in $N_{epoch} = 1253$ per condition for monkey P and $N_{epoch} = 590$ per condition for monkey K.

Before calculating the CSD, epochs were pre-whitened in order to counteract the effects of the $1/f$ characteristic of the power spectra. To do this, data were first down-sampled to 250 Hz. Then the ARfit toolbox was used to fit a first-order autoregressive (AR) model to the data and to calculate a common coefficient r for all channels and all epochs, which is assumed to estimate the $1/f$ component of the data. Each data sample z^w of the pre-whitened time series was calculated from the original data samples z^oas:

$$z_{t+1}^{w} = z_{t+1}^{o} - r z_{t}^{o}$$

where t denotes time steps.

A Multi-Variate Auto-Regressive (MVAR) model with model order 8 was fit to the pre-whitened data. Power, CSD and coherence were calculated on the estimated MVAR coefficients using the BSMART toolbox in Fieldtrip.

Triplet formation

The goal of the previous study (Bosman et al., 2012) was to investigate the case where each of the two stimuli, referred to here as stimulus A and stimulus B, is represented by a separate population in V1, referred to here as V1a and V1b, and both of them send converging inputs to the same population in V4. For this purpose, the authors had selected V1 sites that responded with a gamma-band peak selectively to one of the two stimuli, and V4 sites that responded approximately equally with a gamma-band peak to both stimuli. Since we intended to model the exact effects described by (Bosman et al., 2012) we used the same criteria for the selection of V1 and V4 sites. The main attentional effects observed by (Bosman et al., 2012) are: When attention was directed to the stimulus activating a given V1 site (as compared to when it was directed to the other stimulus), this V1 site showed 1) stronger coherence with V4 sites, and 2) higher gamma-peak frequency.

The goal of the present study was to investigate the generative processes underlying these attention effects by fitting a DCM. To maximize sensitivity, we used triplets of V1a, V1b, and V4 sites, which showed these attention effects particularly strongly. To do this, we first calculated the size of the shift in the V1 gamma-peak frequency and the size of the V1-V4 coherence increase with attention. For the size of the coherence effect, we summed coherence values in the gamma frequency range (monkey P: 50-80Hz, monkey K: 60-90Hz) and subtracted the total “attend out” from the total “attend in” coherence value. Then, we excluded all V1-V4 pairs whose V1 site showed a negative shift in the gamma-peak frequency with attention (monkey P: 0/67 or 0% of all pairs, monkey K: 1/20 or 5% of all pairs) or that had a decrease in V1-V4 coherence with attention (monkey P: 8/67 or 11.9% of all pairs, monkey K: 3/20 or 15% of all pairs). Then we formed all possible V1a-V1b-V4 triplets from the remaining V1-V4 pairs, by considering all combinations of V1a-V4 pairs and V1b-V4 pairs that had a common V4 site (monkey P: 165 triplets, monkey K: 16 triplets). Each triplet had two attentional coherence effect sizes, one for V1a-V4 and one for V1b-V4 coherence. Those effect sizes can be visualized as the coordinates of a point for this triplet on a scatter plot. For each triplet, we calculated the Euclidian distance between this point and the one defined by the maximum observed coherence effect magnitude across all V1a-V4 and V1b-V4 site pairs of a given monkey. Finally, we selected the 10 triplets with the lowest distance (or highest coherence effect size) from the triplets of each monkey for subsequent modelling. Over both monkeys, this resulted in 20 Va-V1b-V4 triplets. For each of those triplets, the data that entered DCM—namely, an array of CSD values for frequencies 0-100 Hz among V1a, V1b and V4—is referred to as CSD data and is made available as specified in the “Key Resources Table”.

Dynamic Causal Model

DCM for CSD is using Bayesian statistical methods to fit a physiologically-informed, neural mass model of neuronal activity to data that has been observed experimentally, in order to infer physiological parameters (Friston et al.,

2012). DCM for CSD is particularly suited to study neuronal oscillations as it is operating in the frequency domain. Here, the frequencies of interest were 0 to 100 Hz.

The model used in this paper consists of 3 nodes modelling activity in cortical areas V1a, V1b and V4. Each node is represented by a canonical cortical microcircuit, which comprises 4 neuronal populations: a pyramidal neuron population in the supragranular layer, ***s***, a pyramidal neuron population in the deep layer, ***d***, an excitatory population in the granular or input layer, ***g***, and a population of inhibitory interneurons, ***i***, that sends afferents to all layers (Bastos et al., 2012) (Auksztulewicz and Friston, 2015). The connections between those neuronal populations within a local microcircuit are referred to as intrinsic connections. The connections between the different microcircuits are referred to as extrinsic connections and come in two flavors: feedforward, from V1 to V4, and feedback, from V4 to V1. For the detailed pattern of intrinsic and extrinsic connections, see **Figure 1C and 2**.

The neuronal model described above is complemented by an observation model that accounts for the measurement of the data. The observation model describes the noise that is inherent to the recording of the data, and how each element of the model contributes to the observed CSD. For the present study, a parameter (w) that accounts for the pre-whitening of the data was added to the observation model.

The neuronal and observation models come together to form the generative model for the CSD data, which is a statistical model of the joint probability of the parameters θ and the data y:

$$\ln P(y,\theta|m) = \ln P(y|\theta,m) + \ln P(\theta,m)$$
$$P(y|\theta,m) = N\big(\Gamma(\theta),\Sigma(\theta)\big)$$
$$P(\theta,m) = N(\lambda,\sigma)$$

The model m is defined by the model structure drawn in **Figure 2** and the priors. Γ expresses the mapping of the model parameters to the model spectral features that are fitted to the experimentally observed data, and therefore encompasses the differential equations that describe the neuronal dynamics shown in **Figure 2**. N denotes a normal distribution with the indicated mean and variance. $\Sigma(\theta)$ and σ are the variance of the parameters and of the hyper-parameters, respectively.

The prior mean and variance for the parameters of the neuronal and the observation model can be found in **Table 1**. The MATLAB toolbox that was used for constructing, fitting and analyzing the DCM was SPM12. The SPM12 functions, where each DCM parameter can be found, are also included in **Table 1**.

Attentional modulation in the DCM

In any given trial, one of the two V1 sites will be attended, and the other V1 site will be unattended. This means that the two conditions need to be modelled simultaneously. In both conditions, attentional modulation is modelled as a multiplier β that modulates the strength of intrinsic V1a and V1b connections. Conceptually, there exists in the model a "baseline" connection strength γ that remains constant between conditions (and that does not correspond to any experimental condition by itself), and is multiplied with β or $-\beta$ to calculate the final connection strength. Thus, the condition-specific connection strength is $e^{\gamma+c\beta}$, where c is the condition effect and is $+1$ for the "attend V1a" and -1 for the "attend V1b" condition respectively.

The ratio of "attend in" over "attend out" connection strength, shown in the schematic representations in **Figure 3D and 5D** is evaluated as:

$$e^{\gamma+c\beta}/e^{\gamma-c\beta} = e^{\gamma+\beta-\gamma+\beta} = e^{2\beta}$$

The estimates for β, as obtained directly from fitting the data, are expected to show opposite effects for V1a and V1b sites, since for V1a sites they represent the modulation from the unattended condition to the attended one, and for V1b they represent the modulation from the attended to the unattended condition. Results from the two

attention conditions were combined as explained below in the Methods section "Quantifying attentional effects across triplets".

Note that to ensure that scale parameters (like rate and time constants) have non-negative values, all DCM parameters (except the electrode gains) have log-normal priors.

Notes on the observation model

As mentioned above, the neuronal model is complemented by an observation model, which describes among others how each element of the model contributes to the observed spectra. This entails that the power spectrum observed from the model could in principle be composed from many different combinations of power spectra across the modeled neuronal populations. We choose to constrain the observation model to reflect exclusively the supragranular pyramidal population. This was based on the fact that our recordings were obtained with ECoG electrodes placed on the surface of the cortex. LFP attenuates with distance following an inverse square law, such that ECoG electrode signals are dominated by supragranular neuron activity (Destexhe and Bedard, 2013; Nunez and Srinivasan, 2006).

Note that in the DCM, only the supragranular pyramidal V1 population projects to V4. Thereby, V1-V4 coherence is directly influenced by the power spectra of the activity in the supragranular pyramidal V1 population (and merely indirectly by other V1 populations). Correspondingly, the supragranular pyramidal V1 power spectra should reflect the observed data faithfully, as obtained by the mentioned constraint on the observation model. Note further that the observed data showed no consistent difference in gamma power between attention conditions, and therefore one of the aims of our modeling effort was to investigate whether the strong attentional effect on V1-V4 coherence could be explained in the absence of attentional effects on the relevant, i.e., supragranular, V1 power.

QUANTIFICATION AND STATISTICAL ANALYSIS

Multi-start Variational Laplace

Variational Laplace (VL) (Friston et al., 2007) performs gradient descent on free energy (see section "Free energy and model comparison" below) in alternating steps of optimizing the parameters to minimize free energy for a set of hyper-parameters and vice versa (optimizing the hyper-parameters to minimize free energy for a set of parameters). The algorithm runs for a given number of steps ($N_{steps} = 128$ in our case) or until convergence, but it is not guaranteed to converge to a global minimum. Therefore, it is common practice to restart the algorithm in search for better parameters. In each restart, the last posterior estimates of the observation and neuronal model parameters are used as an initial point for a new gradient descent on free energy. Meanwhile, the hyper-parameters are reset to their prior values. This allows for a wider search in the parameter space, in the vicinity of the last posterior estimates.

Here, we implemented a multi-start version of this scheme. Every time VL was restarted, uniform noise with a maximum range of 1 or 2 was added to the estimated posterior means obtained from the previous VL execution, before they were used as a starting point for the next execution. After executing VL, using the parameter and hyperparameter priors as an initialization, we run a total of 500 randomizations for each of the two noise levels; each randomization consisted of another 9 consecutive VL restarts for each triplet, with noise added in between. This resulted in $1\ initialization + 9\ restarts * 500\ randomizations * 2\ noise\ levels\ =\ 9001\ models$ (or sets of parameter estimates) for each triplet.

The data features that were considered when fitting the data were a weighted mixture of the CSD, the cross-covariance function and the MVAR coefficients calculated from the data. Their relative weights were 1, 1 and 1/8 correspondingly.

The hyperparameter prior mean was smaller than in previous DCM studies on EEG, because the ECoG data contain less variance.

Free energy and model comparison

As a result of fitting the data with a stochastic version of VL, we could have obtained 9001 models for each triplet. Due to convergence failure we obtained less models for some triplets (1842/360040 or 0.0051% of all possible models missing; more specifically for monkey P: less than 7 models missing for 2 triplets in the DCM with feedback, 8 models missing for 1 triplet in the DCM without feedback; monkey K: less than 183 models missing for all triplets in the DCM with feedback and less than 10 models missing for 3 triplets in DCM without feedback).

For each model, negative free energy is calculated as:

$$F(q) = E_q[\ln p(y|\theta)] - D_{KL}[q||p]$$

where q and p are the approximate posterior and prior distributions respectively (Friston et al., 2007). The first term is the expectation under the posterior probability q of the likelihood of the data given the parameters θ and the model architecture, and it expresses the accuracy of the model and its parameters. The second term is the Kullback-Leibler divergence of the posterior from the prior probability of the parameters; it expresses the complexity of the model, and it imposes a penalty for moving away from the prior parameter values to fit the data.

It follows that models with higher free energy (corresponding to a higher ELBO) afford a better explanation of the data. Therefore, free energy could be used to sort the models obtained for each triplet. For each triplet, we selected the model with the highest evidence and exclusively used those winning models for further analyses.

Quantifying attentional effects across triplets

To calculate the attentional modulation of the 12 connections within a V1 node, the estimates of V1a and V1b modulation needed to be pooled across all 20 triplets from both monkeys.

Some V1 sites appeared in multiple triplets (see labels in **Figure S1**for monkey P). This could introduce a bias when combining the estimated parameters from all triplets, as their V1 spectra would be predicted by similar parameter posterior distributions. To avoid this bias, our first step was to estimate attentional modulation β over the triplets where a site was repeated, using PEB (Friston et al., 2016; Zeidman et al., 2019). Generally, PEB assumes a hierarchical model where the distribution of parameters ($\theta^{(1)}$) in one level (i.e. over subjects) can be modelled by a Gaussian distribution that is defined on the level above. For the attentional modulation parameters β, the PEB model can be defined as:

$$y_k = \Gamma_k^{(1)}\left(\beta_k^{(1)}\right) + X_0 \mathrm{b}_k + \varepsilon_k^{(1)}$$
$$\beta^{(1)} = X\beta^{(2)} + \varepsilon^{(2)}$$
$$\beta^{(2)} = \eta + \varepsilon^{(3)}$$

The first row expresses the fact that the data y_k for each subject k is generated by the DCM with the function $\Gamma_k^{(1)}$, which is a nonlinear mapping from the first-level attentional modulation parameters $\beta_k^{(1)}$ to the data of a subject k. Here, we consider the multiple iterations of a repeated site as subjects. The mean of the signal, which can vary between recordings and is irrelevant to the calculation of the attentional effects, can be modelled by a general linear model (GLM) with design matrix X_0 and parameters b_k. Here, we used a prior variance $C_{b_k} = 1/32$. The additive Gaussian error term $\varepsilon_k^{(1)}$ models the observation noise. The second row of the equation corresponds to the second level of PEB, and it shows that the vector of first-level parameters $\beta^{(1)}$ can be modelled by a GLM with design matrix X and group-level parameters $\beta^{(2)}$ and with the addition of between-subject variability $\varepsilon^{(2)}$. The second-level parameters have their own prior mean η and variability $\varepsilon^{(3)}$, as expressed in the third row.

The estimated second-level PEB parameters correspond to the attentional modulation for a unique V1 site, which can be combined with the estimated DCM parameters of the attentional modulation for the V1 sites that were not

repeated over triplets. This was implemented again using PEB, where all unique sites are considered as subjects and with an assumed prior variance of $C'_b = 1/64$.

This means that the second PEB model combined the results from V1a and V1b sites. However, they were expected to have opposite signs of attentional modulation in each connection due to the way DCM deals with the experimental design. For example, if the V1a supragranular pyramidal self-inhibition was increased to change the V1a node response from the unattended to the attended state, in the same trial, the corresponding V1b connection was expected to decrease to change the V1b response from an attended to an unattended state. To accommodate for this anti-symmetry, we simply inverted the sign of the posterior V1b modulations, since they are log normal and therefore have symmetrical magnitude between conditions. As a simplification, the resulting final estimate of the attentional modulation over all unique V1 sites is referred to in the main text and figures as β or "log modulation".

Hypothesis testing with family-wise model comparison

Reduced models can be calculated efficiently using BMR (Friston et al., 2016) for all possible combinations of the 12 intrinsic V1 connections allowed to change with attention, or not (12 and not 24 because the modulation needs to be applied with opposite sign in both V1a and V1b, see also section "Attentional modulation in the DCM"). Each combination is a reduced model in the sense that its free parameters are a subset of those in the full model that we have fitted to the data (in our case these are the intrinsic V1 connectivity parameters).

Using BMR, the evidence and the parameters of a reduced model can be calculated directly from the full model, without the need to fit the reduced model to data. The set of all reduced models ($N_{models} = 2^{12} = 4096$) is partitioned into non-overlapping subsets of models (families) following a factorial design; models are allocated to families according to whether or not particular connections of interest relevant to a hypothesis change with attention. This means that other connections might also be modulated in some of these models, but this does not affect their allocation to a family for testing this hypothesis. The prior probability of each family is $p(f_l) = 1/L$, where L is the total number of families. This means that we assume uniform priors at the family level to avoid unwanted bias. The posterior probability for the family is calculated by summing the posterior probabilities of the models that belong to the family (Penny et al., 2010):

$$p(f_l|Y) = \sum_{m \in f_l} p(m|Y)$$

By performing inference at the level of families of models we prevent evidence dilution across the many models that satisfy a more general hypothesis about the attentional mechanism.

Finally, we perform BMA for inference at the level of parameters within the winning family (Litvak et al., 2015; Penny et al., 2010). BMA summarizes family-specific parameters by marginalizing the posterior probability of parameters θ over specific models $m \in f_l$.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
CSD files	This paper	
Experimental models: organisms/strains		
M. mulatta	German Primate Center	https://www/dpz.eu/en/home.html
Software and algorithms		
MATLAB	MathWorks	https://www.mathworks.com/
FieldTrip	(Oostenveld et al., 2011)	https://www.fieldtriptoolbox.org/
SPM12	(Ashburner et al., 2014)	https://www.fil.ion.ucl.ac.uk/spm/software/spm12/
ARfit	(Neumaier and Schneider, 2001), (Schneider and Neumaier, 2001)	https://www.mathworks.com/matlabcentral/fileexchange/174-arfit
BSMART	(Cui et al., 2008)	https://brain-smart.org/
Original code	This paper	

References

Akam, T., and Kullmann, D.M. (2010). Oscillations and filtering networks support flexible routing of information. Neuron *67*, 308-320.

Akam, T.E., and Kullmann, D.M. (2012). Efficient "communication through coherence" requires oscillations structured to minimize interference between signals. PLoS Comput Biol *8*, e1002760.

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

Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K.J., Kiebel, S., Kilner, J., Litvak, V., Moran, R., and Penny, W. (2014). SPM12 manual. W.T.C.f. Neuroimaging, ed.

Auksztulewicz, R., and Friston, K. (2015). Attentional Enhancement of Auditory Mismatch Responses: a DCM/MEG Study. Cereb Cortex *25*, 4273-4283.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. Neuron *76*, 695-711.

Bauer, M., Kluge, C., Bach, D., Bradbury, D., Heinze, H.J., Dolan, R.J., and Driver, J. (2012). Cholinergic enhancement of visual attention and neural oscillations in the human brain. Curr Biol *22*, 397-402.

Besserve, M., Lowe, S.C., Logothetis, N.K., Schölkopf, B., and Panzeri, S. (2015). Shifts of Gamma Phase across Primary Visual Cortical Sites Reflect Dynamic Stimulus-Modulated Information Transfer. PLoS Biol *13*, e1002257.

Bichot, N.P., Heard, M.T., DeGennaro, E.M., and Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. Neuron *88*, 832-844.

Börgers, C., and Kopell, N.J. (2008). Gamma oscillations and stimulus selection. Neural Comput *20*, 383-414.

Bosman, C.A., Schoffelen, J.M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., and Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron *75*, 875-888.

Buehlmann, A., and Deco, G. (2010). Optimal information transfer in the cortex through synchronization. PLoS Comput Biol *6*.

Burwick, T., and Bouras, A. (2017). Selective Interareal Synchronization through Gamma Frequency Differences and Slower-Rhythm Gamma Phase Reset. Neural Comput *29*, 643-678.

Buzsáki, G. (2006). Rhythms of the Brain (Oxford University Press).

Buzsáki, G., and Wang, X.J. (2012). Mechanisms of gamma oscillations. Annu Rev Neurosci *35*, 203-225.

Cannon, J., McCarthy, M.M., Lee, S., Lee, J., Börgers, C., Whittington, M.A., and Kopell, N. (2014). Neurosystems: brain rhythms and cognitive processing. Eur J Neurosci *39*, 705-719.

Cardin, J.A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.H., and Moore, C.I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature *459*, 663-667.

Chen, G., Zhang, Y., Li, X., Zhao, X., Ye, Q., Lin, Y., Tao, H.W., Rasch, M.J., and Zhang, X. (2017). Distinct Inhibitory Circuits Orchestrate Cortical beta and gamma Band Oscillations. Neuron *96*, 1403-1418 e1406.

Chen, M., Li, B., Guang, J., Wei, L., Wu, S., Liu, Y., and Zhang, M. (2016). Two subdivisions of macaque LIP process visual-oculomotor information differently. Proc Natl Acad Sci U S A *113*, E6263-e6270.

Cohen, M.X. (2014). Analyzing neural time series data: Theory and practice (MIT Press).

Constantinidis, C., and Steinmetz, M.A. (2005). Posterior parietal cortex automatically encodes the location of salient stimuli. J Neurosci *25*, 233-238.

Cui, J., Xu, L., Bressler, S.L., Ding, M., and Liang, H. (2008). BSMART: a Matlab/C toolbox for analysis of multichannel neural time series. Neural Netw *21*, 1094-1104.

Desimone, R., and Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. Annu Rev Neurosci *18*, 193-222.

Destexhe, A., and Bedard, C. (2013). Local field potential. Scholarpedia *8*, 10713.

DeWeerd, P., Peralta, M.R., Desimone, R., and Ungerleider, L.G. (1999). Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. Nat Neurosci *2*, 753-758.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., and Reitboeck, H.J. (1988). Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. Biol Cybern *60*, 121-130.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex *1*, 1-47.

Ferro, D., van Kempen, J., Boyd, M., Panzeri, S., and Thiele, A. (2021). Directed information exchange between cortical layers in macaque V1 and V4 and its modulation by selective attention. Proc Natl Acad Sci U S A *118*.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci *9*, 474-480.

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

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

Friston, K.J., Bastos, A., Litvak, V., Stephan, K.E., Fries, P., and Moran, R.J. (2012). DCM for complex-valued data: Cross-spectra, coherence and phase-delays. Neuroimage *59*, 439-455.

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B.C.M., Ziegler, G., and Zeidman, P. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. Neuroimage *128*, 413-431.

Friston, K.J., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the Laplace approximation. Neuroimage *34*, 220-234.

Gray, C.M., König, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature *338*, 334-337.

Gregoriou, G.G., Gotts, S.J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science *324*, 1207-1210.

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

Grothe, I., Neitzel, S.D., Mandon, S., and Kreiter, A.K. (2012). Switching neuronal inputs by differential modulations of gamma-band phase-coherence. J Neurosci *32*, 16172-16180.

Herrero, J.L., Roberts, M.J., Delicato, L.S., Gieselmann, M.A., Dayan, P., and Thiele, A. (2008). Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. Nature *454*, 1110-1114.

Herrington, T.M., Masse, N.Y., Hachmeh, K.J., Smith, J.E., Assad, J.A., and Cook, E.P. (2009). The effect of microsaccades on the correlation between neural activity and behavior in middle temporal, ventral intraparietal, and lateral intraparietal areas. J Neurosci *29*, 5793-5805.

Howe, W.M., Gritton, H.J., Lusk, N.A., Roberts, E.A., Hetrick, V.L., Berke, J.D., and Sarter, M. (2017). Acetylcholine Release in Prefrontal Cortex Promotes Gamma Oscillations and Theta-Gamma Coupling during Cue Detection. J Neurosci *37*, 3215-3230.

Ipata, A.E., Gee, A.L., Bisley, J.W., and Goldberg, M.E. (2009). Neurons in the lateral intraparietal area create a priority map by the combination of disparate signals. Exp Brain Res *192*, 479-488.

Kass, R.E., and Raftery, A.E. (1995). Bayes Factors. Journal of the American Statistical Association *90*, 773-795.

Kim, T., Thankachan, S., McKenna, J.T., McNally, J.M., Yang, C., Choi, J.H., Chen, L., Kocsis, B., Deisseroth, K., Strecker, R.E.*, et al.* (2015). Cortically projecting basal forebrain parvalbumin neurons regulate cortical gamma band oscillations. Proc Natl Acad Sci U S A *112*, 3535-3540.

Lee, S., Kruglikov, I., Huang, Z.J., Fishell, G., and Rudy, B. (2013). A disinhibitory circuit mediates motor integration in the somatosensory cortex. Nat Neurosci *16*, 1662-1670.

Litvak, V., Garrido, M., Zeidman, P., and Friston, K. (2015). Empirical Bayes for Group (DCM) Studies: A Reproducibility Study. Front Hum Neurosci *9*, 670.

Lowet, E., Roberts, M., Hadjipapas, A., Peter, A., van der Eerden, J., and De Weerd, P. (2015). Input-dependent frequency modulation of cortical gamma oscillations shapes spatial synchronization and enables phase coding. PLoS Comput Biol *11*, e1004072.

Lowet, E., Roberts, M.J., Peter, A., Gips, B., and De Weerd, P. (2017). A quantitative theory of gamma synchronization in macaque V1. Elife *6*.

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

Ma, G., Liu, Y., Wang, L., Xiao, Z., Song, K., Wang, Y., Peng, W., Liu, X., Wang, Z., Jin, S*., et al.* (2021). Hierarchy in sensory processing reflected by innervation balance on cortical interneurons. Sci Adv *7*, eabf5676.

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

Neumaier, A., and Schneider, T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Transactions on Mathematical Software *27*, 27-57.

Ni, J., Wunderle, T., Lewis, C.M., Desimone, R., Diester, I., and Fries, P. (2016). Gamma-Rhythmic Gain Modulation. Neuron *92*, 240-251.

Nunez, P.L., and Srinivasan, R. (2006). Electric fields of the brain: The neurophysics of EEG (Oxford University Press).

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci *2011*, 156869.

Palmigiano, A., Geisel, T., Wolf, F., and Battaglia, D. (2017). Flexible information routing by transient synchrony. Nat Neurosci *20*, 1014-1022.

Penny, W.D., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K.J., Schofield, T.M., and Leff, A.P. (2010). Comparing families of dynamic causal models. PLoS Comput Biol *6*, e1000709.

Penny, W.D., Stephan, K.E., Mechelli, A., and Friston, K.J. (2004). Comparing dynamic causal models. Neuroimage *22*, 1157-1172.

Peter, A., Stauch, B.J., Shapcott, K.A., Kouroupaki, K., Schmiedt, J.T., Klein, L., Klon-Lipok, J., Dowdall, J.R., Schölvinck, M.L., Vinck, M*., et al.* (2021). Stimulus-specific plasticity of macaque V1 spike rates and gamma. Cell Rep *37*, 110086.

Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., and Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. Nat Neurosci *16*, 1068-1076.

Ray, S., and Maunsell, J.H. (2010). Differences in gamma frequencies across visual cortex restrict their possible use in computation. Neuron *67*, 885-896.

Reynolds, J.H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci *19*, 1736-1753.

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

Richter, C.G., Thompson, W.H., Bosman, C.A., and Fries, P. (2017). Top-Down Beta Enhances Bottom-Up Gamma. J Neurosci *37*, 6698-6711.

Roberts, M.J., Lowet, E., Brunet, N.M., Ter Wal, M., Tiesinga, P., Fries, P., and De Weerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. Neuron *78*, 523-536.

Rohenkohl, G., Bosman, C.A., and Fries, P. (2018). Gamma Synchronization between V1 and V4 Improves Behavioral Performance. Neuron *100*, 953-963 e953.

Sapountzis, P., Paneri, S., and Gregoriou, G.G. (2018). Distinct roles of prefrontal and parietal areas in the encoding of attentional priority. Proc Natl Acad Sci U S A *115*, E8755-e8764.

Schmitz, T.W., and Duncan, J. (2018). Normalization and the Cholinergic Microcircuit: A Unified Basis for Attention. Trends Cogn Sci *22*, 422-437.

Schneider, M., Broggini, A.C., Dann, B., Tzanou, A., Uran, C., Sheshadri, S., Scherberger, H., and Vinck, M. (2021). A mechanism for inter-areal coherence through communication based on connectivity and oscillatory power. Neuron *109*, 4050-4067.e4012.

Schneider, T., and Neumaier, A. (2001). Algorithm 808: ARfit—a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Transactions on Mathematical Software *27*, 58–65.

Sohal, V.S., Zhang, F., Yizhar, O., and Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature *459*, 698-702.

Spyropoulos, G., Dowdall, J.R., Schölvinck, M.L., Bosman, C.A., Lima, B., Peter, A., Onorato, I., Klon-Lipok, J., Roese, R., Neuenschwander, S.*, et al.* (2020). Spontaneous variability in gamma dynamics described by a linear harmonic oscillator driven by noise. bioRxiv.

Stauch, B.J., Peter, A., Schuler, H., and Fries, P. (2021). Stimulus-specific plasticity in human visual gamma-band activity and functional connectivity. Elife *10*, e68240.

Thiele, A., and Bellgrove, M.A. (2018). Neuromodulation of Attention. Neuron *97*, 769-785.

Thompson, K.G., and Bichot, N.P. (2005). A visual salience map in the primate frontal eye field. Prog Brain Res *147*, 251-262.

Veit, J., Hakim, R., Jadi, M.P., Sejnowski, T.J., and Adesnik, H. (2017). Cortical gamma band synchronization through somatostatin interneurons. Nat Neurosci *20*, 951-959.

Veit, J., Handy, G., Mossing, D.P., Doiron, B., and Adesnik, H. (2021). Cortical VIP neurons locally control the gain but globally control the coherence of gamma band rhythms. bioRxiv, 2021.2005.2020.444979.

Veith, V.K., Quigley, C., and Treue, S. (2021). Cholinergic manipulations affect sensory responses but not attentional enhancement in macaque MT. BMC Biol *19*, 49.

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

Wardak, C., Ibos, G., Duhamel, J.R., and Olivier, E. (2006). Contribution of the monkey frontal eye field to covert visual attention. J Neurosci *26*, 4228-4235.

Womelsdorf, T., Schoffelen, J.M., Oostenveld, R., Singer, W., Desimone, R., Engel, A.K., and Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. Science *316*, 1609-1612.

Zeidman, P., Jafarian, A., Seghier, M.L., Litvak, V., Cagnan, H., Price, C.J., and Friston, K.J. (2019). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. Neuroimage *200*, 12-25.

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science *345*, 660-665.

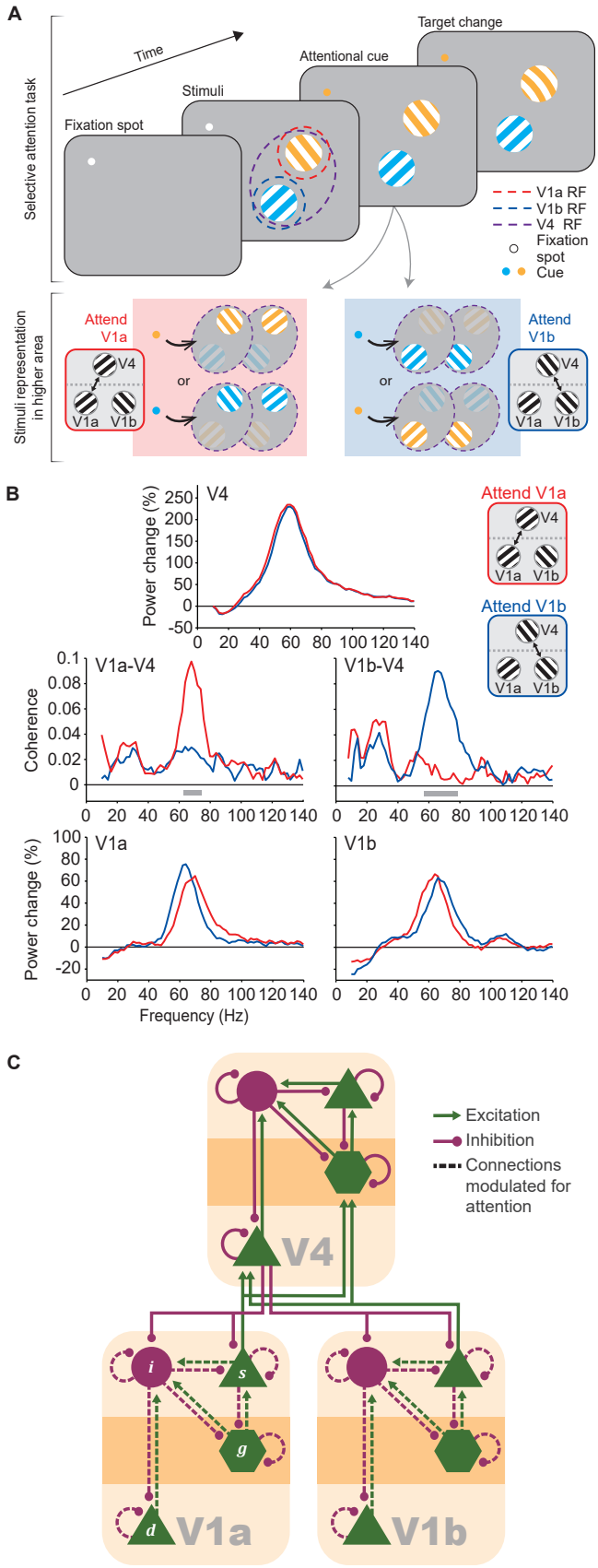


Figure 1. Modelling the effects of selective attention on gamma-band neuronal synchronization in the visual cortex.

(A) Top: Selective attention task (Bosman et al., 2012). The monkey had to maintain fixation on the fixation spot. The stimuli consisted of two iso-luminant gratings, one in the V1a RF, the other in the V1b RF, and both contained in the V4 RF. RFs are illustrated with dashed circles, not visible to the monkey. The color of the fixation splot (cue) indicated which of the two gratings the monkey had to attend.
Bottom: Attentional cueing scheme, and resulting stimuli representation in higher area. Left panel: "Attend V1a" condition. The cue has the same color as the stimulus in the V1a RF, and primarily this attended stimulus is communicated to and represented in the higher area V4, as illustrated in the red-lined inset. The full attentional cueing scheme involves counterbalacing of cue color, stimulus color and stimulus orientation. The attentional filtering out of the unattended stimulus is illustrated by a gray masking of the V1b stimulus. Right panel: Same for the "Attend V1b" condition.
(B) Power and coherence of example V1 and V4 sites from monkey P, while the monkey was instructed to attend the stimulus in the V1a (red lines) or V1b (blue lines) RF. (Top) V4 power relative to baseline; (Bottom) V1 power relative to baseline; (Middle) V1-V4 coherence spectrum. Modified from (Bosman et al., 2012).
(C) The V1a-V1b-V4 microcircuit that was used to model the CSD measured from the data. Each area consisted of 4 subpopulations modelled as neural masses: an excitatory population in the granular layer g, a pyramidal population in the supragranular layers s, a pyramidal population in the deep layers d and a population of inhibitory interneurons i. The pattern of intrinsic (within area) and extrinsic (between areas) connectivity is shown. Note that inhibitory (magenta) connections that appear to stem from excitatory (green) populations can be conceived as being mediated by local inhibitory interneurons. Extrinsic connections are depicted as bifurcating to simplify the illustration, but the strength of each of the 8 extrinsic connections was fitted independently. Dashed lines represent the connections whose strength could vary between the two attentional conditions.

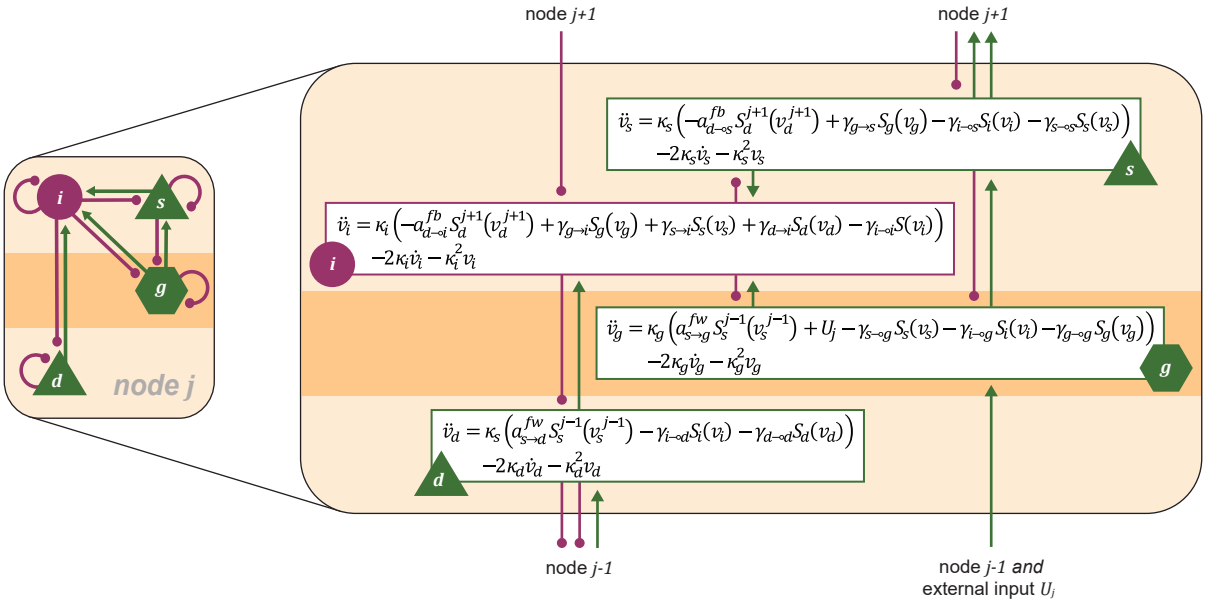


Figure 2. The canonical microcircuit.

Equations of motion for the 4 neural masses within a canonical microcircuit node. v_x, depolarization of population x; κ_x, inverse of time constant of population x; $a^{fw}_{x\to y}$, extrinsic feedforward excitatory connection strength from population x to population y; $a^{fb}_{x\to y}$, extrinsic feedback inhibitory connection strength from population x to population y; $\gamma_{x\to y}$, intrinsic excitatory connection strength from population x to population y; $\gamma_{x\to y}$, intrinsic inhibitory connection strength from population x to population y; S_x, sigmoid function transforming the depolarization of population x to a firing rate; U, external input firing rate; $\dot{v}$, first temporal derivative of depolarization; $\ddot{v}$, second temporal derivative of depolarization.

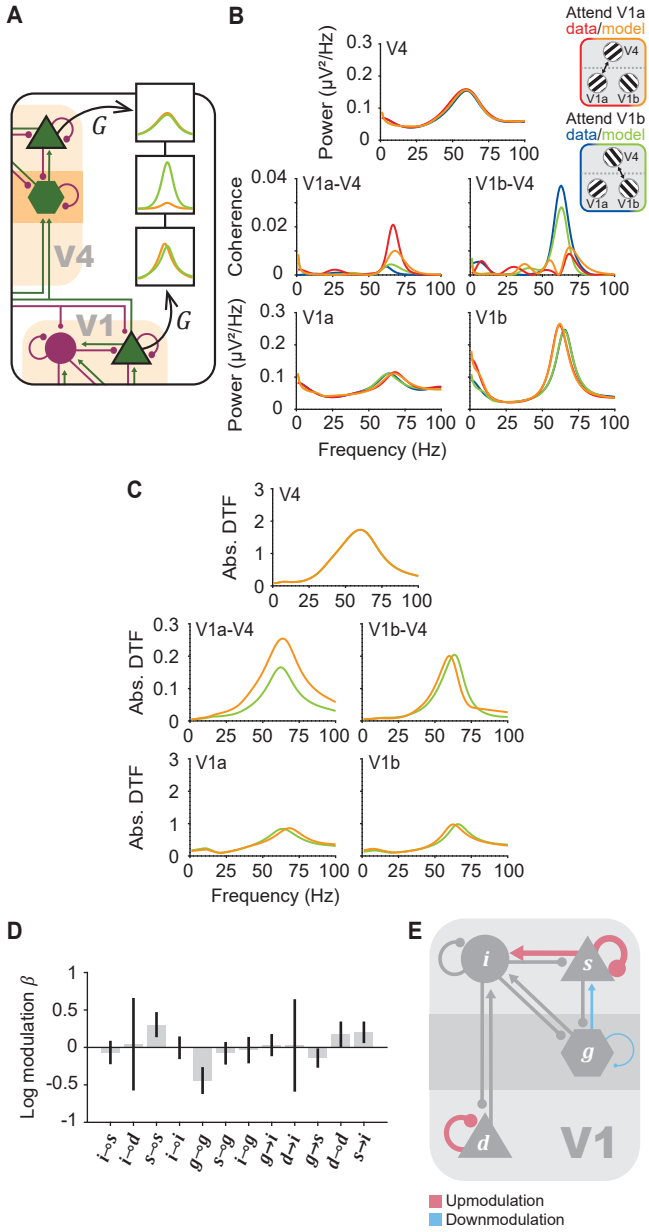


Figure 3. DCM of the attentional effects.

(A) Schematic representation of the observation model used in the present study: The LFP of a node reflects the depolarization of the respective supragranular pyramidal population, weighted by the parameters of the observation model G.
(B) Power spectral density and coherence spectra averaged across the 10 V1a-V1b-V4 triplets of monkey P. Red (orange): attend V1a in the data (model), blue (green): attend V1b in the data (model). Top: V4 power spectrum, bottom: V1a and V1b power spectra, middle: V1a-V4 and V1b-V4 coherence spectrum.
(C) Absolute DTF averaged over the 10 V1a-V1b-V4 DCMs of monkey P.
(D) PEB posterior estimate of the attentional modulation β, combining data from both monkeys. This parameterizes the log ratio of connection strength in the "attend in" over the "attend out" condition for each intrinsic V1 connection. Error bars indicate 95% Bayesian confidence intervals (credible intervals), computed from the leading diagonal of the covariance matrix. Generally speaking, the effects whose credible intervals do not cross the zero line are "significant" in the sense that Bayesian model comparison provides "very strong" evidence for their presence (Kass and Raftery, 1995).
(E) Schematic representation of the "attend in" over the "attend out" connection strength ratio for connections significantly modulated with attention. The ratio is equivalent to twice the β value and is presented as the width of the pink- and blue-colored connections, relative to the width of grey-colored connections for the connections that are modulated by attention. Note that the strengths (γ) of the remaining connections, which are shown as grey lines, have also been fitted to best explain the data features but have the same value for the two attentional conditions (i.e. "attend in" over the "attend out" ratio is equivalent to 1).
See also supplementary Figures S1, S2.

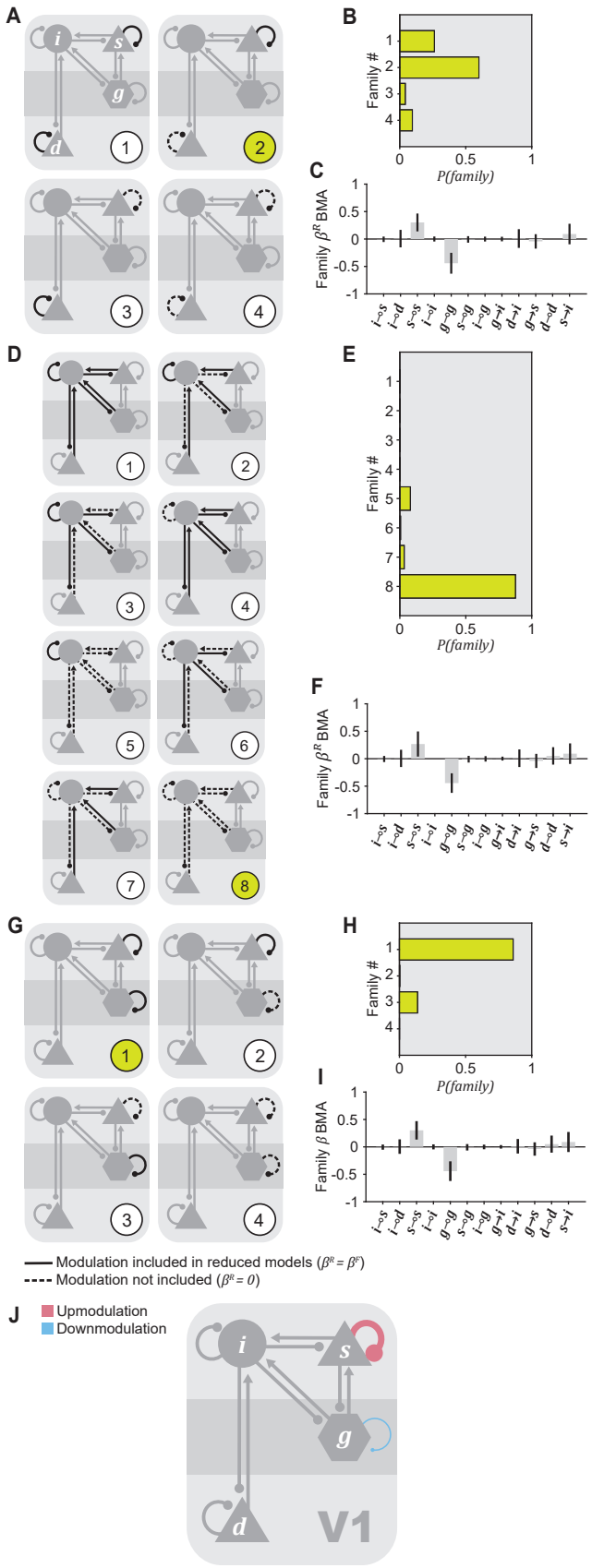


Figure 4. The role of inhibition in attentional modulation.

(A-C) Hypotheses about the role of pyramidal population self-inhibition in the supragranular and deep layers.
(A) Schematic representation of the four hypotheses shown on an exemplar V1 microcircuit. Connections shown in continuous black lines are modulated by attention, connections shown in dashed black lines are not significantly modulated, whereas connections shown in grey may or may or not be modulated.
(B) Posterior probability of each family corresponding to a hypothesis shown in (A). The index of the family with the highest posterior probability is highlighted with a green number in (A).
(C) BMA of the attentional modulation β^R across all reduced models that belong to the winning family.
(D-F) Same as (A-C), but for hypotheses about the modulation of input or response gain of the inhibitory population.
(G-I) Same as (A-C), but for hypotheses about the role of self-inhibition in the supragranular pyramidal and in the granular population.
(J) Schematic representation of the results shown in (G-I), specifically the “attend in” over the “attend out” connection strength ratio for connections significantly modulated with attention. Conventions as in Figure 3E. Here, the significance of attentional modulation for each connection is considered relative to the BMA of β^R.

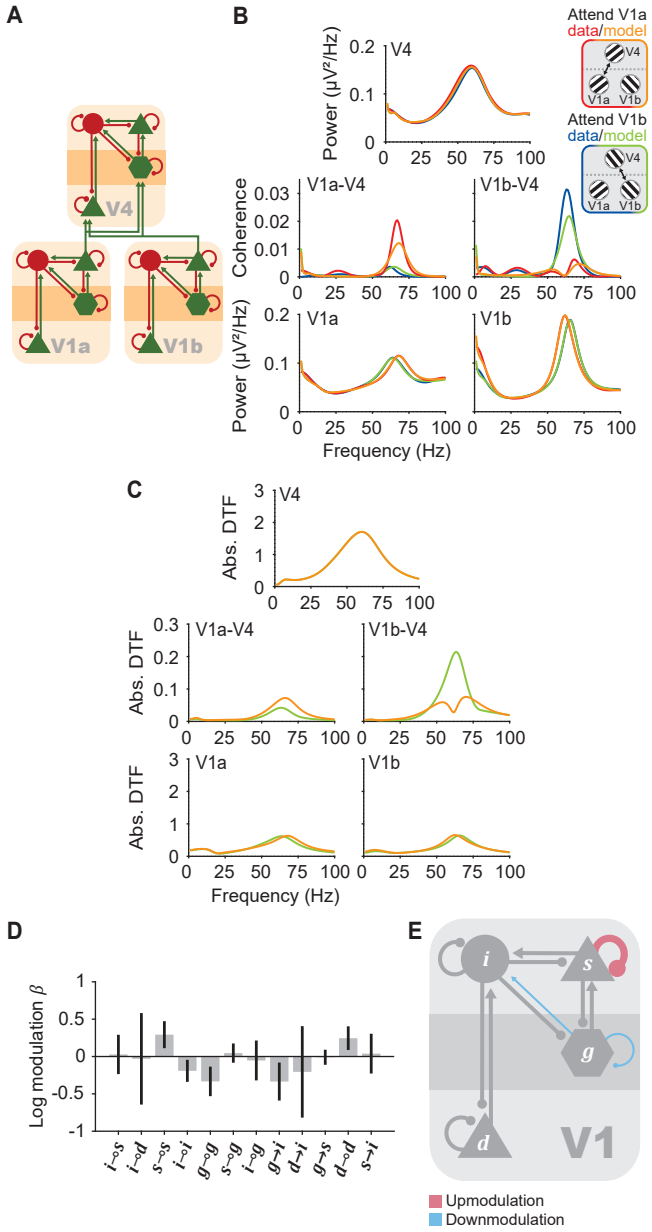
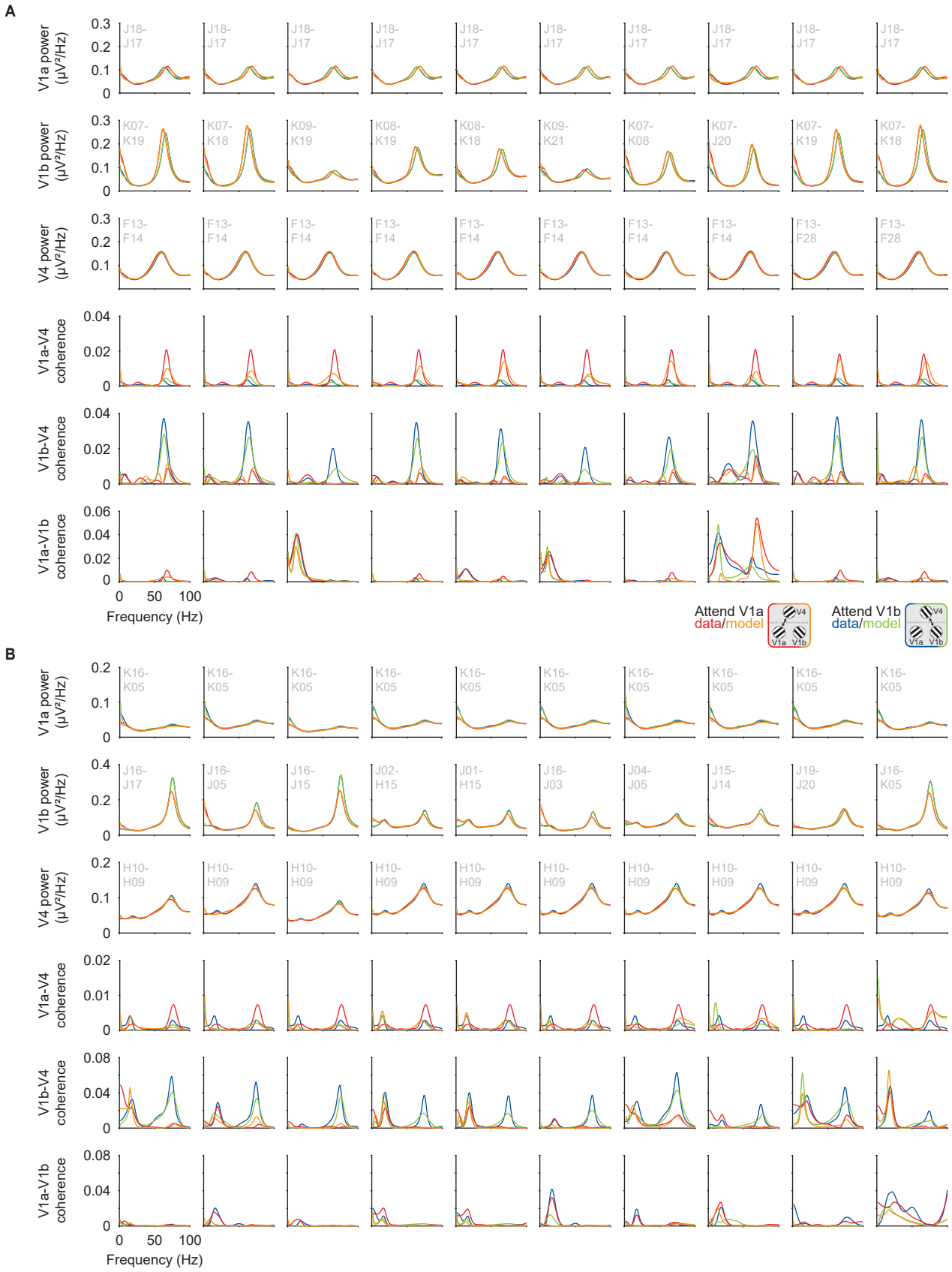


Figure 5. Attentional modulation in the no-feedback architecture.

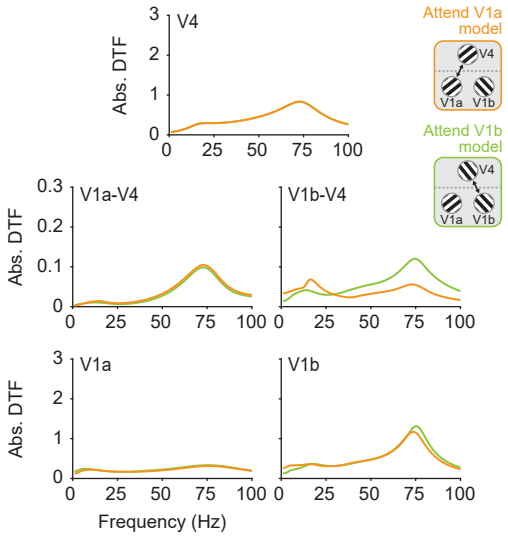
(A) Schematic representation of the V1a-V1b-V4 microcircuit without V4-to-V1 feedback.
(B) Power spectral density and coherence spectra averaged across the 10 V1a-V1b-V4 triplets of monkey P.
(C) Absolute DTFs averaged across the 10 V1a-V1b-V4 DCMs of monkey P.
(D, E) PEB posterior estimate combining data from both monkeys, and schematic representation of the attentional modulation β. Conventions as in Figures 3D and 3E.
See also supplementary Figure S3.

Parameter		Prior mean	Prior variance	SPM 12 function
VL hyperparameters	λ	12	$1/64$	*spm_dcm_csd.m*
Neuronal model				
Intrinsic time constants	$\tau_g, \tau_s, \tau_i, \tau_d$	$2, 2, 16, 18$	$1/64, 1/64, 1/64, 1/64$	*spm_fx_cmc.m, spm_cmc_priors.m*
Intrinsic connection strength	$\gamma_{g\to g}, \gamma_{g\to s}, \gamma_{g\to i}$	$800, 800, 800$	$1/32, 1/32, 1/32$	
	$\gamma_{s\to s}, \gamma_{s\to g}, \gamma_{s\to i}$	$800, 800, 400$	$1/32, 1/32, 1/32$	
	$\gamma_{i\to i}, \gamma_{i\to g}, \gamma_{i\to s}, \gamma_{i\to d}$	$800, 1600, 800, 400$	$1/32, 1/32, 1/32, 1/32$	
	$\gamma_{d\to d}, \gamma_{d\to i}$	$400, 400$	$1/32, 1/32$	
Extrinsic connection strength (forward)	$\alpha^{fw}_{s\to g}, \alpha^{fw}_{s\to d}$	$200, 25$	$1/16, 1/16$	
Extrinsic connection strength (feedback)	$\alpha^{fb}_{d\to s}, \alpha^{fb}_{d\to i}$	$100, 50$	$1/16, 1/16$	
Sigmoid gain	σ	1	$1/32$	
Delays	δ	0	$1/64$	*spm_cmc_priors.m*
Conditional effects	β	0	$1/8$	*spm_dcm_neural_priors.m*
Observation model				
Gain	L	1	64	*spm_L_priors.m*
Contribution to signal	J_g, J_s, J_i, J_i	$0, 1, 0, 0$	$0, 0, 0, 0$	
Neuronal fluctuations	a	0	$1/128$	*spm_ssr_priors.m*
Non-specific noise	b	0	$1/128$	
Channel-specific noise	c	0	$1/128$	
Neuronal innovations	d	0	$1/128$	
Whitening coefficients	w	0	$1/128$	

Table 1. Prior mean and variance of DCM parameters.

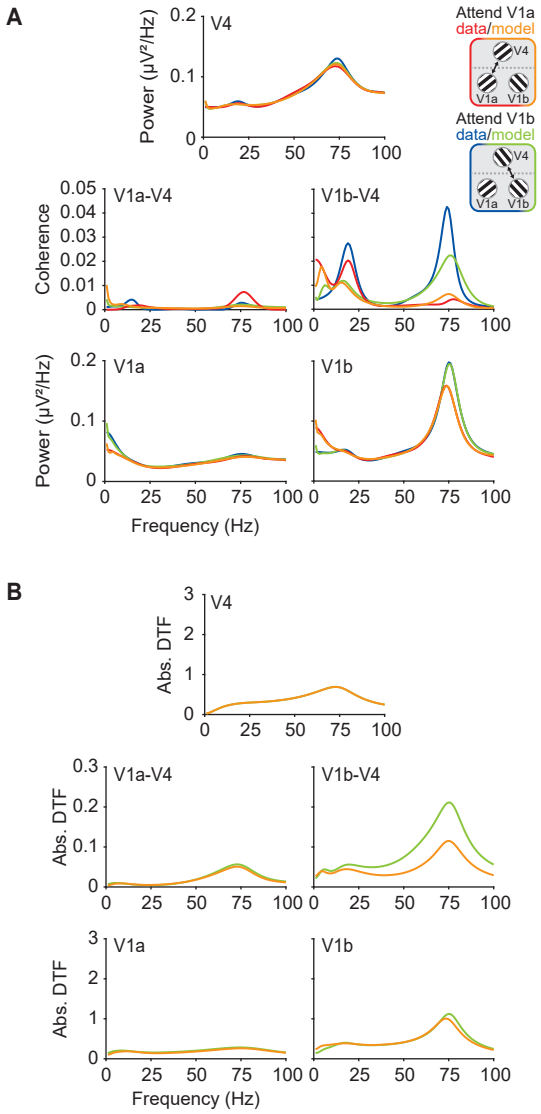


Supplementary Figure 1. Individual triplet DCMs of the attentional effects. Related to Figure 3.
Power spectral density of each node and coherence spectra between nodes for the 10 V1a-V1b-V4 triplets of monkey P (A) and monkey K (B). Each column represents one triplet.



Supplementary Figure 2. DTF of monkey K. Related to Figure 3.

Absolute DTF averaged across the 10 V1a-V1b-V4 DCMs of monkey K.



Supplementary Figure 3. DCM spectra and DTF of monkey K in the no-feedback architecture. Related to Figure 5.

(A,B) Results for DCM without V4-to-V1 feedback connections, for monkey K.
(A) Power spectral density and coherence spectra averaged across the 10 V1a-V1b-V4 triplets of monkey K.
(B) Absolute DTF averaged across the 10 V1a-V1b-V4 DCMs of monkey K.