

Behavioral/Cognitive

Attention Improves Transfer of Motion Information between V1 and MT

Sameer Saproo^{1,2,3} and John T. Serences^{3,4}¹Department of Psychology and ²Center for Neural Science, New York University, New York, New York 10003, and ³Department of Psychology, and⁴Neuroscience Graduate Program, University of California, San Diego 92093

Selective attention modulates activity within individual visual areas; however, the role of attention in mediating the transfer of information between areas is not well understood. Here, we used fMRI to assess attention-related changes in coupled BOLD activation in two key areas of human visual cortex that are involved in motion processing: V1 and MT. To examine attention-related changes in cross-area coupling, multivoxel patterns in each visual area were decomposed to estimate the trial-by-trial response amplitude in a set of direction-selective “channels.” In both V1 and MT, BOLD responses increase in direction-selective channels tuned to the attended direction of motion and decrease in channels tuned away from the attended direction. Furthermore, the modulation of cross-area correlations between similarly tuned populations is inversely related to the modulation of their mean responses, an observation that can be explained via a feedforward motion computation in MT and a modulation of local noise correlations in V1. More importantly, these modulations accompany an increase in the cross-area mutual information between direction-selective response patterns in V1 and MT, suggesting that attention improves the transfer of sensory information between cortical areas that cooperate to support perception. Finally, our model suggests that divisive normalization of neural activity in V1 before its integration by MT is critical to cross-area information coupling, both in terms of cross-area correlation as well as cross-area mutual information.

Key words: correlations; fMRI; gain; information theory; MT

Introduction

Sensory systems have a limited information processing capacity. To optimize behavior given this limited capacity, a subset of incoming sensory information is preferentially processed based on its behavioral relevance, a phenomenon known as selective attention. Most theories of attention focus on neural modulations within a single visual area, such as V1, V4, or MT. For instance, allocating attention to a relevant spatial location or visual feature can increase the mean response of visual neurons (McAdams and Maunsell, 1999; Martinez-Trujillo and Treue, 2004), decrease the overall variability of neural responses (Mitchell et al., 2007; Bressler and Silver, 2010), and modulate the structure of correlated activity across neural populations (Cohen and Maunsell, 2009; Mitchell et al., 2009; Cohen and Maunsell, 2011). Collectively, these modulations are thought to improve the signal-to-noise ratio of population responses in early visual areas, thereby improving representations of attended objects.

The selective improvement in signal-to-noise ratio within a visual area is important for efficient perception; but, selective

attention might also play a role in facilitating the transfer of information (i.e., communication) between functionally specialized regions during information processing. For instance, neurons in MT are known to pool inputs from neurons in V1 to gain complex direction-selective receptive fields (Movshon and Newsome, 1996; Simoncelli and Heeger, 1998; Rust et al., 2006; Beck and Neumann, 2011; Nishimoto and Gallant, 2011). In addition, it has been shown that feedback signals from MT to V1 (Pascual-Leone and Walsh, 2001; Silvanto et al., 2005) are important for the conscious perception of motion. Thus, the quality of the neural representation in V1, as well as the accurate transfer of this information to MT, appears to be critical for motion perception.

From an information-theoretic perspective, the transfer of information between distinct neural populations, such as between V1 and MT, occurs across a noisy conduit that is influenced by the stochastic nature of synaptic signaling (Conti and Wanke, 1975; Schneidman et al., 1998; White et al., 2000; Manwani and Koch, 2001; Vinje and Gallant, 2002). Furthermore, there are limits on the number of bits of information that can be transferred given a fixed expenditure of energy, and any loss of information during cortical information processing in early areas cannot be recovered via additional processing in later areas (the “data processing inequality”; Cover and Thomas, 1991). Given these constraints, the brain should ideally preserve sensory information related to behaviorally relevant stimuli at the expense of information related to irrelevant stimuli as signals are relayed from one neural population to the next. We tested this hypothesis using fMRI to record the BOLD signal simultaneously from areas

Received Aug. 13, 2013; revised Jan. 20, 2014; accepted Jan. 25, 2014.

Author contributions: S.S. and J.T.S. designed research; S.S. performed research; S.S. analyzed data; S.S. and J.T.S. wrote the paper.

This work was supported by National Institute of Mental Health R01-092345 to J.T.S. Portions of this work were published in dissertation form by S.S.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Sameer Saproo, 6 Washington Place, New York, NY 10003. E-mail: saproo@nyu.edu.

DOI:10.1523/JNEUROSCI.3484-13.2014

Copyright © 2014 the authors 0270-6474/14/343586-11\$15.00/0

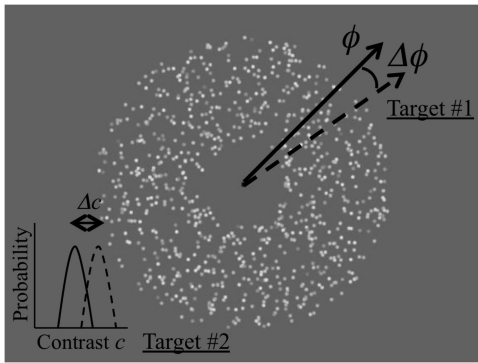


Figure 1. A static snapshot of the stimulus during a trial. On each trial, all dots (with a certain mean contrast c but with some variance in the individual contrast) moved coherently at a constant speed in the direction $\phi \in (0^\circ, 40^\circ, 80^\circ, \dots, 320^\circ)$. On “attend motion” trials, subjects had to detect a brief and transient change in the direction of motion, given by $\Delta\phi$, which lasted 750 ms. On “ignore motion” trials, subjects had to detect a brief and transient change in the mean contrast (Δc) of the display, which also lasted 750 ms. In each trial, both contrast and motion targets were present (i.e., the display briefly changed direction as well as briefly changed mean contrast at different times). However, subjects were instructed to respond to only one target and ignore the other depending on the instructions provided at the beginning of a block of trials. $\Delta\phi$ and Δc were titrated across blocks to approximately match task difficulty for the two conditions.

V1 and MT while subjects attended to moving visual stimuli. We analyzed the impact of attention on the activation of motion-selective responses within each area, as well as its impact on cross-area information transfer. The results were then evaluated through simulations using a linear–nonlinear cascade model of motion computation (Rust et al., 2006) constrained by empirical observations (Cohen and Maunsell, 2011).

Materials and Methods

Subjects. Subjects were 12 neurologically healthy human subjects (5 females) between the ages of 18 and 40 years. All subjects gave written informed consent in accordance with the Institutional Review Board at the University of California, San Diego. Each subject trained on the experimental task for at least one hour before the actual fMRI scanning session. Subjects were compensated at the rate of \$20/h for their time.

Stimulus and task. The stimulus consisted of a circular annulus of grayscale dots (outer radius = 9° , inner radius = 2.5° , and dot size = 0.5° of visual angle) moving on a middle gray background (Fig. 1), also known as a random dot kinetogram. During each 5 s trial, the dots moved in one of 9 possible direction ($0^\circ, 40^\circ, 80^\circ, \dots, 320^\circ$) with 100% coherence and constant speed ($6^\circ/s$). The contrast of each dot was independently sampled from a normal distribution with a mean and variance of $0.41c$ and $0.1c$, respectively (where c is the maximum contrast achievable on the video display with middle gray background). Each dot in the display had a limited lifetime of 100 ms, after which it regenerated at a random location. We kept the average dot density of the display constant at 5 dots/deg². The interval between successive trials was set to 2 s.

There were two types of trials: one where subjects were instructed to respond to a transient change (lasting 0.75 s) in the mean contrast (mean = $0.12 \pm 0.02c$ SEM, across subjects) of moving dots (attend contrast, or ignore motion), and the other where subjects had to respond to a transient change in the direction (mean = $16.5 \pm 0.33^\circ$ SEM, across subjects) of moving dots (attend motion). The timing of each target onset was uniformly randomized for each trial, with the constraint that a target could not occur in the first 1.5 s or the last 1 s of a trial. Each trial contained both targets; however, subjects were to only respond to the relevant target and to ignore the other based on the instructions given at the start of the scan (where a scan was a block of 36 trials, with 4 trials for each direction of motion). Importantly, because the stimulus display was physically identical between attend motion and ignore motion trials, any difference in the BOLD activation between the two conditions can be

attributed to selective attention to the direction of motion. There were 4 scans for each experimental condition (attend motion and ignore motion). The numerical value of the transient change in mean contrast or direction of motion was titrated on a scan-by-scan basis such that task difficulty was approximately matched (behavioral performance, $d' = 0.95 \pm 0.21$ SEM [ignore motion], and $d' = 0.81 \pm 0.20$ SEM [attend motion], paired t test: $t_{(9)} = 1.13, p = 0.29$). Here, a hit is defined as a correct response to the instructed target within 1 s of target onset and a false alarm is defined as a button-press made outside the 1 s response window for the instructed target. We excluded 2 subjects from the computation of behavioral performance, as they inverted the mapping between target stimuli and the appropriate behavioral response. However, even if these two subjects are included, the difference in task difficulty between conditions is still nonsignificant ($p = 0.16$).

Data acquisition and postprocessing. MRI was carried out on a GE MR750 3-Tesla scanner equipped with an 8-channel head coil at the Center for Functional Magnetic Resonance Imaging, University of California, San Diego. Anatomical images were acquired using a SPGR T1-weighted sequence, which yielded images with a 1 mm^3 resolution (TR/TE = 11/3.3 ms, TI = 1100 ms, 150 slices, flip angle = 18°). Functional images were acquired using a 2D gradient EPI pulse sequence, which covered visual cortex using 17 oblique transverse slices. Slices were acquired in ascending interleaved order with 3 mm thickness and 0 mm gap (TR/TE = 1000/30 ms, flip angle = 90° , image matrix = $64 \text{ (AP)} \times 64 \text{ (RL)}$, with FOV = $192 \text{ mm (AP)} \times 192 \text{ mm (RL)}$, voxel size = $3 \text{ mm} \times 3 \text{ mm}$).

Data analysis was performed using BrainVoyager QX (version 1.86; Brain Innovation) and custom time-series analysis routines written in MATLAB (version 10.1; MathWorks). All EPI images were slice-time corrected, motion-corrected (both within and between scans), and high-pass filtered (3 cycles/scan) to remove low-frequency temporal components from the time-series.

Independent region of interest (ROI) localizer scans. Two independent ROI localizer scans were run for each subject to identify voxels in V1 and MT that responded selectively to coherent motion within the spatial position occupied by the stimulus used in the main experiment. The spatial setup of the localizer paradigm was similar to that used in main experimental scans, with the following differences: (1) Rather than having one annulus of moving dots, there were two annuli spanning different eccentricities (outer and inner) that together covered the same spatial region as the annulus used in the main experimental scans. On each trial, either the inner annulus or the outer annulus was displayed in a pseudo-random order. This was done to identify voxels responsive to the outermost and innermost aspects of the stimulus that was used in the main experiment; however, in the final analysis, we collapsed across these two stimulation conditions because the results did not vary significantly as a function of eccentricity. (2) On each 10 s trial, either all of the dots moved coherently in one direction, or the position of each dot was randomly replotted at 60 Hz to produce a “snow” stimulus. Trials were separated by 10 s of blank fixation. The size of dots was 0.3° of visual angle, the speed of coherently translating dots was $5^\circ/s$, and the average dot density was 6 dots/deg². (3) The subject’s task was to respond either to a change in the speed of the moving dots or to a change in the contrast of the “snow” dots, respectively. (4) There were three targets on each trial, and subjects were asked to respond to each. Each target lasted 0.5 s, and the minimum separation between successive targets was 1.5 s.

To identify motion-selective voxels that responded to the retinotopic position of the stimulus aperture, data from the functional localizer scans were analyzed using a general linear model that contained two regressors: one marking motion trials and the other marking snow trials (as mentioned above, trials with inner and outer apertures were collapsed in this analysis). These “boxcar” models of the stimulus time-series were then convolved with the standard hemodynamic response function (HRF) template in BrainVoyagerQX to produce the regressors that were entered into the GLM. Voxels that responded more to epochs of continuous motion compared with epochs of snow were retained for further analysis if they passed a statistical threshold ($p < 0.01$), corrected for multiple comparisons using the false discovery rate algorithm implemented in BrainVoyager QX.

Retinotopic mapping procedures. Retinotopic mapping data were obtained in one to two scans per subject, using a checkerboard stimulus that rotated around a central fixation spot and standard presentation parameters (stimulus flickering at 8 Hz and subtending 60° of polar angle; Engel et al., 1994; Sereno et al., 1995; Wandell et al., 2007). To aid in the visualization of early visual cortical areas, we projected the retinotopic mapping data onto a computationally inflated representation of each subject's gray/white matter boundary. Visual areas were then defined manually according to visual markers of hemi-field and quarter-field reversals. This procedure was used to identify primary visual cortex (V1).

Computation of univariate responses within each ROI. The HRF for each condition was estimated using a finite impulse response (FIR) model-based deconvolution of the z-scored BOLD time series averaged across all voxels in each ROI (Dale, 1999). This model was used to estimate the time course of the response across an interval spanning 0–20 s after stimulus onset (for results, see Figure 2). The y-axis reflects z-scores, with the zero-point reflecting the activation level on “null-trials” in which no stimulus was presented.

Estimating direction-selective responses using a forward model. We assume that BOLD responses from 3 mm × 3 mm × 3 mm voxels in early visual areas, such as V1 or MT, reflect an approximately linear mixture of responses from many subpopulations of neurons with different degrees of selectivity to different directions of motion (Boynton et al., 1996; Heeger et al., 2000; Logothetis and Wandell, 2004; Kahn et al., 2011). To estimate the magnitude of responses in direction-selective neural populations in V1 and MT, we decomposed voxel responses into a set of hypothetical direction-selective channel responses using the method described by Brouwer and Heeger (2009, 2011).

The BOLD response of each voxel for each of the nine motion directions was modeled as a linear sum of weighted responses of nine different direction-selective channels, with the direction selectivity of the channels linearly spaced between 0° and 360° degrees (0°, 40°, 80°, ..., 320°). The tuning profile of each channel was modeled using either a sinusoidal function raised to the fourth power (see Fig. 3) or a Kronecker δ function (see Fig. 4A). The \sin^4 is a steerable filter (Freeman and Adelson, 1991), whereas a δ function forms an orthonormal basis set, the latter being important when computing cross-area correlations to ensure that correlations between channels are not artificially inflated by overlapping basis functions.

The computation then proceeded in two stages. In the first stage, the BOLD time-series for each voxel was shifted by 4 s to account for the hemodynamic lag, and then the BOLD response from 5 time points (TRs) after the onset of each trial was extracted (corresponding to the 5 s stimulus presentation on each trial). The data were then divided into two sets: a “training set” consisting of data from all scans except for two (one attend motion and one ignore motion scan). These remaining two scans, one from each attention condition, were held out to form a “test set.” One scan of each type was held out during model training and testing to ensure a balanced number of trials from each condition in the training set. Thus, the training set consisted of 1080 samples (5 TRs per trial × 36 trials per scan × 6 scans), and the test set consisted of 360 samples (5 TRs × 36 trials × 2 scans). Next, the magnitude of the response, or the weight, of each voxel in response to each of the 9 motion directions was computed using a GLM that took the standard form of the following:

$$B_1 = WC_1 \quad (1)$$

where B_1 is a matrix of voxel responses across all trials in the training set, C_1 is a design matrix comprised of channel responses for all trials, and W is a matrix that provides the estimated weight of the response in each hypothetical motion-selective channel separately in each voxel.

The ordinary least-squares estimate of W is computed as follows:

$$\hat{W} = B_1 C_1^T (C_1 C_1^T)^{-1}. \quad (2)$$

The computations in this first step (Eqs. 1 and 2) are analogous to a standard univariate GLM analysis where regressors, in this case hypothetical direction tuning curves, are fit to the response of each voxel.

In the second stage of the analysis, the pattern of activation across an entire visual area on a given trial in the test set B_2 was used to recover the estimated channel responses C_2 using the previously computed weights W .

$$C_2 = (\hat{W}^T \hat{W})^{-1} \hat{W}^T B_2 \quad (3)$$

Where the estimated channel responses (C_2) is a matrix of size 360 × 9 given the number of samples in the test set (360) and the number of direction-selective channels (9). The estimation of C_2 is a multivariate computation as the entire pattern of voxel responses across an ROI on a given trial is used, in combination with the independently estimated weights that reflect the selectivity of each voxel composition (computed in Eq. 2), to constrain the estimated response in each of the nine direction-selective channels. After the nine channel responses were computed on each trial for a given test set, the entire training/testing procedure was repeated until all unique combinations of two scans (always with one attend motion and one ignore motion scan) were held out in turn. Finally, the channel response profiles computed on each trial were circularly shifted such that the direction of the stimulus that evoked the response profile on each trial was set to 0° offset in the “offset from stimulus” axis used in Figures 3 and 4. As the channel response profiles now had a common center, they could be averaged across trials based on attention condition.

For the computation of channel response profiles in each attention condition (e.g., see Fig. 3), we averaged the channel responses across the 5 TRs estimated on each trial. For the computation of cross-area coupling (both the computation of correlations between channels and for overall mutual information [MI] between all channels, see Figs. 5 and 7), we used channel responses from all 5 TRs in each trial (producing a 720 element vector for each experimental condition, 4 scans × 36 trials/scan × 5 time points/trial).

MT contains fewer voxels than V1 and MT is more anatomically convoluted, therefore partial voluming may have a negative impact on our ability to obtain a clean motion signal. However, the forward model that we used in the present study is relatively robust to this potential problem as it does not require that each voxel have a unimodal (e.g., Gaussian) distribution of motion-selective neural populations. Instead, the method will work so long as the response evoked by each direction of motion is consistent across repeated presentations of the same stimulus (i.e., the voxels have a direction-selective response profile). Thus, even if voxels in MT have highly mixed responses to several different directions, we should still be sensitive to these selective responses as long as they are consistent, and we can exploit them to generate stable channel-tuning functions.

Multivariate MI. We used a multivariate measure of MI, also referred to as multidimensional MI, or MDMI (Chai et al., 2009; Wang et al., 2009), which computed total MI between all direction-selective channel responses in V1 and those in MT. A multivariate MI measure is often more powerful than a univariate MI measure, as MDMI accounts for any synergistic encoding of information in the joint activity of channels (Schneidman et al., 2003) to provide a more accurate picture of total cross-area information coupling. The equation for distance-based nonparametric MI (Wang et al., 2009) between two d -dimensional response vectors X and Y with n observations is as follows:

$$I(X, Y) = -\frac{1}{n} \sum_{i=1}^n \log \left(\frac{p_k(x_i) p_k(y_i)}{p_k(x_i, y_i)} \right) \quad (4)$$

$$p_k(x_i) = -\frac{k}{n-1} \left(\frac{\Gamma(d/2 + 1)}{\pi^{d/2}} \right) \left(\frac{1}{r_k(x_i)^d} \right) \quad (5)$$

Here, Γ is the gamma function, x_i is the i^{th} observation in X , and $r_k(x_i)$ is the distance between x_i and its k^{th} nearest neighbor (similarly for y_i). We set k equal to \sqrt{n} in our analysis because k should scale with the size of the data for accurate estimation of MI in distance-based nonparametric methods, with square-root scaling being a good heuristic (Mnatsakanov et al., 2008; Misra et al., 2010). In our case, the dimensionality d was equal to the number of channels (9), and n was equal to the number of channel

responses (5 TRs \times number of trials, or 720 total responses per attention condition). Because data points from successive TRs are auto-correlated, the resulting channel responses in both V1 and MT should also be auto-correlated. However, this should not degrade the accuracy of MI estimated between channel responses of V1 and MT but instead allows us to exploit information coupling between V1 and MT that occurs across time within a trial (which would not be possible if we had averaged across all TRs in a trial). Bias correction on the resulting MDMI estimates was done by recomputing MDMI 1000 times using pseudo-randomly shuffled data labels and then removing the mean shuffled MDMI value from the original MDMI value obtained using the unshuffled data. This procedure is commonly used in the computation of MI to account for bias in the absolute value of MI measurement resulting from a limited sample size (Chai et al., 2009).

Model simulation of cross-area correlation for MT motion computation. To investigate patterns of cross-area correlations between direction-selective channel responses, we set up 60 large neural pools; each pool was assumed to consist of more than 100 similarly tuned neurons (N) so that the mean and the variance of the pooled response could be reasonably approximated analytically (Abbott and Dayan, 1999). All neurons had a canonical Gaussian tuning function with bandwidth (full width half-maximum) of 90° ; however, each pool of neurons had a unique direction preference chosen from the set of 60 possible directions evenly spaced across 360° ($6^\circ, 12^\circ, 18^\circ, \dots, 360^\circ$). The variance of the neurons in each pool (σ^2) was set equal to their mean firing rate in response to a stimulus (Poisson distribution, although the following logic also holds so long as the variance scales with the mean firing rate). The correlation between neurons within each pool (c) was set to be inversely proportional to their mean firing rate (Cohen and Maunsell, 2011); for example, a firing rate of 40 Hz yielded a correlation coefficient (c) of 0.15, 60 Hz yielded $c = 0.1375$, and 20 Hz yielded $c = 0.1625$. The trial-by-trial variance of the combined response of a neural pool was approximated by $\sigma'^2 = c\sigma^2$, which is a good approximation for the averaged population response when $N \gg 1/c$ as is assumed in our simulation (Abbott and Dayan, 1999).

The covariance matrix for pooled responses was constructed according to the limited range correlation model (Snippe and Koenderink, 1992; Abbott and Dayan, 1999) because correlations between neurons in early visual areas have been shown to depend on the difference between their feature selectivity (Huang and Lisberger, 2009; Cohen and Kohn, 2011). In this model, covariance between the i^{th} and j^{th} neuronal pool is expressed as $Q_{ij} = \sqrt{(\sigma_i^2 \times \sigma_j^2)} \times \rho^{|i-j|}$ where $\rho = \exp(-D/L)$. Here, D is the distance between the tuning preference of neural pools ($\sim 6^\circ$, given 60 pool of neurons with tuning preferences that are evenly spaced across 360°), and L is the correlation length, which was set to 4 keeping $L < D$ (Snippe and Koenderink, 1992). Figure 5B illustrates the covariance matrix of V1 cells using this limited-range correlation model, when the direction of motion in the stimulus is 180° . Using the above expression for covariance and a mean neural response based on a Gaussian tuning function, the response of each V1 neural pool was computed for each stimulus presentation in the simulation using the MATLAB function *mvnrnd*.

The response of direction-selective MT neurons was computed as follows. First, for each stimulus direction ($6^\circ, 12^\circ, \dots, 360^\circ$), the response of each V1 direction-selective neural pool was computed as described above. Because our simulation averages tuning functions across neurons sharing a similar feature preference in each neural pool, with a large enough population ($N > 100$), any heterogeneity in the individual tuning functions would not substantially impact the results of the simulations. Next, the pooled responses from V1 were passed through the linear–nonlinear cascade MT motion computation model (Rust et al., 2006) to yield responses for MT cells tuned to different motion directions for all possible stimulus directions. The model was simulated using the following parameters: MT scaling nonlinearity = 1, MT exponent nonlinearity = 2, V1 untuned normalization factor = 0.1, V1 self-tuned normalization factor = 0.1 (all parameter values referenced from Rust et al., 2006). Figure 5A provides a schematic illustration of the cascade model of motion computation used in our simulation (for more details and equations, see Rust et al., 2006).

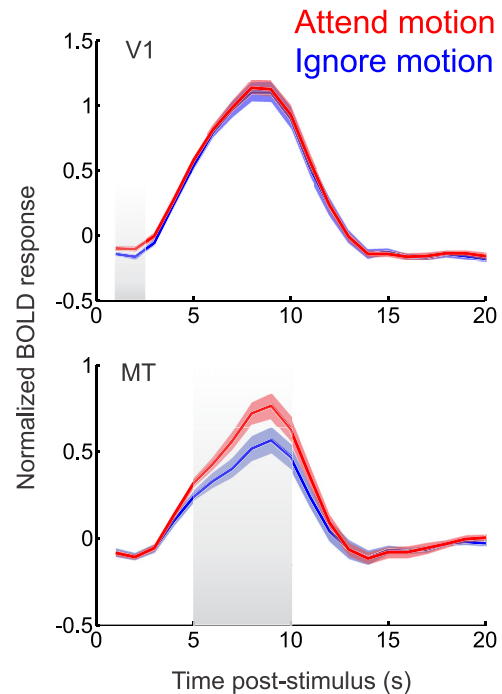


Figure 2. HRF computed from attend motion (red curve) and ignore motion (blue curve) trials in V1 and MT. The gray portion in each panel represents poststimulus time points that had a significant increase in the amplitude of the HRF with attention: 2 s for V1 and 6–10 s for MT ($p < 0.05$). The shaded region around the curves represents between-subject ± 1 SEM ($N = 12$).

The effect of attention on V1 cells was simulated by changing the baseline (β) and amplitude (α) of their tuning function; $R(\theta) = \beta + \alpha \times f(\theta; \kappa)$, where R is the response of a cell for a stimulus of direction θ , and f is a circular Gaussian function. The bandwidth (κ) of tuning was assumed to remain unchanged by attention (McAdams and Maunsell, 1999; Treue and Maunsell, 1999). The tuning parameters for the unattended condition (baseline = 20 Hz, amplitude = 40 Hz) and attended condition (baseline = 10 Hz, amplitude = 70 Hz) were chosen to reflect the proportional strength of attention modulation at the population level according to measured BOLD responses (see Figs. 3 and 4). We get qualitatively similar results from the simulation, albeit of lower magnitude, when using more conservative values for neural firing rates based on single-unit studies (e.g., from Martinez-Trujillo and Treue, 2004). Using the cascade model, we simulated V1 neural responses and the resulting MT neural responses to different motion directions across 100,000 trials. Then we computed the mean response profile of V1 and MT populations (see Fig. 5C,D) as well as the correlation between V1 responses and MT responses for neurons having similar tuning preferences (see Fig. 5E). Based on the assumption that the magnitude of the BOLD channel responses that we estimated is at least monotonically related to the summed output of each simulated neural population, the results of the simulation are qualitatively consistent with our empirical observations (compare Fig. 5C–E with Figs. 3 and 4).

Results

HRF modulation

First, we computed the average HRF for all independently identified voxels in each visual area. There was a relative increase in the amplitude of hemodynamic response after stimulus presentation in area MT when attention was directed to motion direction (paired t tests, all $t_{(11)} > 2.36$, all $p < 0.05$, 5–10 s after stimulus; Fig. 2). On the other hand, the attention modulation in area V1, as indexed by the mean HRF amplitude, was negligible. However, given that feature-based attention can increase the gain of neural populations tuned to a relevant stimulus while simultaneously attenuating the response of other populations (Martinez-Trujillo

and Treue, 2004), it is difficult to interpret a lack of an overall amplitude effect using the average HRF computed across all voxels (Serences and Saproo, 2012).

Channel response modulation

Attending to the direction of motion, as opposed to contrast, led to an increase in the mean response for channels tuned near the stimulus direction (0° , $\pm 40^\circ$ offset), whereas the responses of those channels that were tuned far away from the stimulus direction ($\pm 120^\circ$, $\pm 160^\circ$ offset) were attenuated (Fig. 3). This observation is in agreement with the previous reports of feature-based attention modulation based on the activity of single direction-selective neurons in MT (Martinez-Trujillo and Treue, 2004). However, we found that the attenuation of channels tuned away from the attended feature was more pronounced in V1 compared with MT. The larger attenuation in V1 compared with MT may explain the negligible modulation of V1 HRFs compared with MT HRFs (Fig. 2) as the enhancement and attenuation would have canceled when looking at the aggregated univariate response amplitude across all voxels.

V1-MT correlation

We next investigated the impact of attention on cross-area coupling by computing the correlation between response vectors of like-tuned channels in V1 and MT (e.g., between the channel tuned to 0° in V1 and the channel tuned to 0° in MT). The input vectors for this analysis consisted of estimated responses for each channel in V1 and MT at each TR across the 5 s window corresponding to the time that the stimulus was being viewed by the observer, concatenated across all trials. To ensure that the estimates of cross-area correlations in each channel were independent, we recomputed the forward model using Kronecker δ functions in place of Sine functions for the basis set. This produced a set of orthonormal basis functions that were independent, as opposed to the highly overlapping Sine functions that were not. Importantly, we observe a similar pattern of attention modulation of mean channel responses in V1 and MT using the δ functions (Fig. 4A). Figure 4B shows cross-area correlations as a function of the distance between the direction preference of each channel and the direction of the stimulus that evoked the response on a given trial. In this analysis, we collapsed across channels with corresponding negative and positive offsets from 0° to improve power because the mean response profile (Fig. 4A) was symmetric around the attended direction (i.e., around offset 0°). The data show that, when motion-direction is attended, cross-area correlations decrease for channels tuned to the direction of the stimulus, whereas cross-area correlations increase for channels tuned away from the direction of the stimulus (one-way repeated-measures ANOVA, $F_{(4,44)} = 2.78$, $p < 0.05$). A comparison of the data in Figure 4A, B reveals that attention modulation of cross-area correlation is inversely proportional to the modulation of the mean response for the corresponding channels:

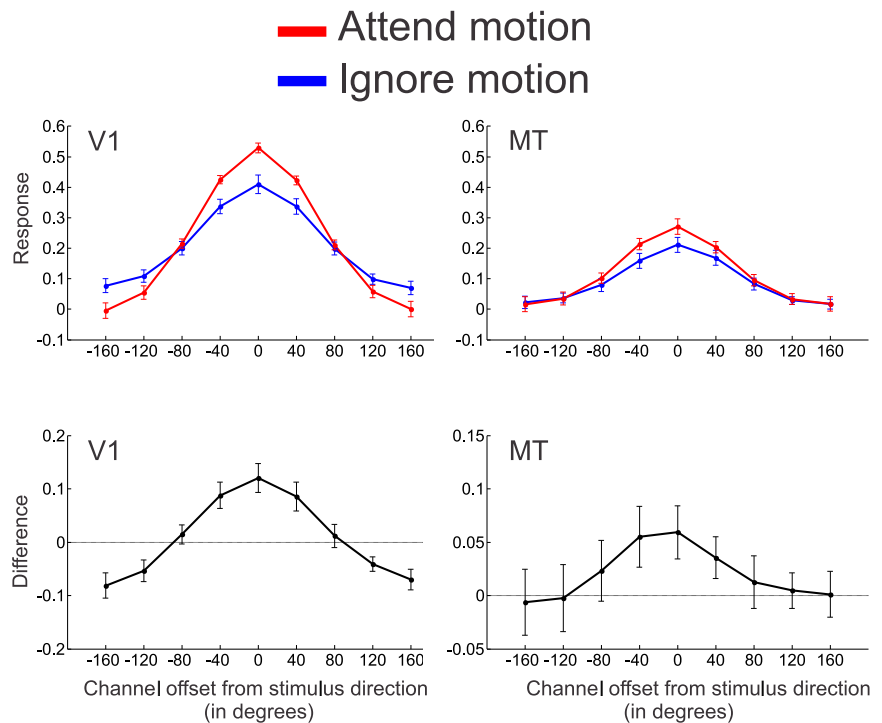


Figure 3. Normalized BOLD response in direction-selective channels in V1 and MT tuned to different direction offsets with respect to the stimulus direction (where stimulus direction is 0° on the x-axis). The blue curve indicates the mean channel response profile when attention was directed away from the stimulus direction (to contrast), and the red curve indicates the mean channel response profile when attention was directed toward the stimulus direction. The response profiles here were generated using a forward model with Sine⁴ basis functions (see Materials and Methods). Bottom, Difference between attend motion and ignore motion conditions, highlighting the effect of attention on the channel response profiles. Error bars indicate between-subject ± 1 SEM ($N = 12$).

cross-area correlations increase for channels whose mean response decreases with attention, and vice-versa.

Simulation results of V1-MT correlations

To better understand this pattern of cross-area correlations, we simulated MT activity using a linear-nonlinear cascade model (Simoncelli and Heeger, 1998; Rust et al., 2006), in which the response of a direction-selective cell in MT is modeled as a probabilistic nonlinear function of the sum of weighted responses from a population of V1 direction-selective cells (see Fig. 5A and Model simulation of cross-area correlation for MT motion computation in Materials and Methods). In addition, we constrained the model in three ways: First, we assumed limited-range correlations between direction-selective cells in V1, such that the strength of correlations between any two direction selective cells is a function of the difference in their direction selectivity (Snippe and Koenderink, 1992; Abbott and Dayan, 1999; Jazayeri and Movshon, 2006; Huang and Lisberger, 2009) (Fig. 5B). Second, we made variance in trial-by-trial responses of V1 cells scale proportionally to their mean response according to a Poisson noise model (Softky and Koch, 1993; White et al., 2000). Third, we constrained the model based on Cohen and Maunsell's (2011) observation that the magnitude of the mean response of a similarly-tuned neural population in area V4 is inversely proportional to the magnitude of pairwise correlations within that population, both with and without attention. We assumed a similar relationship for direction-selective neural populations in V1.

In the simulation, we modulated the mean responses of V1 direction-selective cells according to the pattern observed empirically.

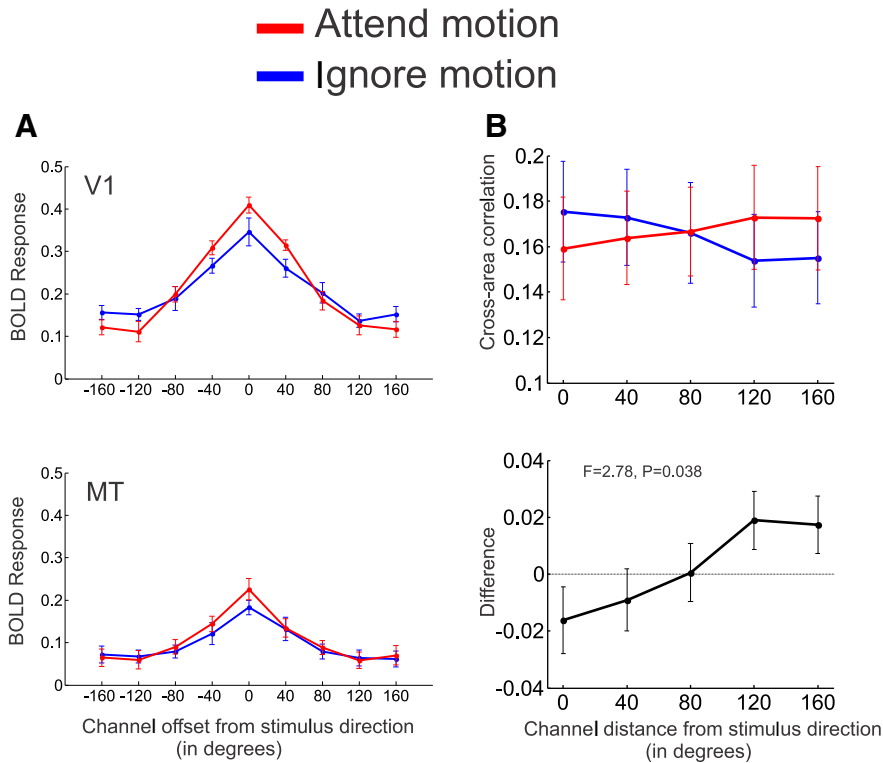


Figure 4. *A*, Mean channel response profiles, similar to Figure 3 but with Kronecker δ basis functions used in the forward model computation. *B*, Top, Cross-area correlations between similarly tuned channels in V1 and MT (with channel responses computed using δ basis functions, as in *A*). Bottom, Change in cross-area correlations with attention. The error bars are quite large in the top because of large differences across subjects. However, the SEM of the difference between the blue and red curves is much smaller, reflecting a consistent effect of attention across subjects (*B*, bottom). The *x*-axis indexes the distance of a channel from the stimulus direction; channels with the same negative and positive offset were averaged to increase statistical power. Error bars indicate between-subject ± 1 SEM ($N = 12$).

ically with attention (Fig. 5C) and derived the response of MT direction selective cells from the model. We simulated a large number of trials with and without attention to the direction of motion. The simulation produced a pattern of attention modulation in MT that is qualitatively similar to the pattern observed empirically (Fig. 5D, compare with Figs. 3 and 4A). Furthermore, the pattern of cross-area correlations produced by the simulation (Fig. 5E) is also qualitatively similar to the empirically observed pattern (Fig. 4B). Attention produced a decrease in the cross-area correlation between direction-selective populations tuned to the attended feature, whereas cross-area correlations increased for populations tuned away from the attended feature. Importantly, we found that the link between cross-area correlation and mean response modulation was dependent on the divisive normalization of the V1 responses before the responses were integrated in MT. Figure 6 shows the cross-area correlation and the mean response of cells when divisive normalization was not performed during model simulation; although the mean response modulation with attention remains qualitatively similar, there is no modulation of cross-area correlation.

Analysis of simulation results

According to the cascade model of MT motion computation (Simoncelli and Heeger, 1998; Rust et al., 2006), the response of an MT cell is modeled as the sum of weighted responses from the entire V1 population as follows: $M_p = \sum_i w_i V_i$, where M_p is the response of a MT cell tuned to direction p , V_i is the response of a V1 cell tuned to direction i , and w_i is the weight for V_i . The

correlation between any two MT and V1 neurons tuned to the same direction p is given by the following:

$$\text{Corr}(M_p, V_p) = \frac{\text{Cov}(M_p, V_p)}{\sigma_{M_p}^2 \sigma_{V_p}^2} \quad (6)$$

where $\text{Corr}(M_p, V_p)$ is the covariance between the cells, $\sigma_{M_p}^2$ is the variance of the MT cell, and $\sigma_{V_p}^2$ is the variance of the V1 cell. The terms can be expanded to the following:

$$\text{Corr}(M_p, V_p) = \frac{\text{Cov}(\sum_i w_i V_i, V_p)}{\sigma_{\sum_i w_i V_i}^2 \sigma_{V_p}^2} = \frac{\sum_i w_i \text{Cov}(V_i, V_p)}{(\sum_i w_i^2 \sigma_{V_i}^2 + 2 \sum_{i,j} w_i w_j \text{Cov}(V_i, V_j)) \sigma_{V_p}^2} \quad (7)$$

Then, the correlation between a V1 and MT cell can be completely expressed in terms of local correlations between V1 cells as follows:

$$\text{Corr}(M_p, V_p) = \frac{\sum_i w_i \sigma_{V_i}^2 \text{Corr}(V_i, V_p)}{\sum_i w_i^2 \sigma_{V_i}^2 + 2 \sum_{i,j} w_i w_j \sigma_{V_i}^2 \sigma_{V_j}^2 \text{Corr}(V_i, V_j)} \quad (8)$$

According to Equation 7, cross-area correlations can be affected by three quantities: the weights w_i , the covariance matrix for V1 cells $\text{Cov}(V_i, V_j)$, and the variance of V1 cells $\sigma_{V_i}^2$. In the model simulation, we first assumed that the linear weights w_i are not affected by short-term changes in attention state because they might reflect synaptic strength and therefore are probably not strongly modulated on a short trial-to-trial time scale. Second, given the observed shape of attention modulation across V1 channels in our data, the number of neural populations that undergo an increase in mean response is approximately equal to the number of neural populations that show a decrease in response, suggesting that the term $\sum_i w_i^2 \sigma_{V_i}^2$ (sum of weighted variances) might remain largely unchanged with attention, assuming Poisson response statistics. In addition, this term is smaller in magnitude compared with the other components, further reducing its impact on attention related changes in correlations between neural populations in V1 and MT.

If we assume that the covariance matrix for V1 cells is not significantly affected by the modulation of population responses with attention, then the main factor that affects cross-area correlations in Equation 7 is the variance of V1 cells. Specifically, it can be hypothesized that $\text{Corr}(M_p, V_p) \propto \frac{1}{\sigma_{V_p}^2}$, (i.e., the change in correlation between similarly tuned V1 and MT cells is inversely proportional to a change in the variance of the V1 cells). Because the variance of neural responses is largely a function of mean firing rate, the implication of this result is that the change in cross-area correlation we observe could simply be a consequence of the observed gain modulation of V1 cells. This notion is also

generally consistent with the observation by Cohen and Maunsell (2011), who found an approximately inverse relationship between neural correlations and mean response, but in a single visual area V4.

One of the candidate mechanisms for allowing covariance between V1 cells to remain largely unchanged despite attention-induced gain in response could be divisive normalization (Heeger, 1992; Reynolds and Heeger, 2009) because its absence in the model simulation significantly affected the pattern of cross-area correlations. We provide an intuitive justification of how divisive normalization could have this effect: from Equation 7, the correlation between an MT and V1 cell is proportional to the sum of the weighted covariance between a V1 cell V_p and all other V1 cells ($V_1 \dots V_n$):

$$Corr(M_p, V_p) \propto \{w_1 Cov(V_1, V_p) + w_2 Cov(V_2, V_p) + \dots + w_n Cov(V_n, V_p)\} \quad (9)$$

The covariance of two V1 cells with responses, N_1 and N_2 , that undergo divisive normalization can be expressed as follows:

$$Cov\left(\frac{N_1}{N_1 + N_2}, \frac{N_2}{N_1 + N_2}\right)$$

Consider the impact of attentional modulation on the covariance between these cells in two simple cases: (1) where the two cells have significantly overlapping tuning functions; and (2) where they do not.

In the case where V1 cells significantly share tuning preference, an attention-induced gain modifier α is applied to the response of both cells, which leads to no change in covariance compared with covariance when no attention-induced gain is applied:

$$Cov\left(\frac{\alpha N_1}{\alpha N_1 + \alpha N_2}, \frac{\alpha N_2}{\alpha N_1 + \alpha N_2}\right) = Cov\left(\frac{N_1}{N_1 + N_2}, \frac{N_2}{N_1 + N_2}\right)$$

In the case where V1 cells significantly differ in their tuning preference, attentional gain is applied only to the cell that is tuned to the stimulus direction. Here, the covariance between cells can be expressed as follows:

$$Cov\left(\frac{\alpha N_1}{\alpha N_1 + N_2}, \frac{N_2}{\alpha N_1 + N_2}\right)$$

However, because the other V1 cell is tuned away from the stimulus feature, it has a much lower response than the cell tuned to stimulus feature, such that $N_1 \gg N_2$. Therefore, the covariance between such cells is negligible with or without attention gain:

$$Cov\left(\frac{\alpha N_1}{\alpha N_1 + N_2}, \frac{N_2}{\alpha N_1 + N_2}\right) = \frac{1}{\alpha} Cov\left(1, \frac{N_2}{N_1}\right) \sim 0$$

This would hold even if we assume that a negative gain ($-\alpha$) is applied to N_2 (Cohen and Maunsell, 2011).

This analysis suggests that attentional modulation would have little impact on the covariance between cells that have highly similar or highly dissimilar tuning, largely resulting from divisive normalization. However, one could argue that attention impacts covariance for intermediate cases in which the cells have partially overlapping tuning functions. According to the cascade model (Rust et al., 2006), the weights assigned to the covariance in Equation 9 are a function of the distance between the tuning of V1 cell V_p and other cells, ranging from a maximum positive value to a maximum negative value, passing through zero. Therefore, ac-

ording to the model, the weight assigned to the covariance of two V1 neurons with partially overlapping tuning functions would be zero (or near zero). This implies that, even if there were a small attention-related change in covariance between V_p and a cell that has a partially overlapping tuning function, it would not contribute to the overall sum of weighted V1 covariances in Equation 9 (or Eq. 7) and therefore would not lead to a modulation of cross-area correlations between MT and V1.

In sum, this analysis demonstrates that divisive normalization is a candidate mechanism that can lead to the empirically observed attention-related modulation of cross-area correlations between V1 and MT. Furthermore, the analysis suggests that the modulation of within or cross-area correlation with attention might arise from attention-driven gain modulations.

Attentional modulation of MI between V1 and MT

We next investigated whether the observed attention-induced modulations of amplitude and cross-area correlation influenced the overall level of information transfer between V1 and MT. Based on Shannon’s channel coding theorem (Latham and Roudi, 2009), if we assume that the conduit of information between V1 and MT forms a noisy channel of unspecified noise characteristics, then the efficacy of neural communication is proportional to the MI between direction-selective channel responses in V1 and MT (Fig. 7A,B). This communication model is consistent with the known characteristics of V1-MT interactions (Simoncelli and Heeger, 1998; Rust et al., 2006) as well as the dense anatomical projections from V1 to MT (Born and Bradley, 2005). To assess the effect of attention on communication efficacy, we used a multivariate metric that captured the MI between all direction-selective channels of V1 and MT on attend motion and ignore motion trials (Chai et al., 2009). This metric accounts for any information embedded in the correlation structure of the channel responses (Brenner et al., 2000; Schneidman et al., 2003), thereby providing a more complete picture of the overall coupling between V1 and MT. When attention is directed to the direction of motion, this metric revealed that MI between V1 direction-selective channels and MT direction-selective channels is significantly higher compared with when motion is ignored ($t_{(11)} = 3.25, p < 0.01$; Fig. 7C). We also find a significant attention-related increase in MI between V1 and MT in the simulation of the cascade model of motion computation when we use exactly the same parameters for the simulation that we described earlier. However, we found that removing divisive normalization from the model simulation, which attenuates changes in cross-area correlation while sparing attentional gain (Fig. 6), reduced cross-area MI between simulated V1 and MT responses by a factor of 4.

Discussion

Here, we investigated the role of selective attention in modulating communication between two visual areas that play a role in processing motion. Mimicking patterns observed using single-unit recording techniques (Martinez-Trujillo and Treue, 2004), we observed an increase in the mean response of direction-selective channels that are tuned to attended stimuli and a decrease in channels tuned away from attended stimuli. However, cross-area correlations were inversely related to the pattern of mean response modulation (i.e., populations tuned to attended stimuli in V1 and MT were less correlated with each other, whereas populations tuned away were more correlated). This result can be explained using a linear–nonlinear cascade model of motion computation (Simoncelli and Heeger, 1998; Rust et al., 2006) that

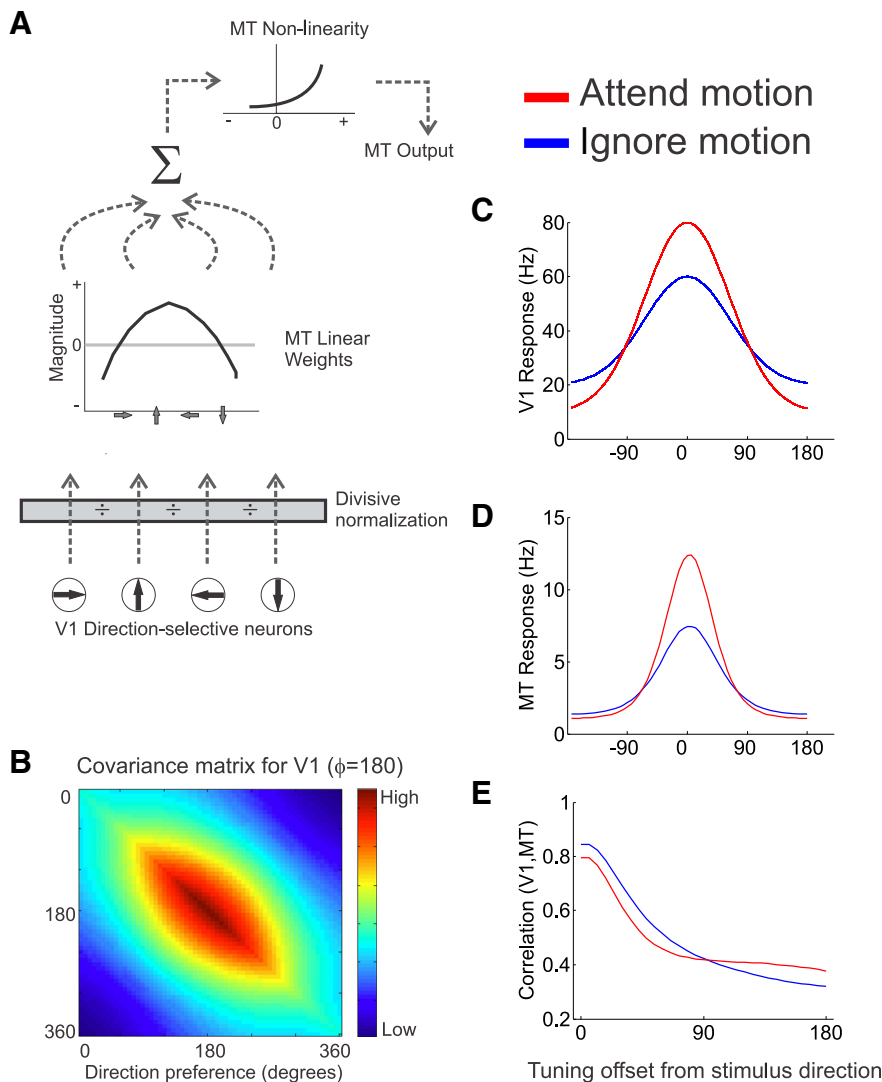


Figure 5. *A*, The cascade model of MT motion computation (from Rust et al., 2006; see Materials and Methods for details). Direction-selective cells in V1 undergo divisive normalization consisting of tuned and untuned components. Next, the response of each V1 cell is weighted using a linear weighting scheme, and then all weighted responses are summed. Finally, passing the summed responses from V1 through a nonlinear function generates the response of direction-selective cells in MT. *B*, An example covariance matrix for V1 cells, based on limited-range correlations. The covariance matrix shown here was generated for a motion direction of 180°. *C–E*, Results of simulating the cascade model to assess the effect of attention on cross-area correlations. *C*, Mean population response profile across V1 direction-selective cells, with and without attention, set by the experimenters according to the observed pattern of modulation in the BOLD data (Fig. 3). *D*, Mean population response profile for MT direction-selective cells, with and without attention, produced by the cascade model. *E*, The simulated cross-area correlation profiles between direction-selective cells in V1 and MT; cross-area correlations decrease for populations that undergo an increase in mean response with attention, whereas the opposite effect is observed for populations that undergo a decrease in mean response with attention. The pattern of simulation results is qualitatively similar to the pattern observed with BOLD data in our experiment.

is constrained by recent empirical observations about the influence of attention on pairwise noise correlations within an area (Cohen and Maunsell, 2011). The data also reveal that the attention-related modulation of response amplitude and cross-area correlation accompany an overall increase in bidirectional information coupling between V1 and MT, as assessed with multivariate MI (Chai et al., 2009). In turn, improving the efficacy of communication between cortical nodes should improve the quality of internal perceptual representations and enhance the speed and accuracy of perceptual decisions. Finally, simulations suggest that attention-mediated changes in cross-area information coupling are dependent on divisive normalization, because removing normalization effectively eliminated the attentional

modulation of cross-area correlations (while sparing the gain modulation) and led to a large reduction in the modulation of cross-area MI.

The larger attention modulation of mean HRFs in MT compared with V1 (Fig. 2) could have multiple plausible explanations. For instance, it could be a consequence of higher direction selectivity in MT, or could be a consequence of pooling several smaller attention modulations from different neural pools in V1 to produce a more pronounced modulation in MT. However, because the HRF is a measure of the average response from all neural subpopulations in an ROI, the observed disparity could also reflect different patterns of attentional modulation playing out in each region. For instance, there are almost an equal number of channels that undergo relative enhancement and suppression in V1 (Fig. 3), which would lead to little effect on the mean HRF. However, the magnitude of suppression in MT is relatively small compared with the magnitude of enhancement, which may contribute to a larger HRF modulation with attention.

The prediction that attention leads to an increase in cross-area information coupling was initially motivated by a previous observation that attention increased the gain of population response profiles in V1 (Sapuro and Serences, 2010). In turn, increases in the gain, or the dynamic range, of responses in V1 increased the MI between responses and stimulus features. One possible consequence of increased MI in V1 is the improved transfer of sensory information to downstream areas, such as MT, as the quality of the code in V1 sets an upper limit on the quality of the cross-area communication. Thus, we might have expected an increase in cross-area MI just on the basis of attention-related changes in gain within V1.

However, our simulations suggest that this issue is more nuanced. For instance, removing divisive normalization from the cascade model of motion computation

(Rust et al., 2006) does not impact overall gain but it does eliminate attention-related changes in cross-area correlations and it dramatically reduces cross-area MI. Thus, the model simulation suggests that attention-related increases in cross-area MI are only observed in the presence of changes in both V1/MT gain and changes in cross-area correlations. However, the apparent dependence of cross-area MI on both factors does not necessarily imply that attention independently modulates neural gain and neural correlations; there may not be separate “volume knobs” for gain and correlation. Instead, even though both types of modulation appear to be important for cross-area coupling, one modulation (in correlation) may be an automatic consequence of the other modulation (in gain) due to divisive normalization. This

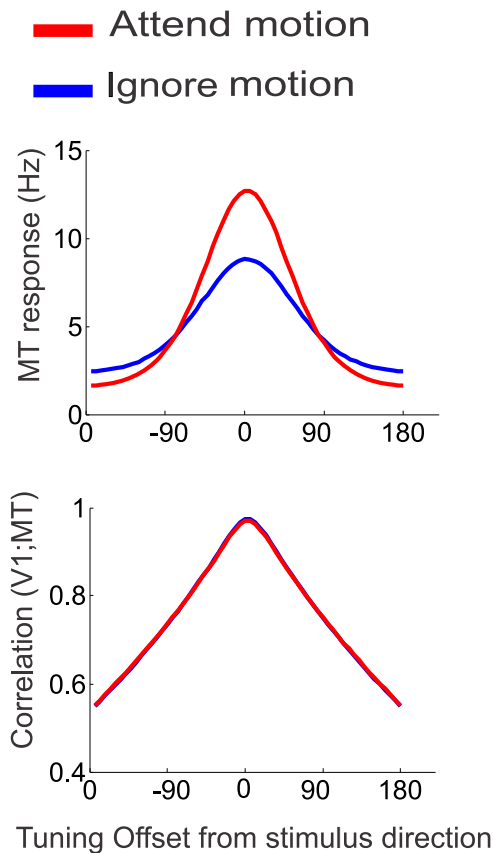


Figure 6. Results produced by the cascade model of MT motion computation in the absence of divisive normalization of responses in V1 cells. Top, Mean response of direction-selective cells in MT, with and without attention. Bottom, Cross-area correlation profiles between V1 and MT direction-selective cells, with and without attention. Without normalization, the mean response profile changes in a manner that is similar to the experimentally observed pattern (compare with Figs. 3 and 4A). However, cross-area correlations remain unchanged as a function of attention (unlike empirical observations; compare with Fig. 4B).

notion is consistent with the results of our simulations, as Equation 8 suggests that cross-area correlations between V1 and MT can be represented based largely on correlations between V1 cells. In turn, if we assume that there is an inverse relationship between gain and correlations in V1 (i.e., Cohen and Maunsell, 2011), then a gain modulation in V1 will impact correlations in V1, which in turn will impact cross-area correlations between V1 and MT (Eq. 8). Thus, in the presence of divisive normalization, this model suggests that gain changes may naturally modulate correlations between similarly tuned populations of V1 and MT cells. The relationship between gain and cross-area correlations, and their joint relationship with cross-area MI, is thus a subtle but important issue to address in future research, and understanding this relationship will likely require the use of spiking network models and converging empirical methodologies that have better temporal and spatial resolution.

The functional impact of changes in the MI between cortical areas (Fig. 7C) can be framed in terms of Shannon's channel coding theorem, which holds that, for a noisy communication channel, there is an upper bound on the efficacy of communication using any encoding scheme, termed channel capacity. Increases in MI between the signals sent and those received over a channel reflect movement toward this hypothetical upper bound. In the present context, the transfer of information between V1 and MT can be viewed as a communication system with two

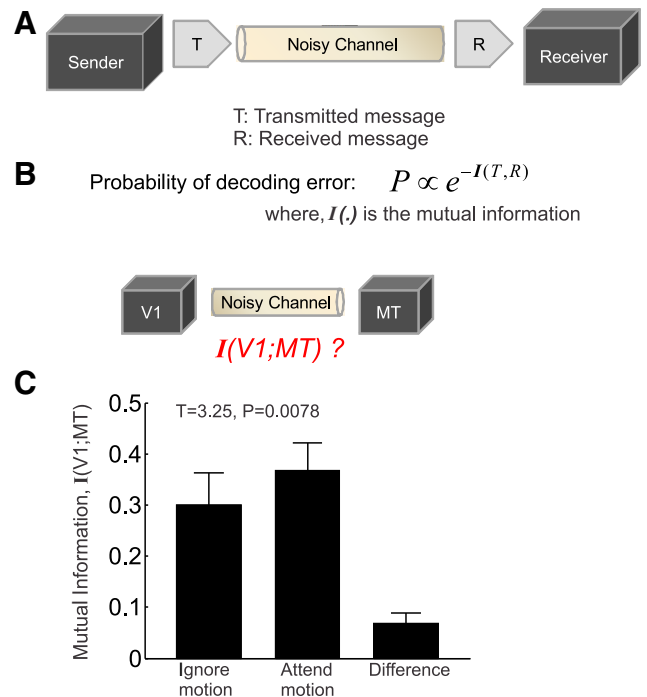


Figure 7. A, A generic communication schema and the probability of error in signal decoding based on Shannon's channel coding theorem. B, The hypothetical communication schema for V1 and MT. The efficacy of communication can be estimated by measuring MI between direction-selective channels in the two areas when they are responding to the same stimulus. C, The MI between direction-selective channels in V1 and MT for ignore motion trials, attend motion trials, and the difference between the two. There is a significant increase in communication efficacy when attention is directed to motion direction ($p < 0.01$). Error bars indicate between-subject ± 1 SEM ($N = 12$).

nodes connected by a channel that has unknown but fixed noise properties, and the measured pattern of direction-specific channel responses in V1 and MT can be viewed as messages that are transmitted from V1 and received in MT. Within the context of this general communication framework, the observation of increased MI between V1 and MT with attention implies an improvement in the efficacy of information transfer between these regions (Fig. 7). That said, our measure of cross-area coupling cannot rule out the influence of a third anatomical structure, such as the pulvinar (Saalmann et al., 2012) or the superior colliculus (Zénon and Krauzlis, 2012; Hafed et al., 2013), which might indirectly modulate activity patterns in both V1 and MT and influence our estimate of cross-area MI. However, even though we cannot rule out that possibility, there are dense connections between V1 and MT (Born and Bradley, 2005) and our model simulation, which does not assume any thalamic mediation, suggests that the direct coupling of V1-MT is at least sufficient to account for the present results. In addition, we cannot assess the possibility that attention improves cross-area communication in part by directly changing the noise characteristics of the communication channel. For example, attention can influence the biophysical properties of the synaptic junctions (Briggs et al., 2013), and this may impact the extent to which noise induced during axonal propagation corrupts neural communication channels.

It has previously been suggested that neurons can communicate or exchange information via synchrony, typically in the gamma band (Gray et al., 1989; Engel et al., 2001; Fries, 2005), and there are emerging data suggesting that attention enhances synchrony between distinct neural populations in different cor-

tical regions (Womelsdorf and Fries, 2007; Gregoriou et al., 2012). If attention increases the oscillatory synchrony between direction-selective neurons in V1 and MT, it might contribute to the increase in cross-area MI that we observed in the present experiment. However, because the BOLD signal has a low temporal resolution, directly evaluating contributions from high-frequency synchrony is not possible using the present data, especially considering that very low-frequency coherence is thought to drive measures of BOLD-based cross-area connectivity (Wang et al., 2012). At such low frequencies, it is also very hard to distinguish between signal and carrier frequency, fundamental concepts in communication-through-coherence theory (Fries, 2005). Thus, future studies using higher-temporal resolution methods will be needed to assess the relationship between synchrony in different frequency bands and the feature-specific information transfer between cortical areas that we report here.

In conclusion, the data processing inequality (Cover and Thomas, 1991) states that any information processing can only reduce the overall amount of information contained in a signal (or data), or at best keep it constant. In the context of neural systems, this inequality suggests that cortical processing diminishes the total sensory information contained in neural signals as they pass through successive stages of the processing hierarchy. Thus, preserving the information about behaviorally relevant objects during information processing, at the expense of behaviorally irrelevant objects, might be a favorable evolutionary strategy. Our results suggest that selective attention implements this strategy in part by changing neural representations associated with relevant features and objects so that they are more robust to channel noise during intercortical communication, and this in turn may lead to more accurate downstream representations.

References

- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101. [CrossRef Medline](#)
- Beck C, Neumann H (2011) Combining feature selection and integration: a neural model for MT motion selectivity. *PLoS One* 6:e21254. [CrossRef Medline](#)
- Born RT, Bradley DC (2005) Structure and function of visual area MT. *Annu Rev Neurosci* 28:157–189. [CrossRef Medline](#)
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4221. [Medline](#)
- Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000) Synergy in a neural code. *Neural Comput* 12:1531–1552. [CrossRef Medline](#)
- Bressler DW, Silver MA (2010) Spatial attention improves reliability of fMRI retinotopic mapping signals in occipital and parietal cortex. *Neuroimage* 53:526–533. [CrossRef Medline](#)
- Briggs F, Mangun GR, Usrey WM (2013) Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. *Nature* 499:476–480. [CrossRef Medline](#)
- Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. *J Neurosci* 29:13992–14003. [CrossRef Medline](#)
- Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. *J Neurophysiol* 106:2108–2119. [CrossRef Medline](#)
- Chai B, Walther DB, Beck DM, Fei-Fei L (2009) Exploring functional connectivity of the human brain using multivariate information analysis. *Adv Neural Information Processing Syst* 22:279–287.
- Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nat Neurosci* 14:811–819. [CrossRef Medline](#)
- Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12:1594–1600. [CrossRef Medline](#)
- Cohen MR, Maunsell JH (2011) Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron* 70:1192–1204. [CrossRef Medline](#)
- Conti F, Wanke E (1975) Channel noise in nerve membranes and lipid bilayers. *Q Rev Biophys* 8:451–506. [CrossRef Medline](#)
- Cover TM, Thomas JA (1991) *Elements of information theory*. New York: Wiley.
- Dale AM (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8:109–114. [CrossRef Medline](#)
- Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716. [CrossRef Medline](#)
- Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI of human visual cortex. *Nature* 369:525. [CrossRef Medline](#)
- Freeman WT, Adelson EH (1991) The design and use of steerable filters. *IEEE Trans Pattern Analysis Machine Intelligence* 13:891–906. [CrossRef Medline](#)
- Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9:474–480. [CrossRef Medline](#)
- Gray CM, König P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338:334–337. [CrossRef Medline](#)
- Gregoriou GG, Gots SJ, Desimone R (2012) Cell-type-specific synchronization of neural activity in FEF with V4 during attention. *Neuron* 73:581–594. [CrossRef Medline](#)
- Hafed ZM, Lovejoy LP, Krauzlis RJ (2013) Superior colliculus inactivation alters the relationship between covert visual attention and microsaccades. *Eur J Neurosci* 37:1169–1181. [CrossRef Medline](#)
- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9:181–197. [Medline](#)
- Heeger DJ, Huk AC, Geisler WS, Albrecht DG (2000) Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? *Nat Neurosci* 3:631–633. [CrossRef Medline](#)
- Huang X, Lisberger SG (2009) Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J Neurophysiol* 101:3012–3030. [CrossRef Medline](#)
- Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–696. [CrossRef Medline](#)
- Kahn I, Desai M, Knoblich U, Bernstein J, Henninger M, Graybiel AM, Boyden ES, Buckner RL, Moore CI (2011) Characterization of the functional MRI response temporal linearity via optical control of neocortical pyramidal neurons. *J Neurosci* 31:15086–15091. [CrossRef Medline](#)
- Latham PE, Roudi Y (2009) Mutual information. *Scholarpedia* 4.
- Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. *Annu Rev Physiol* 66:735–769. [CrossRef Medline](#)
- Manwani A, Koch C (2001) Detecting and estimating signals over noisy and unreliable synapses: information-theoretic analysis. *Neural Comput* 13:1–33. [CrossRef Medline](#)
- Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14:744–751. [CrossRef Medline](#)
- McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19:431–441. [Medline](#)
- Misra N, Singh H, Hnizdo V (2010) Nearest neighbor estimates of entropy for multivariate circular distributions. *Entropy* 12:1125–1144. [CrossRef Medline](#)
- Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55:131–141. [CrossRef Medline](#)
- Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63:879–888. [CrossRef Medline](#)
- Mnatsakanov RM, Misra N, Li S, Harner EJ (2008) kn: nearest neighbor estimators of entropy. *Math Methods Stat* 17:261–277. [CrossRef Medline](#)
- Movshon JA, Newsome WT (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J Neurosci* 16:7733–7741. [Medline](#)
- Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J Neurosci* 31:14551–14564. [CrossRef Medline](#)
- Pascual-Leone A, Walsh V (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292:510–512. [CrossRef Medline](#)

- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185. [CrossRef Medline](#)
- Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9:1421–1431. [CrossRef Medline](#)
- Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337:753–756. [CrossRef Medline](#)
- Saproo S, Serences JT (2010) Spatial attention improves the quality of population codes in human visual cortex. *J Neurophysiol* 104:885–895. [CrossRef Medline](#)
- Schneidman E, Freedman B, Segev I (1998) Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural Comput* 10:1679–1703. [CrossRef Medline](#)
- Schneidman E, Bialek W, Berry MJ 2nd (2003) Synergy, redundancy, and independence in population codes. *J Neurosci* 23:11539–11553. [Medline](#)
- Serences JT, Saproo S (2012) Computational advances towards linking BOLD and behavior. *Neuropsychologia* 50:435–446. [CrossRef Medline](#)
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893. [CrossRef Medline](#)
- Silvanto J, Cowey A, Lavie N, Walsh V (2005) Striate cortex (V1) activity gates awareness of motion. *Nat Neurosci* 8:143–144. [CrossRef Medline](#)
- Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. *Vision Res* 38:743–761. [CrossRef Medline](#)
- Snippe HP, Koenderink JJ (1992) Information in channel-coded systems: correlated receivers. *Biol Cybern* 67:183–190. [CrossRef Medline](#)
- Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* 13:334–350. [Medline](#)
- Treue S, Maunsell JH (1999) Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J Neurosci* 19:7591–7602. [Medline](#)
- Vinje WE, Gallant JL (2002) Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J Neurosci* 22:2904–2915. [Medline](#)
- Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. *Neuron* 56:366–383. [CrossRef Medline](#)
- Wang L, Saalmann YB, Pinsk MA, Arcaro MJ, Kastner S (2012) Electrophysiological low-frequency coherence and cross-frequency coupling contribute to BOLD connectivity. *Neuron* 76:1010–1020. [CrossRef Medline](#)
- Wang Q, Kulkarni S, Verdu S (2009) Divergence estimation of multidimensional densities via k -nearest neighbor distances. *IEEE Trans Information Theory* 55:2392–2405. [CrossRef](#)
- White JA, Rubinstein JT, Kay AR (2000) Channel noise in neurons. *Trends Neurosci* 23:131–137. [CrossRef Medline](#)
- Womelsdorf T, Fries P (2007) The role of neuronal synchronization in selective attention. *Curr Opin Neurobiol* 17:154–160. [CrossRef Medline](#)
- Zénon A, Krauzlis RJ (2012) Attention deficits without cortical neuronal deficits. *Nature* 489:434–437. [CrossRef Medline](#)