

A normalization model suggests that attention changes the weighting of inputs between visual areas

Douglas A. Ruff^{a,1} and Marlene R. Cohen^a

^aDepartment of Neuroscience and Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, PA 15213

Edited by David J. Heeger, New York University, New York, NY, and approved April 12, 2017 (received for review December 5, 2016)

Models of divisive normalization can explain the trial-averaged responses of neurons in sensory, association, and motor areas under a wide range of conditions, including how visual attention changes the gains of neurons in visual cortex. Attention, like other modulatory processes, is also associated with changes in the extent to which pairs of neurons share trial-to-trial variability. We showed recently that in addition to decreasing correlations between similarly tuned neurons within the same visual area, attention increases correlations between neurons in primary visual cortex (V1) and the middle temporal area (MT) and that an extension of a classic normalization model can account for this correlation increase. One of the benefits of having a descriptive model that can account for many physiological observations is that it can be used to probe the mechanisms underlying processes such as attention. Here, we use electrical microstimulation in V1 paired with recording in MT to provide causal evidence that the relationship between V1 and MT activity is nonlinear and is well described by divisive normalization. We then use the normalization model and recording and microstimulation experiments to show that the attention dependence of V1–MT correlations is better explained by a mechanism in which attention changes the weights of connections between V1 and MT than by a mechanism that modulates responses in either area. Our study shows that normalization can explain interactions between neurons in different areas and provides a framework for using multiarea recording and stimulation to probe the neural mechanisms underlying neuronal computations.

attention | normalization | variability

Divisive normalization, which describes nonlinearities in the way that neurons respond to combinations of stimuli, can explain the trial-averaged responses of neurons in many brain areas in many animals and under a wide range of conditions (1). For example, several studies have used normalization models to characterize how visual attention, which improves perception of attended stimuli, scales the responses of neurons in visual cortex (2–7).

Attention, like other processes that divisively scale responses, is also associated with many other changes to the responses of neurons in visual cortex. For example, we recently recorded simultaneously from individual neurons in the middle temporal area (MT) and several dozen neurons in primary visual cortex (V1) (8). In contrast to the well-known attention-related decreases in the correlated variability of the responses of similarly tuned neurons within the same cortical area (8–16), we found that attention increases the extent to which neurons in V1 and MT share trial-to-trial variability (termed spike count or noise correlations, or r_{SC}).

We hypothesized that the attention-related increase in cross-area correlations would place constraints on the neuronal mechanisms underlying attention, but it is consistent with at least three possible non-mutually exclusive mechanisms (Fig. 1). One possibility is that attention modulates the responses of V1 neurons, perhaps through a top-down input. In our previous study, as in many others (17–19), attention slightly increased the firing rates of V1 neurons (8). In this scenario, the increased rates of V1 neurons whose receptive fields overlap the attended location increases their influence on the MT neuron, so the attention-related increase in cross-area correlations is simply inherited from the changes in V1. Second, attention might increase the effective weighting or influence of direct or

indirect connections between V1 and MT. This possibility is consistent both with a long-standing hypothesis that attention increases communication between areas on short timescales (20), and with a model that suggests attention similarly increases the weight of both excitatory drive and divisive inhibition when multiple stimuli are in a neuron's receptive field (6). Finally, attention might influence the membrane potential or some other aspect of the activity of MT neurons directly, changing the proportion of EPSCs from V1 that produce spikes in MT.

The goal of this study is to use a combination of simulations and experiments to determine which hypothesized mechanism best accounts for the attention-related changes in cross-area correlations we observed. Our simulations require a descriptive model that can account for firing rates and correlations of neuronal responses. We considered two models: a normalization model and a linear model. We previously showed that a normalization model of attention that hypothesizes that the tuning of MT neurons comes from their feedforward inputs can account for many aspects of the mean responses of V1 and MT neurons and the V1–MT correlations we measured (21). In particular, the model accurately predicts that attention increases correlations between neurons in V1 and MT, and that cross-area correlations depend on stimulus properties (such as motion direction) that do not affect within-area correlations. Although this model accounts for the mean rates of MT neurons and the V1–MT correlations substantially better than a model in which the responses of V1 and MT neurons are linearly related (21), the linear model does not fail completely, and the better performance of the normalization model is largely quantitative, rather than qualitative.

We realized that the linear and normalization models made qualitatively different predictions about the way that adding extra spikes in V1 would affect responses in MT. We found that electrically stimulating neurons in V1 affected MT responses in a way

Significance

Visual attention dramatically improves the perception of attended stimuli, but the mechanism by which this improvement occurs is unknown. Here we combine simultaneous recordings in multiple visual cortical areas, electrical microstimulation paired with multi-neuron recordings, and a descriptive normalization model to probe the mechanisms underlying attention. We found that the relationship between the activity in different visual areas is well described by normalization and that a mechanism that changes the weights of connections between areas can best describe our physiological observations. This study shows that normalization can explain interactions between neurons in different areas and provides a framework for using multiarea recordings and stimulation techniques to probe the neural mechanisms underlying neuronal computations.

Author contributions: D.A.R. and M.R.C. designed research; D.A.R. performed research; D.A.R. and M.R.C. contributed new reagents/analytic tools; D.A.R. and M.R.C. analyzed data; and D.A.R. and M.R.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. Email: ruffd@pitt.edu.

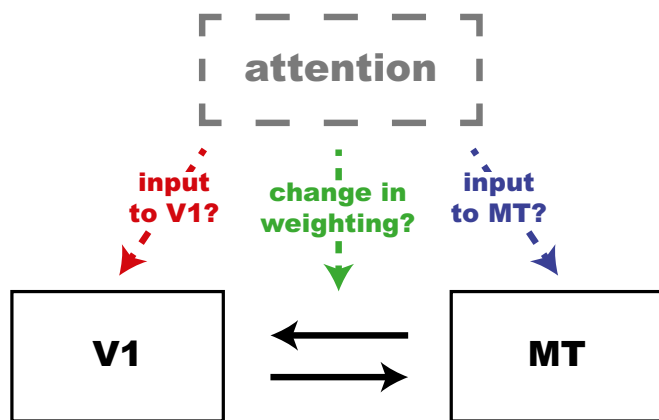


Fig. 1. Schematic of non-mutually exclusive possible mechanisms that could account for the attention-related increase in r_{SC} between V1 and MT. First, attention could modulate the responses of V1 neurons, perhaps through a top-down input. Second, attention might increase the effective weighting or influence of direct or indirect connections between V1 and MT. Finally, attention might influence the activity MT neurons directly.

that was consistent with the predictions of the normalization model and was inconsistent with the predictions of a linear model.

This affirmation of the normalization model, combined with our previous observations that the normalization model can account for many physiological observations, provides a strong rationale for using the normalization model to dissociate between the possible explanations for the attention-related increase in cross-area correlations. We used simulations to show that the attention-related increase in V1–MT correlations is inconsistent with the hypothesis that the correlation increase is a by-product of attention-related changes within V1. However, the simulations showed that our observations are well explained by the idea that attention changes the weighting, or influence, of direct or indirect connections between V1 and MT. We then used recording and microstimulation experiments to show that the weight change is more likely to be caused by a change in the efficacy of direct or indirect synapses between V1 and MT than by changes to the response properties of the MT neurons themselves. Our study therefore suggests that attention acts at least in part by improving the communication of neurons representing the attended parts of a visual scene. More generally, our study expands current models to show that normalization can capture the transformation of sensory information between brain areas and provides a framework for using correlated variability to probe the neural mechanisms underlying canonical neural computations.

Results

A Causal Test of the Normalization Model of Attention. The goal of the current study is to use the normalization model to investigate the neuronal mechanisms underlying attention; we did so below using simulations that are based on the normalization model. The logic behind this approach requires accepting that normalization provides a good account of the way that mean firing rates and cross-area spike-count correlations depend on the visual stimulus or task condition.

Divisive normalization is well known to provide a good account of the trial-averaged responses of neurons in a wide variety of model systems, brain areas, and stimulus and task conditions (1). We recently showed that normalization can also account for cross-area correlations: a simple extension of a normalization model in which the responses of pools of V1 neurons are hypothesized to represent feedforward inputs to modeled MT neurons can also account for the way that spike count correlations between V1 and MT depend on the visual stimulus and on the animal's attentional state (21) (Eqs. 1–3). The hypothesis underlying that extension of

the normalization model is that the tuned component of the MT neuron's response would come from neurons whose responses are correlated on a trial-to-trial basis with the average response of the two groups, or pools, of V1 neurons we recorded.

The results of our simulations depended strongly on the normalization terms in the denominator of Eqs. 1–3, so it is important to justify their inclusion in the model. We showed previously that normalization provides a better account of trial-averaged rates and correlations than a linear model that does not include those normalization terms (21). However, neither model perfectly succeeds nor perfectly fails to account for our observations. In our data set, the normalization model (Eqs. 1–3) accounts for 92% of the variance in the mean responses of the MT units we recorded across stimulus and attention conditions, and 65% of the variance in the correlations between the MT units and V1 pools. A linear model in which the response of the MT neuron is modeled as the numerators of Eqs. 1–3 perform significantly worse, but still accounts for the majority of the variance in both cases (85% of the variance in mean responses and 53% of the variance in V1–MT correlations).

To further justify the normalization terms in our model, we used a causal manipulation to differentiate between the normalization and linear models. The normalization model predicts that the effect of extra spikes in a pool of V1 neurons on MT responses should depend on stimulus contrast, because the MT unit's response depends on the relative activity of the pools of V1 neurons whose receptive fields overlap the stimuli moving in the MT unit's preferred and null directions. For example, adding ε extra spikes to the activity of the pool of V1 neurons whose receptive fields overlap the MT unit's preferred stimulus [so that the activity of that pool is given by $V1_P(c_P) + \varepsilon$] while the animal is attending toward the stimulus moving in the MT unit's preferred direction would increase the response of the MT unit by $\beta s_P \varepsilon / (\beta c_P + \alpha c_N + \sigma)$.

When contrast is low, the denominator is small, so the normalization model predicts a bigger increase than when the contrast is high. In contrast, the linear model predicts that adding ε extra spikes to the V1 pool will produce the same number of extra spikes in the MT unit (given by $\beta s_P \varepsilon$) for any stimulus contrast.

We tested this prediction by measuring the number of extra MT spikes produced using subthreshold microstimulation in V1 while the animal performed the direction change detection task on visual stimuli presented at different contrasts (Fig. 2). Although stimulation artifacts prevented us from measuring the effect of V1 electrical stimulation directly, we hypothesized that the effect of V1 microstimulation on V1 responses would not depend on the contrast of the visual stimulus; if so, the normalization model predicts that electrically stimulating V1 should have a greater net effect on MT responses (whether that effect is suppressive or excitatory, depending on whether the number of extra V1 spikes ε is positive or negative) when the visual stimulus is low rather than high contrast. Furthermore, because the attention term β appears in both the numerator and denominator of Eqs. 2 and 3, the normalization model predicts that the difference between the number of MT spikes produced by V1 microstimulation at low and high contrast should not depend strongly on attention.

The linear model makes very different predictions. This model predicts that the number of extra MT spikes produced by V1 microstimulation should not depend on contrast. The model therefore predicts that the difference in the number of extra MT spikes when the stimulus is low and high contrast should be near zero regardless of attention.

To visualize these predictions, we fitted the normalization model to the 64 MT units/stimulus conditions for which we simultaneously recorded the responses of several dozen V1 units (*Methods*). In a previous experiment, we observed that V1 microstimulation could have either a suppressive or excitatory effect on MT rates during the time period of microstimulation (Fig. 3), and we could not predict the sign or magnitude of these effects from the properties of the

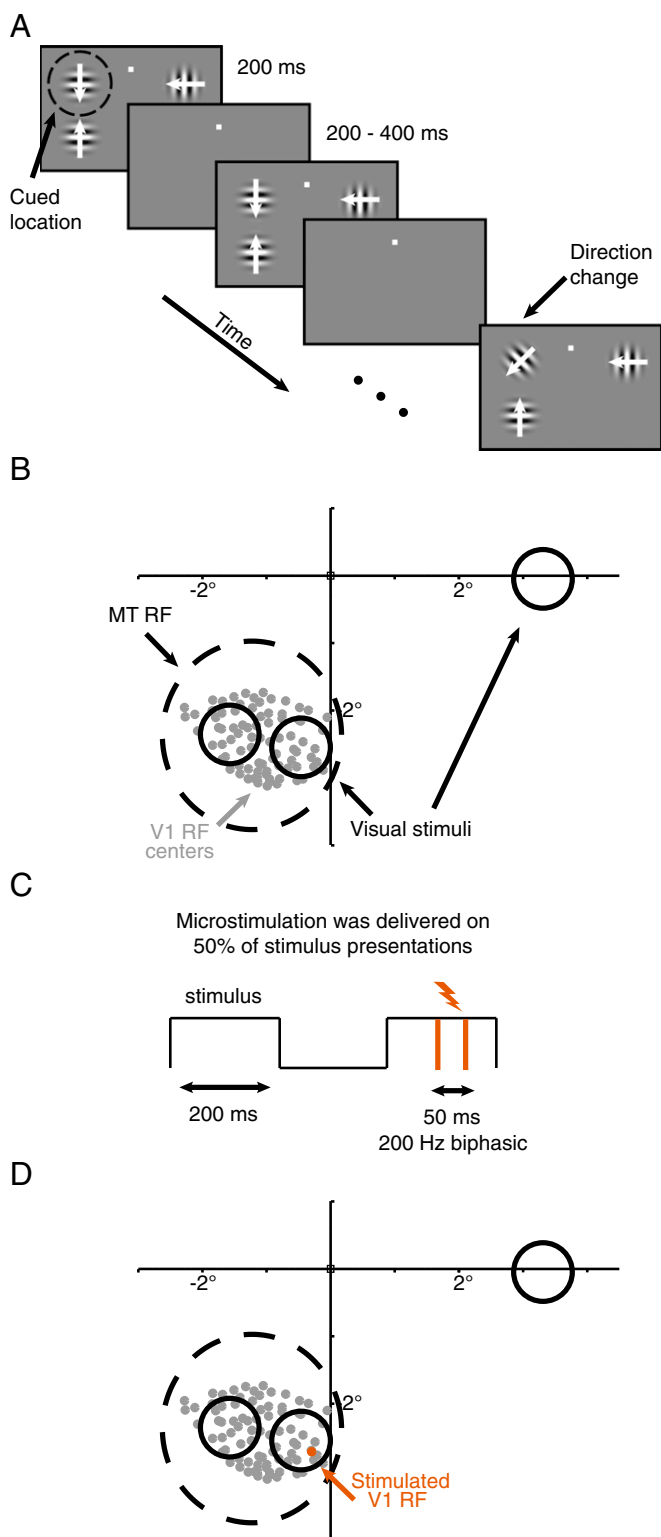


Fig. 2. Recording, microstimulation, and psychophysical methods. (A) Schematic of the motion direction change detection task used in all experiments. Once the monkey fixated a central spot, either two or three small Gabor stimuli synchronously flashed on for 200 ms and off for a randomized 200- to 400-ms period. Two of the stimuli were positioned inside the joint receptive fields of the MT and V1 neurons we recorded, and the third (when present, which was on one out of every three blocks of trials) was placed in the opposite hemifield. After an unsignaled and randomized number of stimulus presentations (picked from an exponential distribution, minimum 2, mean 6, maximum 14 stimulus presentations), the direction of one of the stimuli changed. The monkeys were cued in

neurons we measured (8). We therefore randomly assigned a microstimulation effect ε to each pool of V1 units and assumed that this effect of V1 stimulation on V1 rates varies from trial to trial according to Poisson statistics.

Fig. 4 plots the predictions of the normalization and linear models concerning the number of extra MT spikes elicited by V1 microstimulation when the visual stimulus has low or high contrast (Fig. 4 A and C) and the effect of attention on the difference between the number of extra MT spikes elicited by V1 microstimulation during presentations of low- and high-contrast visual stimuli (Fig. 4 B and D). These simulations are imperfect in several ways: they are based on different V1 and MT units than those whose responses we recorded during the microstimulation experiments; they involve assumptions about the lack of dependence of the effect of V1 microstimulation on V1 rates on stimulus contrast; and they involve randomly assigning a microstimulation parameter ε to each pool of V1 units. However, it is clear that the normalization model and the linear model make qualitatively different predictions.

Our observations strongly support the predictions of the normalization model. The effect of V1 microstimulation on MT responses was strongly correlated for low- and high-contrast visual stimuli ($R = 0.59$, $P < 10^{-17}$), but the slope of the best-fit line relating the two is significantly less than 1 (slope = 0.37, bootstrap, $P < 10^{-4}$; Fig. 4E). The difference in the number of extra MT spikes elicited during low- and high-contrast visual stimuli when the animal attended toward vs. away from the receptive fields of the stimulated V1 units was also strongly correlated ($R = 0.96$, $P < 10^{-103}$), but the slope of the best-fit line was not significantly different from 1 (slope = 0.97, bootstrap, $P = 0.48$; Fig. 4F). These data, although not identical to the predictions of the normalization model, are much more similar to the predictions of the normalization model than those of the linear model.

Our simulations assume that the number of extra V1 spikes elicited by microstimulation did not depend on the contrast of the visual stimulus. One possibility, however, is that the firing rates of the V1 neurons were closer to saturation during presentations of high- than low-contrast visual stimuli. In this case, a ceiling effect might have reduced the number of extra V1 spikes produced by microstimulation on high-contrast trials. Stimulation artifacts prevent us from measuring the number of extra V1 spikes elicited, so we used simulations to determine how this sort of ceiling effect

blocks of 50–100 trials to detect changes in (and therefore attend to) one of the stimuli and ignore motion direction changes in the other stimulus locations. The monkeys were rewarded for making a saccade to the attended stimulus within 500 ms of the stimulus change. Distractor changes were never rewarded. The contrasts of the stimuli were either all 8% or all 100%, and contrast was randomly interleaved on each stimulus presentation. The two stimuli within the receptive field moved in opposite directions (the preferred and null directions of the MT cell under study for the experiments used to measure spike count correlations and fit the models), and which of the two stimuli moved in the preferred direction varied randomly from trial to trial. When present, the stimulus in the opposite hemifield moved in a direction that was orthogonal to the directions of the stimuli inside the MT units' receptive fields. (B) Receptive fields and visual stimulus locations from an example experimental session. We recorded simultaneously from a 96-channel chronically implanted microelectrode array in area V1 and a single electrode or a movable 24-electrode probe in area MT. We selected MT units whose receptive fields (black dashed circle in this example) overlapped the envelope of receptive fields of the neurons we recorded in V1 (centers denoted by the gray dots). The locations and approximate sizes of the visual stimuli are denoted by the solid black circles. For the electrical microstimulation experiments, (C) microstimulation was delivered for 50-ms pulses at 200 Hz during a randomly selected 50% of stimulus presentations in each trial. (D) Channels were selected for microstimulation if the receptive fields of the neurons on them clearly overlapped with just one stimulus, as the orange receptive field center does, and if microstimulation led to an observable change in MT firing rates during microstimulation.

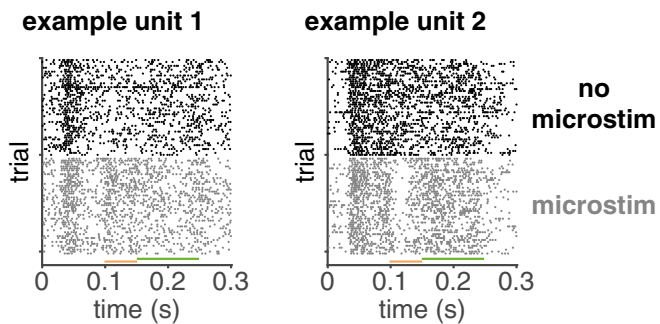


Fig. 3. Example raster plots from two example MT units of activity with and without electrical microstimulation in V1. The rasters show the first 50 non-microstimulated (black) and microstimulated (gray) high-contrast stimulus presentations from trials in which the monkey directed its attention toward the stimulus overlapping the receptive fields of the microstimulated V1 neurons. The visual stimulus came on at time 0, and the orange bar depicts time period during which the microstimulation occurred. Fig. 4 is based on spikes during the microstimulation period, and the green bar depicts a 100-ms period following microstimulation during which spikes were counted for the analyses presented in Fig. 7B.

would change the predictions of the linear model (Fig. 4 C and D). To match the slope of the line relating the effect of V1 microstimulation on MT responses to low- and high-contrast visual stimuli using the linear model, we needed to assume that microstimulation produces 3.84 times as many V1 spikes on low- than high-microstimulation trials; this is a large difference, but it is possible if the responses of most V1 neurons to high-contrast stimuli are near saturation.

Two factors make us think the scenario that a ceiling effect combined with a linear relationship between V1 and MT responses cause the results in Fig. 4 is unlikely. First, the responses of most V1 units to high-contrast visual stimuli were likely not especially

close to saturation because the second stimulus was in the suppressive surround of most units. Second, we found previously that the normalization model captures substantially more of the stimulus and attention variability in mean rates and V1–MT spike count correlations (21). We therefore conclude that the most plausible interpretation of the results in Fig. 4 is that they provide causal support for the argument that normalization provides a good account of the relationship between activity in V1 and in MT.

Attention Affects the Responses of V1 Neurons and Interactions Between V1 and MT. Our primary goal was to use observations about attention-related changes in the rates, variability, and spike count correlations between pairs of V1 units or between units in V1 and MT to evaluate hypothesized mechanisms underlying attention. We previously reported that during this task, attention increased the trial-averaged firing rates of V1 units, decreased the spike count correlations between pairs of V1 units, and increased the spike count correlations between V1–MT pairs (8).

Because our normalization model takes as inputs the summed activity of pools of V1 units whose receptive fields overlap the visual stimulus moving in either the MT neuron's preferred or null direction, we calculated the effects of attention on various measures of the activity of pools of V1 neurons as a whole (see *Methods* for details of pool assignment). Directing attention to the receptive fields of the pool of V1 neurons was associated with modest firing rate increases [mean = 1.84 spikes per second (sp/s), which is significantly greater than zero, *t* test, $P < 10^{-8}$]. The overall decrease in the spike count correlations between pairs of V1 units corresponds to a decrease in the variance of the activity of the pool as a whole (mean = $-2.87 \text{ sp}^2/\text{s}$, which is significantly less than zero, *t* test, $P < 0.05$). Because the responses of the V1 units that made up each pool were overwhelmingly positively correlated, the attention-related increase in V1–MT spike count correlations was greater for the average of the pool, mean = 0.039, which is significantly greater than zero, *t* test, $P < 0.03$) than for the individual units (0.018) (8).

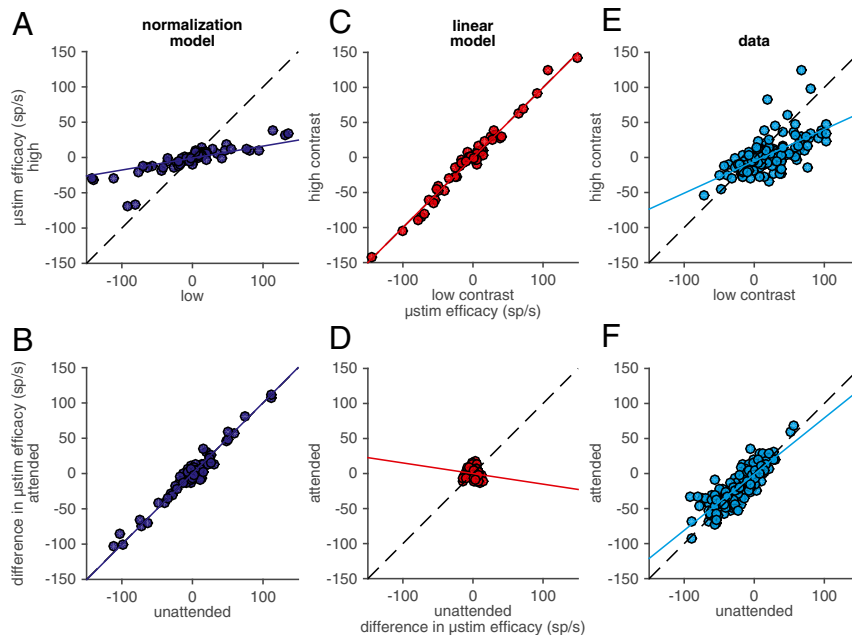


Fig. 4. Electrical microstimulation results support the normalization model. (A) Predictions of the normalization model for the number of extra MT spikes elicited by V1 microstimulation when the visual stimulus was high contrast (y axis) vs. low contrast (x axis). The dashed line is unity, and the solid line is the best fit line to the model's predictions. (B) Predictions of the normalization model for the difference between the numbers of extra MT spikes elicited by V1 microstimulation during high- and low-contrast visual stimulus presentations when the animal attends (y axis) or ignores the visual stimulus overlapping the receptive fields of the stimulated V1 neurons. Conventions as in A. (C and D) Same as A and B for the linear model. (E and F) Same as A and B for actual data.

The Normalization Model Suggests That the Attention-Related V1–MT Correlation Increase Comes from an Increase in the Weighting of V1 Inputs to MT. The normalization model allows us to distinguish between two of the possible sources of the attention-related increase in V1–MT correlations we observed. The model lets us evaluate the feasibility of the ideas that the attention-related increase in cross-area correlation is simply inherited from attention-related changes in V1 (quantified as changes in either the rates or variance of the activity of the pool of V1 units) or comes from changes in the influence of V1 neurons on MT responses. If attention were associated with sufficiently large increases in the rate or increases in the variance of the attended pool of neurons, those neurons would have a proportionally larger effect on the activity of the MT neuron than those in the unattended pool, which would cause a correlation increase. The hypothesis seems unlikely because the V1 rate changes we observed were small and attention was associated with decreases, rather than increases, in the variance of the attended pool of neurons. Alternatively, the V1–MT correlation increase could be caused by attention-related changes in the weighting, or influence, of the V1 neurons that represent the attended and unattended stimuli (which corresponds to the parameter β in our model). (We address a third possibility, that attention modulates the responses of MT units directly, using experimental results below.) To distinguish between the first two possibilities, we simulated the responses of pools of V1 neurons (*Methods*) to determine the impact of changing their responses or changing the weighting parameter β on the attention-related changes in V1–MT correlations predicted by the model.

Although the actual attention-related changes we observed in the V1 neurons we recorded were very small (8), we varied the magnitude of the simulated attention-related changes in V1 to allow for the possibility that subsets of V1 units whose responses are unusually modulated by attention have a proportionally bigger or smaller impact on the responses of MT units. Attention has at least two effects on populations of visual neurons, multiplicatively scaling firing rates (17–19) and changing the variance or covariance of neuronal responses (8–16, 22). Changes in either variance or covariance would change the variance of the mean response of the pool of neurons. We therefore tested the effect of changing the mean (x axis in Fig. 5*A*) and variance (y axis in Fig. 5*A*) of the modeled V1 neurons on the predicted attention-related change in V1–MT correlation (color in Fig. 5*A*).

Even when we simulated pools of V1 neurons with physiologically large (3,200%) changes in mean and/or variance, the predicted attention-related change in V1–MT correlation was smaller than we observed (actual change in correlations between pools of V1 units and MT = 0.039, which is the brightest yellow in the color scale of Fig. 5*A* and never appears in the main plot; actual change in V1 rate/variance is the white X in Fig. 5*A*, which corresponds to a very small change in V1–MT correlation). These simulations suggest that although it is possible that changes within V1 contributed to the attention-related changes in V1–MT correlation, V1 changes were not the primary source of the observed correlation increase.

We also addressed the possibility that the attention-related increase in V1–MT correlation was caused by changes in the weighting or influence of V1 neurons representing the attended stimulus on MT responses. This idea is consistent with the long-standing hypothesis that attention changes communication between visual areas (for review, see ref. 20). To determine the effect of changing the weighting of particular V1 inputs on V1–MT correlations, we varied the parameter β in Eqs. 2 and 3 (x axis of Fig. 5*B*; *Methods*).

Varying the parameter β had a large effect on the model's predicted attention dependence of V1–MT correlations (y axis in Fig. 5*B*). The mean fitted value of β was 1.96 (vertical line in Fig. 5*B*), and the observed attention-related change in V1–MT correlation was 0.039 (horizontal line in Fig. 5*B*). The fact that the black line intersects the crossing of the dashed lines so closely

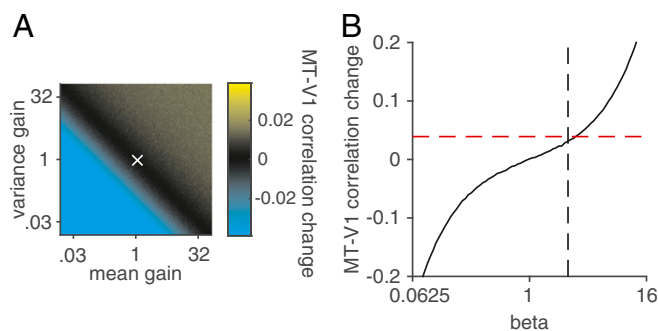


Fig. 5. The attention-related increase in correlations between V1 and MT is more consistent with changes in the weighting of V1 inputs than with changes within V1. (*A*) The effect of changing the mean (x axis) and variance (y axis) of the responses of the V1 pools on predicted V1–MT correlations. Color represents the predicted change in V1–MT correlation. The bright yellow at the top of the color bar (which never appears in the main plot) represents the observed attention-related change in V1–MT correlation, and the white x represents the observed attention-related changes in mean response and variance of the V1 pools. (*B*) Predicted V1–MT correlation change as a function of the parameter β in Eqs. 1–3. The red dashed line represents the observed attention-related change in V1–MT correlation, and the black dashed line represents the fitted value of β .

indicates that the attention-related change in V1–MT correlation can be almost completely accounted for in the model by changes in the parameter β .

The simulations in Fig. 5 rely on a quantitative comparison between the model's prediction and the data. However, in the model, the only source of variability in MT responses is inherited from V1. As a result, the raw V1–MT correlations are unrealistically high and the Fano factor of the modeled MT unit is unrealistically low. In a previous study (21), we compensated for this by adding independent, zero-mean Gaussian noise to the responses of the MT unit until we quantitatively matched the mean V1–MT correlation across all conditions. This modified model was inaccurate in a different way, however, because quantitatively matching the correlations required adding so much noise that the average Fano factor of the modeled MT units was 106, which is much larger than for the recorded MT units (which was 1.41).

Importantly, however, the modified model made quantitatively nearly identical predictions about how attention changes V1–MT correlations. For example, the original model predicts that the attention-related change in correlation will match the data when $\beta = 1.86$, whereas the modified (high noise) model predicts a match to the data when $\beta = 1.91$, neither of which is statistically distinguishable from the median fitted β , which was 1.96 (t tests, $P > 0.05$). Therefore, both versions of the normalization model suggest that the most parsimonious explanation of our results is that attention changes the influence of the V1 neurons that represent the attended stimulus on MT responses.

Experimental Evidence Suggests That Attention Improves Synaptic Efficacy Between V1 and MT.

Our simulations suggest that attention changes the effective influence of the V1 units whose receptive fields overlap the attended stimulus; this could be accomplished by changing the efficacy of the synapses that either directly or indirectly connect V1 to MT or by changing the membrane voltage or some other aspect of the response properties of the MT neuron that would make spikes from V1 inputs more or less likely to elicit spikes in MT. Because our normalization model is descriptive rather than mechanistic, these potential mechanisms are indistinguishable in our model (and are also indistinguishable in a previous study that measured attention-related changes in the way that thalamic inputs influence the activity of V1 neurons) (23). However, the two mechanisms make different predictions that are testable experimentally.

The efficacy, or weighting, of synapses that either directly or indirectly connect V1 to MT is monotonically related to both the correlation between V1 and MT and the effect of microstimulating V1 on MT responses. Fig. 5B shows the monotonic relationship between the parameter β and the predicted attention-related change in V1–MT correlation. Because we calculate V1–MT correlation for individual stimulus and attention conditions, the synaptic efficacy is proportional to the parameter β for an individual condition. (The other parameters, including the denominator, are constants for an individual condition.) Therefore, Fig. 5B also shows that for an individual condition, V1–MT correlation is monotonically related to the synaptic efficacy. An attention-related increase in synaptic efficacy should always increase V1–MT correlation, regardless of whether attention increases or decreases the rate of the MT cell. By similar logic, an attention-related increase in synaptic efficacy should always increase the effect of microstimulating V1 on the firing rate of the MT unit.

In contrast, the hypothesis that the attention-related changes to V1–MT correlation or to microstimulation efficacy that we observed are caused by changes to the MT cell makes different predictions. Fig. 6 shows the relationship between the MT cell's baseline membrane potential and V1–MT correlation. Borrowing a procedure from previous work (24, 25), we simulated correlated membrane potentials for the V1 and MT cells by picking values for each trial from a bivariate normal distribution (Fig. 6A and B,

Left; the panels in A and B are identical except for the mean of the putative MT neuron's membrane potential). We simulated the relationship between the membrane potential of the MT cell and its firing rate using a nonlinearity used in previous studies (26); Fig. 6A and B, Center; the exact shape of the nonlinearity does not qualitatively affect these results).

Our simulations show that when the MT neuron's mean membrane potential is above threshold (Fig. 6A), the spiking correlation is similar to the membrane potential correlation. When the mean membrane potential is low, however (as in Fig. 6B), the spiking correlation is substantially lower than the membrane potential correlation. The relationship between the MT neuron's membrane potential and the spiking correlation is summarized in Fig. 6C. The quantitative relationship depends on experimentally inaccessible parameters such as the synaptic weight, the spiking threshold, or the shape of the threshold nonlinearity. These simulations show, however, that the relationship between the mean membrane potential of the MT unit and the V1–MT correlation and, by extension, the effect of adding spikes to V1 on the MT response, is monotonic.

Because our recordings were extracellular, we were unable to measure the effect of attention on the mean membrane potential of the MT unit. However, because the relationship between membrane potential and firing rate is monotonic, we can use the sign of the attention-related change in firing rate as a rough estimate of the change in membrane potential. Because on average,

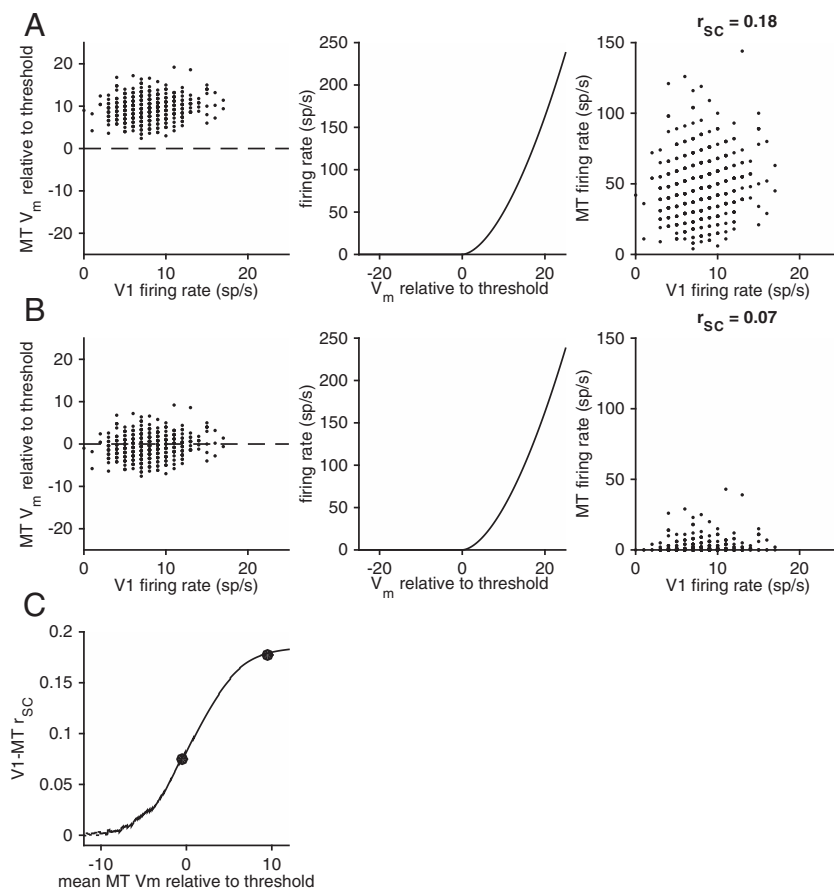


Fig. 6. V1–MT correlations depend monotonically on the MT neuron's membrane potential. (A and B, Left) Membrane potentials of simulated MT neurons that differ only in their membrane potential relative to threshold as a function of the firing rate of a simulated V1 neuron. In both cases, the correlation between V1 and MT membrane potentials was picked to be 0.2. (A and B, Center) The nonlinear relationship between the neuron's membrane potential and firing rate (picked to be realistic, although the exact shape of the nonlinearity does not qualitatively affect the results). (A and B, Right) When the mean membrane potential of the MT neuron is substantially above threshold, the spiking correlation between the simulated V1 and MT neurons is similar to the membrane potential correlation. When the mean membrane potential of the MT neuron is low, however, the spiking correlation is lower than the membrane potential correlation. (C) Summary of the relationship between the relative membrane potential of the MT neuron and the spiking correlation.

attention slightly increases the rate of V1 units, attention should increase V1–MT correlations when attention increases the rate of the MT unit and decrease correlations when attention decreases the rate of the MT unit.

Our data are consistent with the idea that attention changes correlations between V1 and MT or the effect of V1 microstimulation on MT responses by changing the effective weight between V1 and MT and are inconsistent with the hypothesis that these changes are the result of membrane potential changes to MT units. Attention increased V1–MT correlations regardless of whether attention increased or decreased the rate of the MT unit (median attention-related correlation change for MT units for which attention decreased firing rate = 0.0172, which is significantly greater than zero, Wilcoxon signed-rank test, $P < 0.0001$; right bar of Fig. 7A; median attention-related correlation change for MT units for which attention increased firing rate = 0.0197, which is significantly greater than zero, Wilcoxon signed-rank test, $P < 10^{-5}$; left bar of Fig. 7A; attention-related change r_{SC} was not significantly different for the two groups, $P = 0.72$, Wilcoxon rank sum test). Furthermore, attention increased the number of extra MT spikes that followed V1 microstimulation both when attention increased or decreased the rate of the MT cell (Fig. 7B, which plots the attention-related change the number of extra MT spikes elicited by V1 microstimulation minus the attention related change in the response of the MT unit). This number, termed microstimulation efficacy, was positive for both subgroups and the two subgroups did not significantly differ ($P = 0.33$, Wilcoxon rank sum test).

A final possibility is that attention increases the strength or activity of a common, perhaps top-down, input to both V1 and MT. Such an increase could account for the attention-related increase in V1–MT r_{SC} , but there are two problems with this hypothesis. First, an increase in the strength or activity of a common input to the two areas would likely increase correlations both between V1 and MT and within each area, which is contrary to the results of many studies, including ours, that show attention reduces correlations within each area (8–16). Second, it is difficult to imagine how increasing a common input could account for the attention-related increase in the effect of V1 microstimulation on MT responses we observed.

Together, our data and simulations suggest that the most likely explanation for the attention-related increases in spike count correlations between pairs of V1 and MT neurons come from improvements in the efficacy of direct or indirect synapses between V1 and MT. Future experiments measuring the strengths of these synapses directly would be necessary to verify this hypothesis.

Discussion

We used a combination of simulations and correlative and causal experiments to investigate the neuronal mechanisms underlying the attention-related increase in V1–MT correlations that we observed. We used electrical microstimulation to causally test the hypothesis that normalization provides a good description of the transformation of signals between cortical areas. We then used a normalization model to show that attention-related changes in correlations between V1 and MT are unlikely to result from changes within V1 and are most consistent with a mechanism in which attention changes the weighting of connections between visual areas. Finally, we used correlation and microstimulation data to argue against the idea that this weighting increase comes from changes within MT.

Normalization as a Description of How Signals Are Transformed Between Cortical Areas. A large and growing literature shows that normalization provides a good description of the trial-averaged responses of neurons in a variety of species, brain areas, stimuli, and task conditions, and it has been proposed to be a canonical neural computation (for review, see ref. 1). In particular, normalization accurately accounts for the divisive (or multiplicative) scaling of

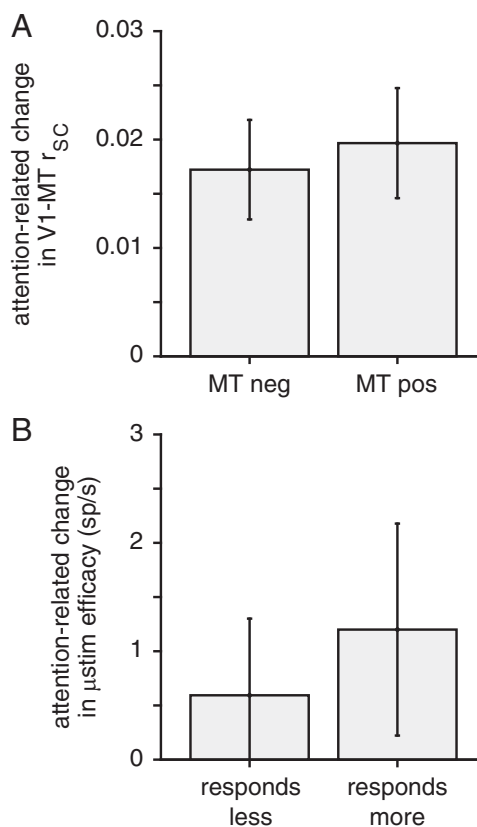


Fig. 7. Correlative and causal evidence that the attention-related changes in the weighting between V1 and MT are not accomplished through changes within MT. (A) Median attention-related change in V1–MT r_{SC} under conditions when attention is associated with decreases in the rate of an MT unit (Left) or increases in the rate of an MT unit (Right). Error bars represent SEM. (B) Median number of extra spikes elicited by V1 electrical microstimulation in MT units that respond less (Left) or more (Right) to the visual stimulus overlapping the receptive fields of the stimulated V1 units. Conventions as in A.

neuronal responses associated with a wide variety of modulatory processes such as the contrast or orientation of visual stimuli (1, 27, 28), multisensory integration (29), reward (30), and attention (2–7).

The present study, combined with our previous work (21), suggests that normalization provides a good description for the way that visual signals are transformed from V1 to MT as well. Substituting the responses of groups of V1 neurons for the linear, stimulus-dependent terms in conventional models accounts for correlated variability that is shared across areas (21). The same model predicts that adding spikes to one pool of V1 neurons should affect MT responses in a way that depends on the contrast of the visual stimuli, and we verified this prediction using electrical microstimulation paired with multineuron recordings (Fig. 4).

The ability of the normalization model to account for the way that signals are transformed between V1 and MT makes early visual cortex an important system for studying the neural mechanisms underlying what may be a canonical neural computation. In the future, experiments using paired recordings and causal manipulations (ideally targeting specific neuronal subtypes) in the two areas using a variety of visual stimuli and behavioral tasks may begin to uncover the circuit mechanisms underlying normalization. Additionally, because of the task and stimulus demands associated with recording from a population of V1 neurons and because neurons in V1 tend to show small modulations due to cognitive and behavioral factors (18, 31), it will be important to replicate these findings in other sensory areas where larger effects can be observed.

Implications for the Neural Mechanisms Underlying Attention. Our results suggest that attention changes the weighting of either direct or indirect inputs from V1 to MT. This idea is broadly consistent with a long-standing hypothesis that attention acts by increasing communication between cortical areas (20), although this theory is mostly based on measurements of synchrony between visual cortex and higher-order cortex (32–37) or between thalamus and primary visual cortex (23) on very fast (millisecond-level) timescales. The results of our microstimulation experiments and simulations suggest that attention changes communication between areas on longer timescales as well. This finding is also consistent with predictions from an earlier model that was designed to explain changes in firing rate in V4 neurons during situations when attention is split between one of two stimuli in a neuron's receptive field (6).

Many studies have shown that attention is associated with several changes in the responses of populations of neurons such as changes firing rates (17–19) and spike count correlations between pairs of neurons in the same cortical area (8–16). Although the relationship between these changes within a cortical area and population coding is a matter of current study and debate (38–43), the attention-related changes in rates and correlations are in a direction that would be expected to improve the information that populations of neurons can encode about visual stimuli (9, 11, 15).

Our simulations using the normalization model and our correlation and microstimulation data suggest that the between area attention-related correlation increase is not likely explained by changes to the properties of neuronal populations in either V1 or MT, which might be accomplished by top-down inputs to the two areas. One possibility is that the two mechanisms (increases in cross-area communication and improvements in population coding) work in concert to improve perception. In the future, it will be interesting to see whether the effects of the two mechanisms on perception can be dissociated.

Normalization is perhaps the most widespread description of a neural computation that spans many model systems. Our study expands current models to show that normalization can capture the way that sensory information is transformed between cortical areas. The success of this model in accounting for correlated variability and manipulations using electrical microstimulation, and using these observations to narrow down potential mechanisms, bodes well for future work using recordings from groups of neurons to probe the neural mechanisms underlying canonical neural computations.

Methods

Electrophysiological Recordings. The data in this manuscript come from two distinct experiments, and a subset of these data have been presented previously (8, 21) and outlined in more detail below. The subjects in the paired V1–MT recording experiments (Figs. 4 A–D, 5, 6A, and 7A) were two adult male rhesus monkeys (*Macaca mulatta*, 8 and 9 kg), and the subject in the microstimulation experiments (Figs. 4 E and F and 7B) was the 8-kg monkey. All animal procedures were approved by the Institutional Animal Care and Use Committees of the University of Pittsburgh and Carnegie Mellon University. Before behavioral training, we implanted each animal with a titanium head post. After the animal learned the behavioral task (Fig. 2A, described below), we implanted a 10 × 10 chronic microelectrode array (Blackrock Microsystems) in V1 and a recording chamber that gave us access to area MT. The V1 array was connected to a percutaneous connector that allowed simultaneous recordings from 96 electrodes. The distance between adjacent electrodes on the array was 400 μm, and each electrode was 1 mm long. We identified area V1 using stereotactic coordinates and by visually inspecting the sulci.

The data used to fit the two models (Figs. 4 A–D and 5) and concerning the effects of attention on the responses of V1 neurons and on the correlations between V1 and MT come from recordings from the V1 array in which we simultaneously recorded from a single electrode (Fred Haer Corporation, Inc.) in MT. We identified MT using the gray and white matter transitions and the characteristic receptive field properties of MT units. Different analyses of these data were presented previously (8, 21).

The microstimulation data (Figs. 4 E and F and 7B; stimulation details are below) come from a separate set of experiments in which we electrically stimulated from the V1 array while recording from a movable multielectrode

probe (Alpha Omega or Plexon) in area MT. These experiments were performed in the one animal whose V1 array lasted long enough for us to attempt these additional experiments. Different aspects of data from these experiments (using a subset of the analysis time windows used here, using only high-contrast visual stimuli, and using very different data analysis methods) were presented previously (8).

We performed these experiments during daily experimental sessions for several months in each animal. We included experimental sessions for analysis when the MT unit's receptive field largely overlapped the envelope of the receptive fields of the units we recorded on the V1 array (Fig. 2B) and when the animal completed at least 150 behavioral trials (mean for the joint V1–MT recording sessions is 648 completed trials from 32 recording sessions; 12 from monkey 1 and 20 from monkey 2; mean for the microstimulation experiments is 519 completed trials from 16 experimental sessions). In the joint V1–MT experiments, we optimized the direction and speed of the visual stimuli for the tuning properties of the MT unit.

Visual Stimuli and Behavior. Our methods for presenting visual stimuli and monitoring behavior have been described elsewhere (8, 21). Briefly, we presented visual stimuli using custom software (written in Matlab using the Psychophysics Toolbox) (44, 45) on a cathode ray tube (CRT) monitor (calibrated to linearize intensity; 1024 × 768 pixels; 120 Hz refresh rate) placed 57 cm from the animal. We monitored eye position using an infrared eye tracker (Eyelink 1000; SR Research) and recorded eye position and pupil diameter (1,000 samples/s), neuronal responses (30,000 samples/s), and the signal from a photodiode to align neuronal responses to stimulus presentation times (30,000 samples/s) using hardware from Ripple.

We trained the animals to perform a motion direction change detection task (Fig. 2A) that required them to shift attention between three possible locations. A trial began when the monkey fixated a small spot within a 1° square fixation window in the center of the video display. The visual stimuli were achromatic Gabors whose size and location were picked so that two stimuli lay within the receptive field of the single MT unit under study (Fig. 2B) or in the joint receptive field of the group of MT units for the microstimulation experiments. The visual stimuli flashed on for 200 ms and off for a randomized interval (200–400 ms between each stimulus presentation, picked from a uniform distribution). During blocks of trials when attention was directed to one of the two stimuli within the MT receptive fields, the stimuli were either both 8% contrast (referred to as low-contrast stimulus presentations) or both 100% contrast (high contrast), and contrast was randomly interleaved on each stimulus presentation. The stimuli drifted at the same speed, which was selected from a range between 6 and 12° per second (picked to elicit large responses in the MT units), the two stimuli within the receptive field of the MT unit moved in opposite directions (the preferred and null directions of the MT cell under study), and which of the two stimuli moved in the preferred direction varied randomly from trial to trial.

In separate blocks of trials, the animals were instructed to direct their attention to a third stimulus in the opposite hemifield ("attend opposite" blocks). In these trials, the two stimuli in the MT receptive field were independently presented at either 0%, 50%, or 100% contrast. The third, attended, stimulus was presented at either 8% or 100% contrast, with its contrast randomly interleaved on each stimulus presentation and independently selected from the contrasts of the stimuli in the MT receptive field. When the third stimulus was present, it moved in an orthogonal direction to those in the opposite hemifield. We used responses from all attention conditions to fit the parameters of the normalization and linear models.

After an unsigned number of stimulus presentations picked from an exponential distribution (minimum 2 stimulus presentations, mean 6 stimuli, maximum 14 stimuli), the direction of one of the stimuli changed. During each experimental session we selected a single magnitude of the direction change designed to get the animal to perform near psychophysical threshold (range: 10–45°). The probability of direction change was independent at each location (the unattended stimuli each changed on ~12% of trials). Before the start of each block of trials, the monkey performed 5–10 instruction trials (which were not included in any of the analyses) in which there was only a single stimulus. The location of this stimulus constituted a cue as to the attended location. In the upcoming block of trials, if the attended stimulus was the one that changed, the monkey was given a liquid reward for making a saccade to that stimulus within 500 ms of the change. To account for saccadic latency and to avoid rewarding the monkey for guessing, the monkey was rewarded only for saccades beginning at least 100 ms after the change. If no change occurred within the maximum 14 stimulus presentations, the monkey was rewarded simply for maintaining fixation. Attention was cued to one of the three stimulus locations in blocks of 50–100 trials. The monkey was never rewarded for making a saccade to distractor changes.

Overall, the monkeys correctly detected the stimulus changes on 66% of completed trials during the 32 experimental sessions where we recorded neuronal data from both V1 and MT. When alternating attention between two stimuli placed within the receptive field of the MT unit, the monkeys detected 90% of the full-contrast direction changes and 42% of the low-contrast direction changes that occurred at the attended location. The monkeys responded to 21% of the unattended orientation changes, but were not rewarded for these responses. During blocks of trials when the animals were instructed to direct attention to the hemifield opposite the joint receptive fields of the V1 and MT units ("attend away" conditions, which were much easier for the animals because the distractors were far away from the attended stimuli), they detected direction changes at full contrast 99% of the time and responded to direction changes at the unattended location 5% of the time.

Normalization Model of MT Responses. We showed previously that a simple extension of a classic normalization model can largely account for the stimulus and attention dependence of V1–MT correlations (21). We modeled the response R of an MT neuron to a combination of stimuli moving in its preferred direction (P) and null direction (N) as

$$R_{P,N} = \frac{s_P \sqrt{V_{1P}(C_P)} + s_N \sqrt{V_{1N}(C_N)}}{C_P + \alpha C_N + \sigma}, \quad [1]$$

where $R_{P,N}$ is the mean response of the MT neuron and $\sqrt{V_{1P}(C_P)}$ and $\sqrt{V_{1N}(C_N)}$ are the trial-averaged (measured) responses of the pools of simultaneously recorded V1 neurons whose receptive fields overlap the preferred or null stimuli whose contrast is given by C_P or C_N . The parameters s_P and s_N are scaling parameters that reflect the direction tuning of the MT neuron. The parameter α represents a tuned normalization parameter that accounts for the observation that MT neurons vary in the extent to which they exhibit normalization (4), and σ represents a semisaturation constant.

Attention is instantiated in the model using the scaling parameter β such that the response of the MT neuron is given by

$$R_{P,N} = \frac{\beta s_P \sqrt{V_{1P}(C_P)} + s_N \sqrt{V_{1N}(C_N)}}{\beta C_P + \alpha C_N + \sigma}, \quad [2]$$

when attention is directed to the stimulus moving in the MT neuron's preferred direction, and

$$R_{P,N} = \frac{s_P \sqrt{V_{1P}(C_P)} + \beta s_N \sqrt{V_{1N}(C_N)}}{C_P + \beta \alpha C_N + \sigma}, \quad [3]$$

when attention is directed to the stimulus moving in the null direction. In this model, attention acts to scale both the sensory evidence (numerator in Eqs. 2 and 3) and the relevant normalization terms (denominator in Eqs. 2 and 3). Together, Eqs. 1–3 have five free parameters (s_P , s_N , α , β , and σ).

The linear model (Fig. 4 C and D) is identical to the normalization model except that the denominators of Eqs. 1–3 are set to 1.

We fit the models to the trial-averaged responses of the MT unit in 10 stimulus and attention conditions, using the measured trial-averaged responses of the V1 units whose receptive fields overlapped each stimulus and therefore comprised a "pool." We then predicted the response of the MT unit on each trial using the fitted model parameters and the actual responses of the V1 units recorded on that trial. We used the predicted MT responses to calculate predicted spike count correlations between the modeled MT unit and the recorded V1 units.

We fit our model to each MT unit's responses to each stimulus configuration separately (e.g., preferred stimulus at location 1 and null at location 2, or the opposite). Our data set therefore consisted of 64 MT units/conditions (32 units recorded in separate sessions, each with two different stimulus configurations). None of our results failed to reach significance, and none of our conclusions differed if we restricted analysis to one stimulus configuration per unit. In contrast, we felt it was most conservative to consider both stimulus configurations to account for differences in the location of each stimulus within the MT unit's receptive field or the number of V1 units whose receptive field overlapped each stimulus.

We included a V1 unit for analysis, and assigned it to a pool if it responded significantly more to a full contrast stimulus at one location than the other (t test, $P < 0.01$) during the attend-opposite blocks. For example, we assigned a V1 unit to the preferred pool if it responded more to a visual stimulus at the location where the preferred stimulus was presented compared with the location of the null stimulus. The activity of pools of V1 units (used to calculate correlations and as inputs to the model in Fig. 4) was defined as the average response of all such V1 units. Across the 64 recording sessions/stimulus configurations, 3,262 V1 units

satisfied this criterion (1,631 unique units). The mean number of V1 units per pool from each recording session was 25.

Microstimulation Experiments. Our microstimulation procedure has been described previously (8). During the microstimulation experiments (Fig. 2 C and D), we measured the difference between the number of MT spikes elicited by a visual stimulus combined with electrical stimulation in V1 and the number of spikes elicited by a visual stimulus alone (Fig. 3). We stimulated through the V1 array using trains of 200-Hz biphasic pulses (pulse duration = 0.2 μ s). We increased the current (range: 30–65 μ A) and number of electrodes (range: 1–3 electrodes) until the V1 microstimulation produced a readily observable modulation of the responses of the simultaneously recorded MT units (Fig. 3). If microstimulation in V1 did not modulate the responses of the MT neurons, we moved the multielectrode probe in MT until we found a site whose responses were modulated by stimulating V1. We stimulated for a 50-ms period during randomly interleaved stimulus presentations in the direction change detection task. Microstimulation never occurred on the first visual stimulus presentation in a trial, and occurred with 50% probability on subsequent stimulus presentations. In different experiments, we began microstimulation at either 50 or 100 ms after stimulus onset (always for a 50-ms period). The effects of V1 microstimulation on MT responses did not qualitatively depend on when the stimulation began, so we combined all of the data.

We saw evidence of electrical artifacts in our V1 but not our MT recordings. In V1, we saw large artifacts despite fast-settle amplifiers and other hardware technologies (Ripple) designed to minimize artifacts from electrical stimulation during the microstimulation experiments. These artifacts were obvious because they were huge in voltage and were time-locked to the stimulation pulses. For that reason, we only report findings from data collected in MT that were collected using a separate front-end amplifier from the one that was used to stimulate. Because we used separate (isolated) amplifiers and different electrodes, we were able to record during stimulation in this separate area. We saw no evidence of the sort of artifacts that occurred in V1 in our MT recordings. We took several additional steps to be sure that our MT results were not contaminated by more subtle artifacts, and these steps have been described previously (8).

Data Analysis. All spike sorting was done offline manually using Offline Sorter (v3.3.2; Plexon Inc.). We based our analyses on both single units and multiunit clusters and use the term "unit" to refer to either. Using chronically implanted microarrays (as we used for our V1 recordings), it is nearly impossible to tell whether we recorded from the same single units or multiunit clusters on subsequent days. However, the MT units were recorded on different electrodes each day, so each MT unit or V1–MT pair was unique.

The analysis of the microstimulation data in Fig. 4 included spikes counted during a 50-ms window during microstimulation, shifted 1 ms after microstimulation onset. The analysis of the microstimulation data in Fig. 7B included spikes counted during a 100-ms window after microstimulation, shifted 1 ms after microstimulation stopped (see Fig. 3 for reference). Our other analyses of neuronal data are based on spike counts from 30 to 230 ms after stimulus onset for V1 and from 50 to 250 ms after stimulus onset for MT to account for the visual latencies of neurons in both areas. Using identical windows for both areas led to qualitatively similar results to those presented here.

We computed spike count correlations between pairs of units or between the averaged activity of pools of V1 units and the MT unit using a standard Pearson's correlation coefficient for each stimulus condition separately. Because this measure is sensitive to outliers, we excluded stimulus presentations on which either unit (or group of units) in the pair responded more than three SDs differently than its mean (according to the convention in ref. 46). Stimulus presentations where a microsaccade was detected anywhere between 10 ms before until 10 ms after the stimulus was shown were excluded from analysis. We identified microsaccades using a velocity detection algorithm (47).

Simulations. To assess the predicted changes in V1–MT correlations caused by changes within V1, we simulated the responses of pools of V1 neurons (Fig. 5). To begin, we calculated the mean and variance of the V1 pool during blocks of trials in which the monkey directed attention to the stimulus inside the MT receptive field that did not overlap the receptive fields of the V1 units in the pool (e.g., to assess predicted correlations between the MT unit and the pool of V1 neurons whose receptive fields overlap the preferred stimulus, we began with responses in the attend-null condition). For each trial, we drew the responses of the pool from a Gaussian distribution with the same mean and variance as the neurons in the pool.

We then simulated the effects of attention on V1 neurons. Attention is known to multiplicatively scale the mean rates of neurons in visual cortex (17–19), so we scaled the mean responses of neurons in the attended and unattended pools by multiplying their responses by scaling factors k and $1/k$, respectively (sampling

values of k in an unphysiologically large range 1/32–32). Attention is also associated with changes in the variance (22) and covariance (8–16) of visual neurons, both of which would affect the variance of the mean of the pool of neurons. We therefore simulated these changes by scaling the variance of the distributions of simulated responses of the attended and unattended pools by scaling factors ν and $1/\nu$, respectively (sampling values of ν in an unphysiologically large range 1/32–32; Fig. 5A).

To assess the predicted changes in V1–MT correlations caused by changes in the weighting of attended V1 inputs to MT (Fig. 5B), we changed the parameter β in Eqs. 2 and 3 in the range 1/32–32.

Other than this manipulation of the parameter β in Fig. 5B, all of our simulations (including those in Fig. 4) were based on the fitted parameters of the MT units we recorded concurrently with V1 responses. The median parameters

for the normalization model were: S_P : 3.7043; S_N : 0.43; α : 0.43; β : 1.96; and σ : 0.72. The median parameters for the linear model were: S_P : 1.74; S_N : 0.06; and β : 1.29 (the linear model does not have parameters α or σ).

ACKNOWLEDGMENTS. We thank David Montez for assistance with animal training and recordings; Karen McCracken, Joshua Alberts, and Jen Symmonds for technical assistance; and Amy Ni for comments on an earlier version of the manuscript. Support for this work was provided by NIH Grants 4R00EY020844–03 and R01 EY022930 (to M.R.C.); a training grant slot on NIH Grant 5T32NS7391–14 (to D.A.R.); a Whitehall Foundation Grant (to M.R.C.); a Klingenstein-Simons Fellowship (to M.R.C.); a grant from the Simons Foundation (to M.R.C.); a Sloan Research Fellowship (to M.R.C.); a McKnight Scholar Award (to M.R.C.); and NIH Core Grant P30 EY008098.

- Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. *Nat Rev Neurosci* 13:51–62.
- Lee J, Maunsell JHR (2009) A normalization model of attentional modulation of single unit responses. *PLoS One* 4:e4651.
- Boynton GM (2009) A framework for describing the effects of attention on visual responses. *Vision Res* 49:1129–1143.
- Ni AM, Ray S, Maunsell JHR (2012) Tuned normalization explains the size of attention modulations. *Neuron* 73:803–813.
- Verhoef BE, Maunsell JHR (2016) Attention operates uniformly throughout the classical receptive field and the surround. *Elife* 5:1–26.
- Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19:1736–1753.
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
- Ruff DA, Cohen MR (2016) Attention increases spike count correlations between visual cortical areas. *J Neurosci* 36:7523–7534.
- Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12:1594–1600.
- Cohen MR, Maunsell JHR (2011) Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron* 70:1192–1204.
- Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63:879–888.
- Zénon A, Krauzlis RJ (2012) Attention deficits without cortical neuronal deficits. *Nature* 489:434–437.
- Herrero JL, Gieselmann MA, Sanayei M, Thiele A (2013) Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron* 78:729–739.
- Gregoriou GG, Rossi AF, Ungerleider LG, Desimone R (2014) Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4. *Nat Neurosci* 17:1003–1011.
- Ruff DA, Cohen MR (2014) Attention can either increase or decrease spike count correlations in visual cortex. *Nat Neurosci* 17:1591–1597.
- Luo TZ, Maunsell JHR (2015) Neuronal modulations in visual cortex are associated with only one of multiple components of attention. *Neuron* 86:1182–1188.
- Yantis S, Serences JT (2003) Cortical mechanisms of space-based and object-based attentional control. *Curr Opin Neurobiol* 13:187–193.
- Maunsell JHR, Cook EP (2002) The role of attention in visual processing. *Philos Trans R Soc Lond B Biol Sci* 357:1063–1072.
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647.
- Womelsdorf T, Fries P (2007) The role of neuronal synchronization in selective attention. *Curr Opin Neurobiol* 17:154–160.
- Ruff DA, Cohen MR (2016) Stimulus dependence of correlated variability across cortical areas. *J Neurosci* 36:7546–7556.
- Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55:131–141.
- Briggs F, Mangun GR, Usrey WM (2013) Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. *Nature* 499:476–480.
- de la Rocha J, Doiron B, Shea-Brown E, Josić K, Reyes A (2007) Correlation between neural spike trains increases with firing rate. *Nature* 448:802–806.
- Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nat Neurosci* 14:811–819.
- Carandini M (2004) Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS Biol* 2:E264.
- Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9:1421–1431.
- Ruff DA, Alberts JJ, Cohen MR (2016) Relating normalization to neuronal populations across cortical areas. *J Neurophysiol* 116:1375–1386.
- Ohshiro T, Angelaki DE, DeAngelis GC (2011) A normalization model of multisensory integration. *Nat Neurosci* 14:775–782.
- Louie K, Khaw MW, Glimcher PW (2013) Normalization is a general neural mechanism for context-dependent decision making. *Proc Natl Acad Sci USA* 110:6139–6144.
- Nienborg H, Cumming BG (2006) Macaque V2 neurons, but not V1 neurons, show choice-related activity. *J Neurosci* 26:9567–9578.
- Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291:1560–1563.
- Bichot NP, Rossi AF, Desimone R (2005) Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308:529–534.
- Womelsdorf T, Fries P, Mitra PP, Desimone R (2006) Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* 439:733–736.
- Saalmann YB, Pigarev IN, Vidyasagar TR (2007) Neural mechanisms of visual attention: How top-down feedback highlights relevant locations. *Science* 316:1612–1615.
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320:110–113.
- Gregoriou GG, Gotts SJ, Zhou H, Desimone R (2009) High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* 324:1207–1210.
- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.
- Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J Neurosci* 18:3870–3896.
- Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7:358–366.
- Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. *J Neurosci* 31:14272–14283.
- Moreno-Bote R, et al. (2014) Information-limiting correlations. *Nat Neurosci* 17:1410–1417.
- Kohn A, Coen-Cagli R, Kanitscheider I, Pouget A (2016) Correlations and Neuronal Population Information. *Annu Rev Neurosci* 39:237–256.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* 10:437–442.
- Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J Neurosci* 25:3661–3673.
- Engbert R, Kliegl R (2003) Microsaccades uncover the orientation of covert attention. *Vision Res* 43:1035–1045.