Title: Simultaneous multi-area recordings suggest a novel hypothesis about how attention improves performance

Authors: Douglas A. Ruff* and Marlene R. Cohen

Affiliations: Department of Neuroscience and Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, Pennsylvania 15260; Correspondence to: ruffd@pitt.edu

Summary – Visual attention dramatically improves subjects' ability to see and also modulates the responses of neurons in every known visual and oculomotor area, but whether those modulations can account for perceptual improvements remains unclear. We measured the relationship between populations of visual neurons, oculomotor neurons, and behavior, which was inconsistent with all published hypotheses about how attention improves perception. Our results support the novel hypothesis that attention reshapes the visual representation of attended stimuli to more effectively influence the premotor neurons that guide decisions without explicitly changing the readout or decision process. Our results suggest a path toward understanding the neural underpinnings of perception and cognition in health and disease by analyzing neuronal responses in ways that are constrained by behavior and interactions between brain areas.

Keywords – *attention, decoding, population analyses*

Introduction –

Each of the huge number of psychophysical and physiological studies of visual attention show that attention profoundly affects subjects' perceptual abilities and also modulates the responses of populations of neurons at every stage of visual and oculomotor processing (Carrasco, 2011), Despite these oft replicated observations, whether any of the observed neuronal modulations can account for the improvements in psychophysical performance remains unknown. Three hypotheses have dominated the literature (Figure 1A): that attention 1) improves visual information coding (Cohen and Maunsell, 2009; Mitchell et al., 2009; Ruff and Cohen, 2014), 2) improves communication between visual and oculomotor brain areas (Fries, 2015; Gregoriou et al., 2009; Ruff and Cohen, 2016; Ruff and Cohen, 2017), or 3) reduces uncontrolled fluctuations in the animals' cognitive state, thereby improving performance and rendering observed effects of attention on response variability as epiphenomena (Ecker et al., 2016).

The studies used to support these hypotheses were limited by available data and analysis methods, which primarily involved the responses of single neurons, pairs of simultaneously recorded neurons in the same brain area, or coarse measures of population activity like local field potentials or measures of blood flow. We evaluated these hypotheses using the responses of groups of simultaneously recorded neurons in multiple stages of visuomotor processing, psychophysics, and data analysis methods that leverage that unique combination. We recorded simultaneously from groups of neurons in area MT, which encodes motion information (Born and Bradley, 2005) and the superior colliculus (SC), where neuronal responses are either visual, oculomotor, or intermediate (Gandhi and Katnani, 2011; Sparks, 1999; Wurtz and Albano, 1980) and have been hypothesized to be involved in computing perceptual decisions (Crapse et al., 2018; Horwitz and Newsome, 1999). When we analyzed the responses of single neurons or pairs of neurons, we replicated previous observations. However, constraining our analyses by the animals' behavior and the simultaneous recordings from both areas allowed us to reject each prior hypothesis.

We found that none of the previous hypotheses account for observed attention-related perceptual improvements. Instead, our results suggest a novel hypothesis: that the well-known effects of attention on firing rates and shared response variability in visual cortex reshape the representation of attended stimuli such that they more effectively drive downstream neurons and guide decisions without explicitly changing the weights relating sensory responses to downstream neurons or behavior (Figure 1B). Our results suggest that the long observed attention-related changes in the responses of visual cortical neurons account for perceptual improvements. More generally, this study shows that leveraging the ability to record simultaneously from neurons at different stages of neural processing and constraining analyses by the subjects' behavior can greatly clarify the relationship between many sensory, cognitive, and motor processes, neuronal responses and behavior.

Results –

We compared evidence for and against three hypothesized attention mechanisms using neuronal responses collected while two rhesus monkeys performed a widely studied motion direction change-detection task (Figure 1C; Cohen and Maunsell, 2009; Ni et al., 2018; Ruff and Cohen, 2016). The animals' performance was greatly affected (Figure 1D) by a cue instructing them to shift spatial attention between a stimulus within the same or opposite hemifield as the joint receptive fields of several dozen neurons that were recorded on multielectrode probes in MT (Figure 1E, red points) and the SC (blue points). MT and the SC represent different stages of perceptual decision-making and therefore provide the opportunity to evaluate each hypothesized attention mechanism. MT contributes to motion perception (Born and Bradley, 2005; Parker and Newsome, 1998). The SC is thought to play many roles in visually guided tasks (Gandhi and Katnani, 2011; Sparks, 1999; Wurtz and Albano, 1980), decision-making (Horwitz and Newsome, 1999; Horwitz and Newsome, 2001) and attention(Krauzlis et al., 2013).

Figure 1

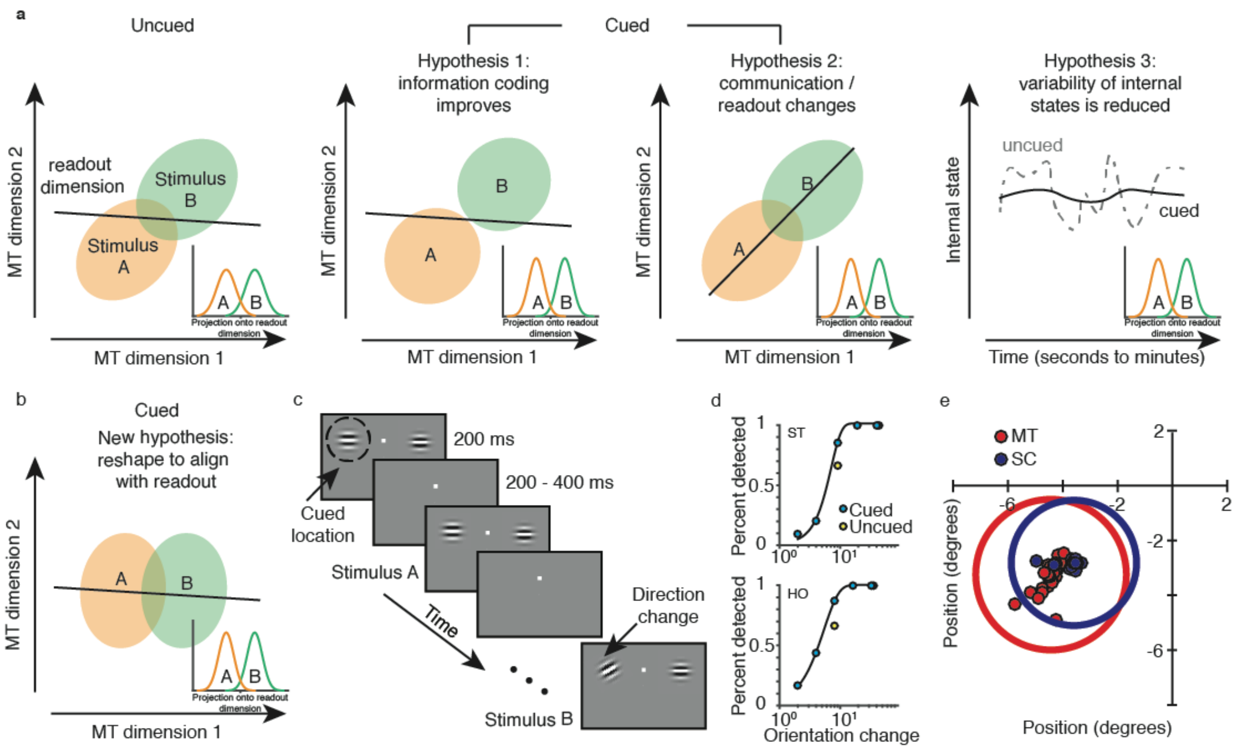


Figure 1. Hypotheses and methods. (A) Schematics describing predominant hypotheses about links between attention, visual cortical activity, and behavior. The left plot depicts MT population responses to two visual stimuli plotted along two dimensions in population response space (e.g. the first two principal components; see Methods) and a readout dimension which represents the visual information that is communicated to neuronal populations involved in planning behavior during the uncued condition. The insets depict projections of the population responses onto the readout dimension. Hypothesis 1 is that the MT representations of the two stimuli become more easily distinguishable (e.g. by separating the distributions of responses to the two stimuli). In this scenario, the distributions of projections along even a suboptimal readout axis may also be more separable. Hypothesis 2 suggests that attention changes the

communication or readout dimensions, such that projections of responses to the two stimuli are more separable. Hypothesis 3 proposes that attention-related changes in visual cortex are epiphenomenal. Instead, it suggests that attention reduces uncontrolled variability in the animals' internal states, which might produce less variable neuronal population responses and therefore more separable projections onto the readout dimensions. (B) Our new hypothesis: attention reshapes population responses so they are better aligned with relatively static readout dimensions. This alignment could be a direct result of widely observed attention-related changes in firing rates and response variability. (C) Direction change-detection task with cued attention. The stimuli before the change can be thought of as stimulus A and the changed stimulus can be thought of as stimulus B in the schematics in (A). (D) Psychometric curves from two example sessions (monkey ST, top, monkey HO, bottom) with best-fitting Weibull functions. Attention improved detection of median difficulty trials by 25% on average across all experiments. (E) Receptive field (RF) centers of recorded units from the same example session as in the top plot in (D). Dots represent the RF center (red, MT; blue, SC). The circle represents the size and location of the median RF from each area.

The three predominant attention hypotheses make different predictions about how attention should affect MT and the SC in our task. The first (information coding) hypothesis predicts that attention improves the motion direction information encoded in MT. The second (communication) hypothesis posits that attention improves the stimulus information that is shared between MT and the SC. The third (internal states) hypothesis suggests that attention reduces uncontrolled fluctuations in the animals' cognitive states, allowing them to focus on their psychophysical task. In this scenario, the widely observed attention-related reduction in response variability in visual cortex (Cohen and Maunsell, 2009; Mitchell et al., 2009; Ni et al., 2018) would be an epiphenomenon of decreased variability in cognitive states.

Population recordings replicate previously observed effects of attention

Past studies have evaluated these hypotheses by analyzing the responses of individual neurons or pairs of neurons, which typically lack the statistical power to reveal a strong link to behavior. Using our data set, we replicated the observations that have been used as evidence in favor of all three hypotheses. Consistent with previous studies evaluating the information coding hypothesis

(Maunsell, 2015; Reynolds and Chelazzi, 2004; Ruff et al., 2018), we found that attention increased the trial-averaged responses of neurons in both MT and the SC (Supplemental Figure 1A and B). Consistent with studies evaluating both the information coding and internal states hypotheses (Ecker et al., 2016), we found that attention decreased the extent to which the trial to trial fluctuations in neuronal responses to repeated presentations of the same stimulus are shared between pairs of MT neurons(Cohen and Maunsell, 2009; Mitchell et al., 2009; Ni et al., 2018) (quantified as the average spike count or noise correlation, or r_{SC}(Cohen and Kohn, 2011); Supplementary Figure 1C). Consistent with studies evaluating the communication hypothesis, attention increases correlated variability between the two areas (Ruff and Cohen, 2016; Ruff and Cohen, 2017) (Supplemental Figure 1C). This attention-related increase was weakly dependent on the visual responsivity of SC neurons (Supplemental Figure 2).

Most previous studies aim to connect the observed attention-related changes in single neurons or pairs of neurons to behavior via an explicit or implicit model. Recordings from neuronal populations provide the opportunity to do so directly. We reasoned that analyzing the relationship between populations of simultaneously recorded neurons in multiple brain areas with the animals' behavior would provide insight into the relative importance of each hypothesized mechanism.

Interactions between brain areas do not support the internal states hypothesis

The internal states hypothesis provides a challenge for the approach of linking populations of visual neurons with behavior because it proposes that there is no link between attention-related changes in MT or other visual areas and performance. Instead, it posits that attention limits mind wandering, and the changes in performance and neuronal responses simply reflect improved stability in internal states (Ecker et al., 2016). There are two reasons this hypothesis seems unlikely to account for our data. First, the effects of spatial attention are spatially specific (e.g. correlated variability increases in one hemisphere while decreasing in the other, even when neurons in the two hemispheres are simultaneously recorded(Cohen and Maunsell, 2009)), meaning that reductions in the variability of global cognitive processes like arousal and motivation are unlikely to account for the attention-related changes in visual cortex. Further, it is not obvious how reductions in fluctuations in internal states could account for the attention-related increases in firing rates observed in spatial attention studies like ours (Supplementary Figure 1), let alone the more complex firing rate changes associated with feature attention (Treue and Martinez-Trujillo, 1999).

This hypothesis can also be addressed using a population-analysis approach, by using the responses of MT and SC neurons to attempt to quantify the variability in the animals' internal states. We reasoned that fluctuations in internal states would 1) often occur at timescales longer than the 400-600 ms between stimulus presentations in our task and 2) affect the covariability of neurons in many brain areas, including both MT and the SC. We performed principal components analysis on population responses to the identical visual stimuli that occurred before the direction change on each trial (e.g. stimulus A in Figure 1; see Methods) in each attention condition. Because the only variability in those population responses is internally generated, the first principal component (PC) represents the axis of greatest shared variability in the population of neurons in each brain area.

Figure 2

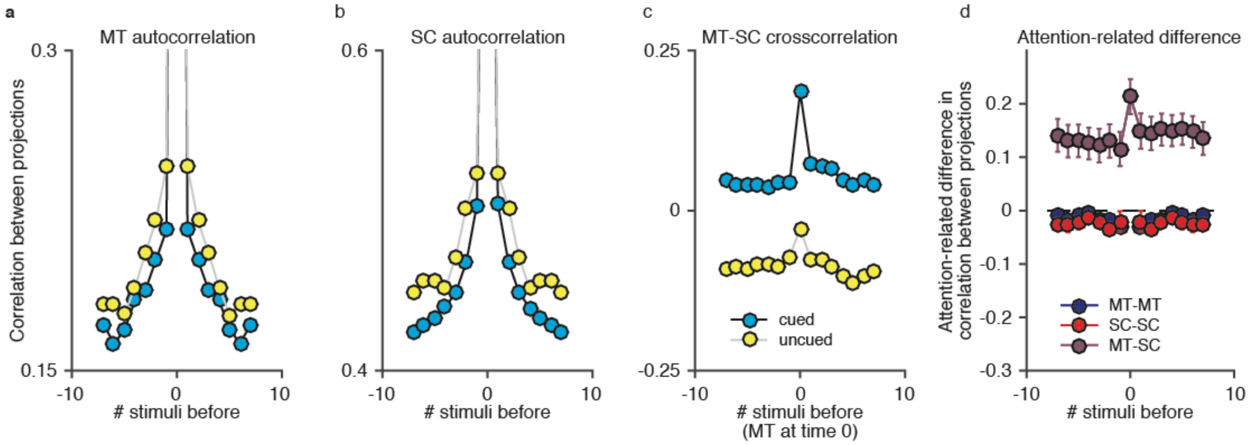


Figure 2. Attention has opposite effects on slow fluctuations in neuronal population responses within and across areas. (A, B) Autocorrelations between projections onto the first PCs of population responses to repeated presentations of the same visual stimulus in (A) MT, and (B) the SC. The x-axis plots time lag in units of stimulus presentations (400-600 ms; see Methods). (C) Cross correlation between projections onto the first PCs in MT and the SC (same data and plotting conventions as in A and B). (D) Attention-related difference in autocorrelation or cross correlations between the projections in the previous plots. Error bars represent standard error of the mean. Attention was associated with a statistically significant decrease in autocorrelation overall (t-tests, p<.05) in both areas and in 11/15 individual MT data sets and 9/15 SC data sets (t-tests, p<0.05 with a Bonferroni correction) and a significant increase in cross correlation overall (t-test, p<.001) and in 11/15 individual data sets.

The autocorrelation functions of projections onto the first PC show that there is indeed response variability in each area that fluctuates slowly and is reduced by attention (Figures 2A, B), which is not as readily observable by computing noise correlations between pairs of neurons (Supplementary Figure 1C). However, the cross-correlation of projections onto the first PC in MT and the SC showed a qualitatively different time course than the autocorrelation function. Furthermore, not only did attention not reduce the covariability of these signals between these two areas (and presumably brain wide), attention *increased* the slow variability that is shared between areas (Figure 2C and Supplementary Figure 1C). These results are in conflict with the idea that the attention-related decrease in covariability within each area is a byproduct of a decrease in uncontrolled fluctuations in internal states, because such a decrease should be brain-wide.

Neuronal population decoding methods fail to support the information coding or communication hypotheses

Figure 3

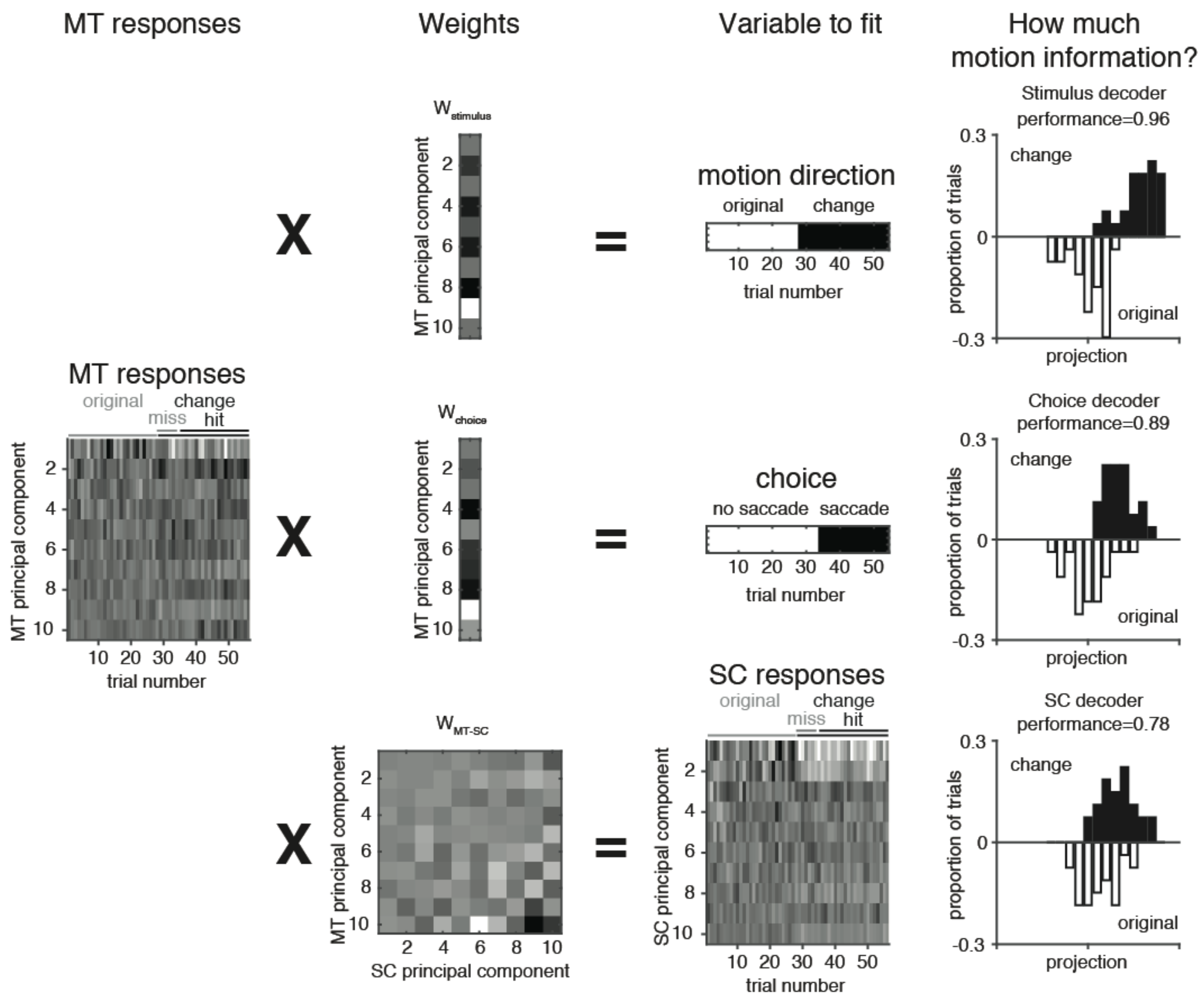


Figure 3 – Schematic of our decoding procedure. We used linear regression to find the weights (second column) that best relate the first ten principal components of the MT population's response *(left) to the visual stimulus (Stimulus decoder; top row), the animal's choice (Choice decoder; middle row), or the projections of the responses of the population of simultaneously recorded SC neurons (SC decoder; bottom row). We assessed the performance of each decoder by decoding stimulus information from MT responses on a separate set of trials using each set of weights (right column). See methods for detailed decoding and cross validation procedures.*

To investigate the remaining hypotheses, we determined whether attention affects the amount of stimulus information that can be decoded from the population of MT neurons using linear decoders that are optimized to a) dissociate between the original and changed stimuli (Stimulus decoder in Figure 3), b) predict the animals' choices (whether or not they made an eye movement; Choice decoder), or c) predict the activity of the population of SC neurons we recorded (SC decoder).

The information coding hypothesis posits that attention improves the stimulus information that could be gleaned by an optimal stimulus decoder. However, attention did not significantly affect the performance of an optimal decoder in our data set, even when we used a decoder optimized separately for each attention condition (Figure 4A, left bars). Furthermore, theoretical work suggests that the effects of attention on the stimulus information that can be decoded from small neuronal populations are likely to be even more minimal for larger populations (Kanitscheider et al., 2015; Kohn et al., 2016; Moreno-Bote et al., 2014).

The communication hypothesis posits that attention improves communication between visual and decision areas, meaning that it should change the weights relating MT responses to either behavior or SC responses. We found that attention had larger effects on the stimulus information that is related to the animals' choices on individual trials (Figure 4A, middle bars) or that is shared with the SC (Figure 4A, right bars). However, this could arise from either a weight change (Figure 1A) or a change within MT, to better align MT responses with static readout weights (Figure 1B).

Figure 4

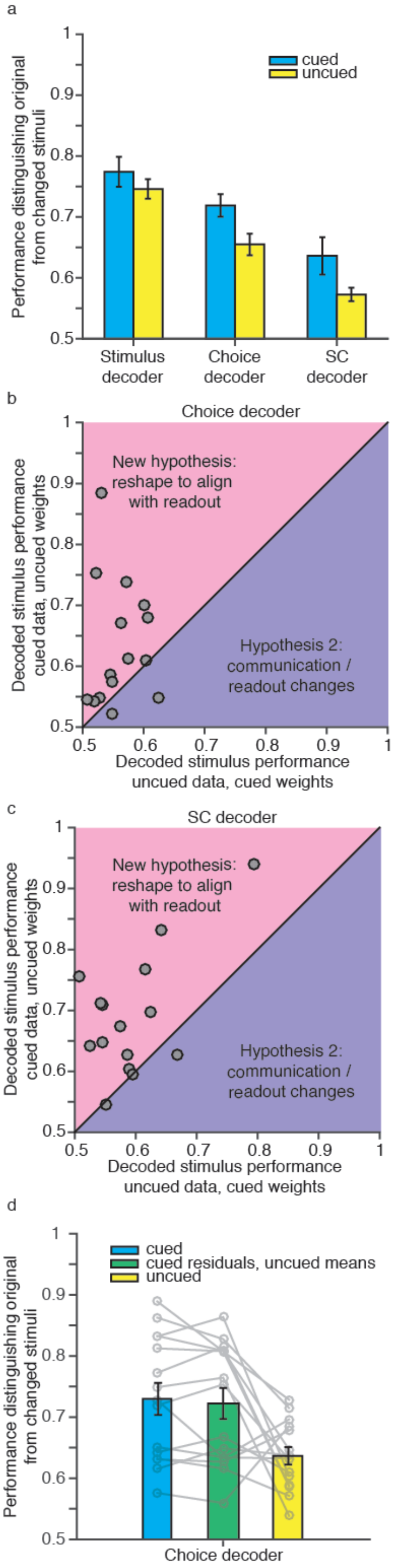


Figure 4. Effects of attention on the stimulus information that can be decoded from small populations of MT neurons. (A) Ability of a cross-validated linear decoder to distinguish the original from changed stimuli for each decoder. Error bars represent SEM. The effect of attention was significant for the Choice and SC decoders (paired t-tests, p<.05) but not for the Stimulus decoder (p=0.28). The effects of attention on the Choice and SC decoders were greater than for the stimulus decoder (paired t-tests, p<.05), but not significantly different from each other (p=0.21). (B) Weight swapping analysis demonstrates that decoding performance was typically better using the MT responses from the cued condition and the choice decoder weights from the uncued condition (y-axis) than using the MT responses from the uncued condition and the Choice decoder weights from the cued condition (x-axis; paired t-test, p<.05). (C) Same, using the weights from the SC decoder (paired t-test, p<.05). (D) The reshaping of the MT representation of the attended stimulus can be accomplished as a result of attention-related changes in response variability (e.g. noise correlations). The amount of stimulus information that can be decoded using a single choice decoder whose weights are determined from data from both attention conditions is indistinguishable for the cued data and data constructed using the mean responses from the uncued condition and the residuals from the cued condition (paired t-test, p=0.84).

A new hypothesis: attention reshapes sensory activity so that it more effectively guides decisions

We found no evidence that attention changes weights relating MT responses to SC responses or behavior. Both the Choice and SC decoders gleaned more stimulus information from MT responses in the attended than unattended condition, even when we used the weights computed in the opposite attention condition for which they were calculated (Figures 4B and 4C). Together, these neuronal population analyses that use the animals' behavior and the activity of downstream neurons to assess the hypothesized attention mechanisms reveal that none of the three previously hypothesized mechanisms provides a satisfactory account of the observed attention-related behavioral improvement.

Our data support the hypothesis that attention reshapes the representation of attended stimuli to more effectively guide behavior (Figure 1B). In this scenario, the critical changes are in visual cortex. However, this does not result in a large improvement in the stimulus information that can be gleaned by an optimal stimulus decoder. Instead, the modulated neuronal activity in MT better aligns with the readout dimensions using relatively static weights.

How could a reshaping of the representation of an attended stimulus be implemented? The simplest mechanism would make use of the signatures of attention that have been observed for many years (e.g., changes in firing rate gain or noise correlations; Maunsell, 2015). We investigated the possibility that these simple response changes can account for the attention-related improvement in the stimulus information decoded using the choice decoder in two stages. First, to verify the prediction of the weight-swapping analyses (Figures 4B and 4C), we constructed a single choice decoder for both attention conditions and determined that it captured the attention-related improvement in decoded stimulus information (compare the blue and yellow bars in Figure 4D). Second, we used those same weights to decode stimulus information from population responses constructed using the mean rates from the uncued condition but the residuals from the cued condition. We found that by simply using residuals (which incorporate both response variability that is private to each neuron and that which is shared between neurons) from the cued condition was enough to completely account for the attention-related improvement in decoded stimulus information (Figure 4D).

Discussion –

We measured population spiking activity in MT and the SC while monkeys performed a difficult visual task in which we manipulated attention. We used these responses and the monkey's behavior to test existing hypotheses about how attention improves perception. Our data did not support any of these hypotheses.

Instead, our data are consistent with the novel hypothesis that attention reshapes population activity so that it better aligns with the dimensions in population space along which visual information is read out to guide behavior. We decoded the visual information that is most closely related to the animals' choices or to the activity of neurons in the superior colliculus. These analyses support the idea that attention changes the activity of visual neurons so that the population activity that aligns with relatively static readout dimensions is more informative about the visual stimulus.

The idea of realigning sensory information seems like it would require much more exotic mechanisms than the other hypothesized attentional mechanisms. However, we showed that previously observed effects of attention on neuronal response variability were sufficient to reshape the representation of attended stimuli so that they more effectively influence behavior. Further, we showed recently that the covariability of a population of neurons can be readily changed by altering the balance of inhibition to excitation (Huang et al., 2017). It may be that changing covariability realigns sensory responses, and that this change is in fact mechanistically simple compared with mechanisms required to improve the information encoded in a population of neurons or changing communication between areas.

The idea that changing correlated variability better aligns sensory responses to a fixed readout is also consistent with our recent observation that in this task, monkeys' choices are well-aligned with the axis in population space that explains the most correlated noise (Ni et al., 2018). One exciting possibility is that the correlated variability axis represents the fixed readout dimension, perhaps because it is well-positioned to decode the motion direction of the broad set of stimuli that animals encounter outside the limited environment of most laboratory tasks (Ruff et al., 2018). If so, reducing noise correlations and increasing firing rate gains would improve the stimulus information projected along that readout axis (following the intuitions in (Averbeck et al., 2006)).

By combining psychophysics and multi-neuron, multi-area recordings, our study provides a framework for understanding the neuronal mechanisms underlying a wide variety of sensory, cognitive, and motor processes. For example, attention-related changes in gains and correlations have been observed for many years. The key insight that they may improve performance by aligning the responses of populations of visual neurons to fixed readout dimensions was only possible using population analyses constrained by the animals' behavior rather than focusing on one or two neurons.

In the future, it will be interesting to use the same approach to determine whether a similar mechanism can account for behavioral changes associated with other cognitive processes (e.g. feature attention) that might seem more likely to change the weights relating stimulus information to downstream neurons or behavior. Further, many neuropsychiatric disorders (including disorders of attention, Autism, and schizophrenia) are thought to involve changes in the same computations thought to underlie attention (Carandini and Heeger, 2012). An exciting possibility is that these changes might be identified and potential therapies evaluated in animal models using the combination of behavioral evaluation and multi-neuron, multi-area recordings that we described here.

Acknowledgements
We are grateful to Karen McKracken for providing technical assistance as well as Faisal Baqai, Matt Getz, Jay Hennig, Chengcheng Huang, Amy Ni and John Maunsell who provided comments on an earlier version of this manuscript.

Author Contributions

Both authors conceived and designed the experiments, analyzed the data and wrote the manuscript. D.A.R. collected the data.

Competing interests

The authors declare no competing interests.

References:

Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nature Reviews Neuroscience *7*, 358-366.

Born, R.T., and Bradley, D.C. (2005). Structure and function of visual area MT. Annual review of neuroscience *28*, 157-189.

Brainard, D.H. (1997). The Psychophysics Toolbox. Spatial vision *10*, 433-436.

Carandini, M., and Heeger, D.J. (2012). Normalization as a canonical neural computation. Nature Reviews Neuroscience *13*, 51-62.

Carrasco, M. (2011). Visual attention: the past 25 years. Vision Res *51*, 1484-1525.

Cohen, M.R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. Nature Neuroscience *14*, 811-819.

Cohen, M.R., and Maunsell, J.H.R. (2009). Attention improves performance primarily by reducing interneuronal correlations. Nature Neuroscience *12*, 1594-1600.

Crapse, T.B., Lau, H., and Basso, M.A. (2018). A Role for the Superior Colliculus in Decision Criteria. Neuron *97*, 181-194 e186.

Ecker, A.S., Denfield, G.H., Bethge, M., and Tolias, A.S. (2016). On the Structure of Neuronal Population Activity under Fluctuations in Attentional State. J Neurosci *36*, 1775-1789.

Engbert, R., and Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. Vision Research *43*, 1035-1045.

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

Gandhi, N.J., and Katnani, H.A. (2011). motor functions of the superior colliculus. Annual Review of Neuroscience *34*, 205-231.

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

Horwitz, G.D., and Newsome, W.T. (1999). Separate Signals for Target Selection and Movement Specification in the Superior Colliculus. Science *284*, 1158-1161.

Horwitz, G.D., and Newsome, W.T. (2001). Target selection for saccadic eye movements: prelude activity in the superior colliculus during a direction-discrimination task. Journal of neurophysiology *86*, 2543-2558.

Huang, C., Ruff, D.A., Pyle, R., Rosenbaum, R., Cohen, M.R., and Doiron, B. (2017). Circuit models of low dimensional shared variability in cortical networks. bioRxiv *217976*.

Kanitscheider, I., Coen-Cagli, R., and Pouget, A. (2015). Origin of information-limiting noise correlations. Proceedings of the National Academy of Sciences of the United States of America *112*, E6973-6982.

Kohn, A., Coen-cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and Neuronal Population Information. Annual Review of Neuroscience *39*, 237-256.

Krauzlis, R.J., Lovejoy, L.P., and Zénon, A. (2013). Superior Colliculus and Visual Spatial Attention. Annual Review of Neuroscience *36*, 165-182.

Maunsell, J.H.R. (2015). Neuronal Mechanisms of Visual Attention. Annu Rev Vis Sci *1*, 373-391.

Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron *63*, 879-888.

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nature Neuroscience *17*, 1410-1417.

Ni, A.M., Ruff, D.A., Alberts, J.J., Symmonds, J., and Cohen, M.R. (2018). Learning and attention reveal a general relationship between population activity and behavior. Science *359*, 463-465.

Parker, A.J., and Newsome, W.T. (1998). Sense and the single neuron: probing the physiology of perception. Annual review of neuroscience *21*, 227-277.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial vision *10*, 437-442.

Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. Annual Review of Neuroscience *27*, 611-647.

Ruff, D.A., and Cohen, M.R. (2014). Attention can either increase or decrease spike count correlations in visual cortex. Nature Neuroscience *17*, 1591-1597.

Ruff, D.A., and Cohen, M.R. (2016). Attention Increases Spike Count Correlations between Visual Cortical Areas. Journal of Neuroscience *36*, 7523-7534.

Ruff, D.A., and Cohen, M.R. (2017). A normalization model suggests that attention changes the weighting of inputs between visual areas. Proc Natl Acad Sci U S A *114*, E4085-E4094.

Ruff, D.A., Ni, A.M., and Cohen, M.R. (2018). Cognition as a Window into Neuronal Population Space. Annu Rev Neurosci *41*, 77-97.

Sparks, D.L. (1999). Conceptual issues related to the role of the superior colliculus in the control of gaze. Current opinion in neurobiology *9*, 698-707.

Treue, S., and Martinez-Trujillo, J.C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. Nature *399*, 575-579.

Wurtz, R.H., and Albano, J.E. (1980). Visual-motor function of the primate superior colliculus. Annual review of neuroscience *3*, 189-226.

Materials and Methods

Methods Summary

The subjects were two adult male rhesus monkeys (*Macaca mulatta*, 8 and 9 kg). All animal procedures were approved by the Institutional Animal Care and Use Committees of the University of Pittsburgh and Carnegie Mellon University. Using linear 24 channel moveable probes (Plexon), we simultaneously recorded extracellular activity from direction-selective neurons in area MT and neurons in the superior colliculus that responded either visually, prior to a saccade, or both. Before beginning the experiment, we searched for neurons in both areas that had overlapping spatial receptive fields (Figure 1E) as determined by mapping with both drifting gratings and a delayed saccade task. The monkeys performed a direction change-detection task that commenced upon fixation of a central spot (Figure 1C). Two drifting Gabor stimuli, whose direction was selected to drive the recorded population of MT neurons well, flashed on and off until the direction of one stimulus changed at a random, unsignaled time. The monkeys signaled detection of the change by making a saccade to the changed stimulus within 450 ms of its onset. The location of the stimulus change was cued using instruction trials prior to each block of trials and the cue was valid 80% of the time. On the other 20% of trials, the change happened at the uncued location. In order to encourage fixation on longer trials, catch trials, where no stimulus changed direction and monkeys were rewarded for maintaining fixation, were intermixed. We analyzed spiking activity in response to all visual stimuli except the first stimulus in each trial.

Materials and Methods

The subjects were two adult male rhesus monkeys (*Macaca mulatta*, 8 and 9 kg). All animal procedures were approved by the Institutional Animal Care and Use Committees of the University of Pittsburgh and Carnegie Mellon University.

We presented visual stimuli using custom software (written in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) on a CRT monitor (calibrated to linearize intensity; 1024 768 pixels; 120 Hz refresh rate) placed 54 cm from the animal. We monitored eye position using an infrared eye tracker (Eyelink 1000; SR Research) and recorded eye position and pupil diameter (1000 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.

Behavioral Task

As previously described (Cohen and Maunsell, 2009), a trial began when the monkey fixated a small, central spot within a 1.25° per side, square fixation window in the center of a video display while two peripheral Gabor stimuli (one overlapping the receptive fields of the recorded neurons, the other in the opposite visual hemifield) synchronously flashed on (for 200 ms) and off (for a randomized period between 200-400 ms) until, at a random, unsignaled time, the direction of one of the stimuli changed from that of the preceding stimuli (Figure 1C). The monkey received a liquid reward for making a saccade to the stimulus that changed. Attention was cued in blocks of 50-100 trials, and alternated between blocks where attention was cued to either the left or the right stimulus. In each block, the direction change occurred at the cued stimulus on 80% of trials, and at the uncued stimulus in 20% of trials (all uncued changes used either the middle or largest orientation change, Figure 1D). Catch trials, where no stimulus changed direction and the monkey was rewarded for maintaining fixation, were randomly intermixed throughout each block. Psychometric data were fit with Weibull functions. Before recording commenced, the monkeys were extensively trained to have stable thresholds across a range of spatial locations (3-6 months). During recording, the size, location, spatial frequency and direction of the Gabor stimuli were selected to drive the recorded units well and the direction of all of the stimuli prior to the orientation change was constant throughout a recording session.

Electrophysiological Recordings

We simultaneously recorded extracellularly from single units and sorted multiunit clusters (the term “unit” refers to either). The dataset consisted of a total of 306 responsive MT units and 345 responsive SC units total (36-58 units per session, mean 20 in MT, 24 in the SC for Monkey HO; 36-53 units per session, mean 21 in MT, 22 in SC for Monkey ST) in both MT and the SC in the right hemisphere using moveable, linear 24-channel V-probes (Plexon; inter-electrode spacing in MT = 50μm, SC = 100μm). We presented visual stimuli and tracked eye position as previously described (Ruff and Cohen, 2016). The data presented are from 6 days of recording for Monkey HO and 9 days of recording for Monkey ST. Each day consisted of multiple blocks of the attention task (Figure 1C; mean 1015 of trials for Monkey HO, 745 for Monkey ST) preceded by receptive field mapping using a delayed saccade task and direction tuning during passive fixation.

Data Analysis

All spike sorting was done offline manually using Offline Sorter (version 3.3.5; Plexon). We based our analyses on both single units and multiunit clusters and use the term "unit" to refer to either. All neuronal analyses prior to Figure 3 used spike count responses between 50-250 ms after stimulus onset to account for visual latencies in the two areas. To remove response contamination from eye movements during change stimuli, data presented in the decoding analyses in Figure 3 and 4 used shorter response windows. Responses to both unchanged and changed stimuli were measured from 50-185 ms after stimulus onset for monkey HO and 50-220 ms for monkey ST. These times were selected based on the distribution of each animal's reaction times, with the goal of maximizing the number of trials that could be included in the analyses. Trials with reaction times that began during those windows were excluded. To minimize the impact of adaptation on our results, we did not analyze the first stimulus presentation in each trial. We only analyzed a recorded MT unit if its stimulus-driven firing rate was 10% higher than its firing rate as measured in the 100 ms prior to the onset of the first stimulus. We only analyzed a recorded SC unit if its stimulus-driven firing rate was 10% higher than its firing rate as measured in the 100 ms prior to the onset of the first stimulus or if its response during a 100 ms epoch prior to a saccade on correct trials to the contralateral side was 10% larger than that same baseline. Stimulus presentations during which a microsaccade was detected were excluded from analyses (Engbert and Kliegl, 2003; Ruff and Cohen, 2016)).

For firing rate analyses in Supplemental Figure 1A and B, attention indices were calculated using average spike counts on the stimulus presentation prior to correct detections of the intermediate change amount depending on whether attention was directed into or out of the receptive fields of the recorded neurons using the formula $(\text{attend}_{in} - \text{attend}_{out})/(\text{attend}_{in} + \text{attend}_{out})$. Significance of individual units was determined by a paired t-test ($p<0.05$).

Noise correlations

We defined the correlated variability of each pair of simultaneously recorded units (quantified as spike count correlation or r_{SC} (Cohen and Kohn, 2011)) as the Pearson correlation coefficient between the responses of the two units to repeated presentations of the same stimulus. This measure of r_{SC} represents noise correlations rather than signal correlations because the responses used in this analysis were always to an identical visual stimulus. For Supplemental Figure 1C, we included responses from stimulus presentations 2 though 10 from trials that ended with either a hit, miss or correct catch trial and that were immediately followed by the maintenance of fixation and continuation of the trial. We z-scored responses as a function of the stimulus presentation number in each trial and then pooled data across stimulus presentations before calculating noise correlations. Results did not qualitatively change if we did not perform this z-score procedure. For Supplemental Figure 1D, we included data from all stimulus presentations prior to the change stimulus (except the first) and sorted them depending on what the behavioral outcome was on the subsequent stimulus presentation. Pairs of units that were recorded on the same electrode were not included in correlation analyses. The data presented in Supplementary Figures 1C consisted of 3,285 MT pairs, 3,948 SC pairs and 6,934 between area pairs.

Slow timescale auto- and cross-correlations

To test the hypothesis that attention affects uncontrolled fluctuations in internal states, we created a procedure to identify slow fluctuations in population responses. We plotted the responses of the populations of simultaneously recorded MT or SC neurons to the stimuli before the motion direction change (excluding the first stimulus) in a high dimensional space in which each neuron's response comprises one axis. We performed principal components analysis (PCA) on this cloud of points. Because the visual stimuli in this subset of the data were all identical, the first principal component (PC) represents the axis that captures the most variance in non-stimulus related population responses. We searched for slow fluctuations in internal states by measuring the auto- and cross-correlations in projections onto this first PC in each area.

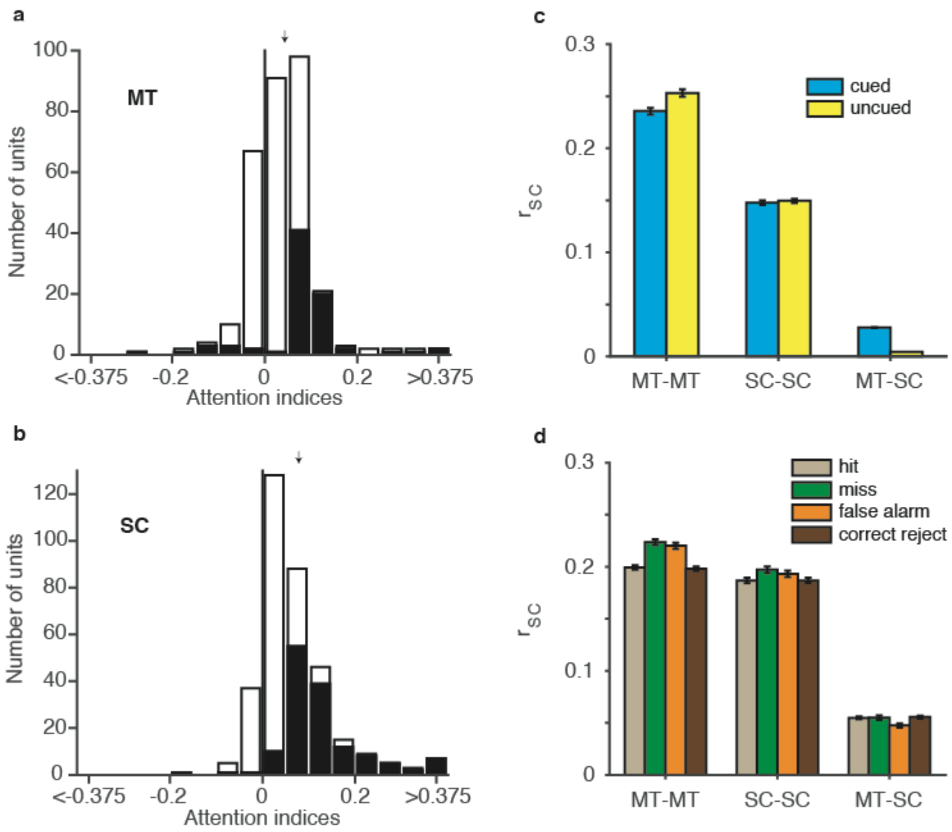
Decoding

We focused our decoding analyses on trials in which the third largest (middle) direction change occurred, because changes of that magnitude occurred in both attention conditions. We used the decoding strategy schematized in Figure 3. We began by constructing four matrices in each attention condition: 'MT responses' (a # MT neurons x 2*# trials matrix of MT responses to the stimuli before the direction change and the changed stimulus on the relevant trials), 'motion direction' (a 1 x 2*# trials vector of zeros for the stimulus before the change, referred to as 'original', and ones for the changed stimulus, referred to as 'change'), 'choice' (a 1 x 2*# trials vector of zeros for stimulus presentations on which the animal did not make an eye movement, referred to as 'no saccade', and ones when the animal made an eye movement, referred to as 'saccade'), and 'SC responses' (a # SC neurons x 2*# trials matrix of SC responses to the stimuli before the direction change and the changed stimulus on the relevant trials).

We cross validated by holding out the two stimulus presentations (for the original and changed stimuli) from one trial at a time to perform the rest of our analyses. To reduce the number of weights we needed to fit and therefore improve our confidence in the weights we did fit, we performed PCA on the MT and SC responses to find the first 10 PCs in each area. We then performed linear regression to find the weight vectors (for the Stimulus and Choice decoders) or weight matrices (for the SC decoder) that related projections along the first ten MT PCs plus a vector of ones to ‘motion direction’, ‘choice’, or projections along the first 10 SC PCs in each attention condition.

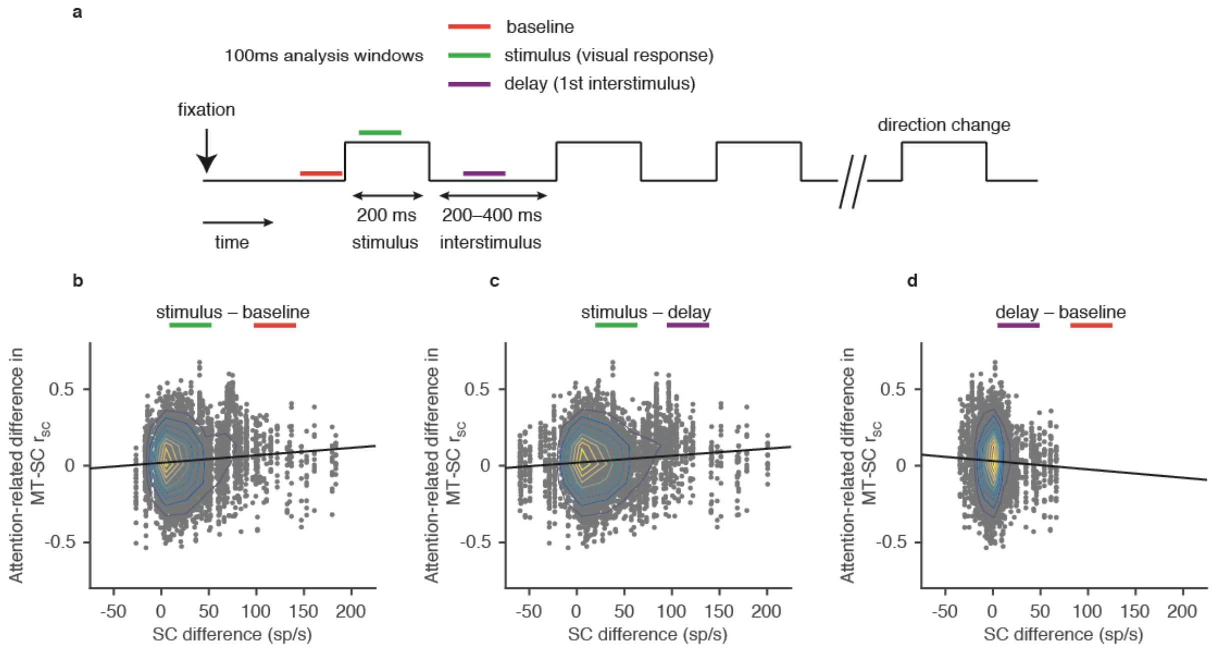
We assessed the stimulus information in each decoder (Figure 4) by multiplying projections of MT responses to the original and changed stimuli from the held-out trial by the fitted weights and either determining whether those weighted sums correctly classified the stimuli as original or changed (Stimulus and Choice decoders) or whether a linear classifier correctly classified those stimulus presentations on the basis of the predicted SC responses (SC decoder). For the decoding analysis in Figure 4D, we took a similar approach to the previously described choice decoder, except that we combined data from both the cued and uncued conditions to calculate decoding weights. We then decomposed the responses of the population responses to each stimulus in each attention conditions into mean responses and residuals (R=M+S, where R is the number of neurons x number of trials matrix of spike count responses to one stimulus in one attention condition, M is a matrix of mean responses for each neuron, and S is the matrix of residuals). We tested the hypothesis that attention-related changes in the residuals account for the improvement in stimulus information used to guide behavior by decoding stimulus information from responses created by using the mean responses from the uncued condition and residuals from the cued condition.

Supplemental Figure 1



Supplemental Figure 1. Effects of attention on common analyses of individual units and pairs of units (A) Attention increases firing rates in MT, quantified as the difference in firing rates in the different attention conditions divided by the sum. Units with significant differences in average responses for the two conditions are specified by black bars (t-test, p<.05). This distribution (mean = 0.04, median = 0.04) is significantly different from zero (Wilcoxon signed rank test, $p < 10^{-21}$*). (B) Same as A, for SC data. This distribution (mean = 0.073, median = 0.05) is significantly different from zero (Wilcoxon signed rank test,* $p < 10^{-43}$*). (C) Within and between area noise correlations calculated from spike counts during stimulus presentations that preceded successful maintenance of fixation from trials that ended with either a hit or miss or were a successful catch trial. Attention decreases average correlations within MT (Wilcoxon signed rank test,* $p < 10^{-12}$*), not in the SC (Wilcoxon signed rank test, p=0.8) and increases them between the two areas (Wilcoxon signed rank test,* $p < 10^{-41}$*). Error bars are standard error of the mean. (D) Within and between area noise correlations calculated from spike counts that immediately preceded different behavioral outcomes during cued trials. Misses and false alarms are associated with higher correlations within MT (t-test,* $p<10^{-3}$*) and SC (t-test,* $p<10^{-3}$*) but not between the two areas (t-test, p=0.23). Error bars are standard error of the mean.*

Supplemental Figure 2



Supplemental Figure 2. Relationship between SC responses during different task epochs and attention-related correlation changes with MT. (A) Schematic of task timing depicts the three 100ms epochs used to count spikes in SC units. The baseline period began 100ms before the first stimulus appeared, which is after stable fixation had been acquired. The stimulus period was shifted 30 ms after the appearance of the visual stimulus, to account for the earliest visual latencies observed in the SC. The delay period began 100ms after the first stimulus turned off and always ended prior to the onset of the second stimulus. (B) Attention-related changes in MT-SC r_{SC} *plotted against the difference between each SC unit's response during the stimulus and baseline periods. There are multiple MT-SC correlation differences measured for each SC unit. Correlations between MT and SC were calculated using the same data and methods as Supplementary Figure 1C (Pearson correlation, rho=0.087,* $p< 10^{-12}$*). Isolines depicting the decile boundaries are overlaid over the individual data points. (C) Similar to B, but data are now sorted by the difference between each SC unit's response during the stimulus and delay periods (Pearson correlation, rho=0.092,* $p< 10^{-14}$*). (D) Similar to B, but data are now sorted by the difference between each SC unit's response during the delay and baseline periods (Pearson correlation, rho=-0.042,* $p< 10^{-4}$*).*