nature neuroscience

ARTICLES
https://doi.org/10.1038/s41593-019-0477-1

Simultaneous multi-area recordings suggest that attention improves performance by reshaping stimulus representations

Douglas A. Ruff* and Marlene R. Cohen

Visual attention dramatically improves individuals' ability to see and modulates the responses of neurons in every known visual and oculomotor area, but whether such modulations can account for perceptual improvements is unclear. We measured the relationship between populations of visual neurons, oculomotor neurons and behavior during detection and discrimination tasks. We found that neither of the two prominent hypothesized neuronal mechanisms underlying attention (which concern changes in information coding and the way sensory information is read out) provide a satisfying account of the observed behavioral improvements. Instead, our results are more consistent with the hypothesis that attention reshapes the representation of attended stimuli to more effectively influence behavior. 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.

Each of the large number of psychophysical and physiological studies of visual attention show that attention profoundly affects individuals' perceptual abilities and also modulates the responses of populations of neurons at every stage of visual and oculomotor processing[1–4]. Despite these often replicated observations, whether any of the observed neuronal modulations can account for the improvements in psychophysical performance is unknown. Two nonmutually exclusive hypotheses have dominated the literature (Fig. 1a): (1) that attention improves visual information coding[5–7]; or (2) that it improves the efficiency with which visual information is read out by the premotor neurons involved in decision-making[8–11]. The studies used to support these hypotheses were limited by the available data and analysis methods, which primarily involved the responses of single neurons or pairs of simultaneously recorded neurons in the same brain area. 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 the middle temporal area (MT or V5), which encodes motion information[12,13], and the superior colliculus (SC), where neuronal responses are either visual, oculomotor or intermediate, contribute to gaze control[14–16] and are involved in computing perceptual decisions[17–19]. When we analyzed the responses of single or pairs of neurons, we replicated previous observations, including the results from two of our previous studies, which focused on visual area V4 in two different tasks with spatial attention components: (1) an orientation change detection task[5]; and (2) a contrast discrimination task[6]. However, constraining the analyses of our MT dataset or of both V4 datasets by the animals' behavior and the simultaneous recordings from both areas made it clear that neither prior hypothesis constitutes a satisfying account of the observed attention-related improvements in performance.

Our results suggest that on the timescale of perceptual decisions, across two visual areas and during both detection and discrimination tasks, spatial attention does not act primarily by improving information coding or by changing the way visual information is read out. Instead, the long-observed attention-related changes in the responses of visual cortical neurons account for perceptual improvements, but they do so by reshaping the representation of attended stimuli such that they more effectively drive downstream neurons and guide behavior (Fig. 1b). Our study provides a framework for leveraging multi-neuron, multi-area recordings and controlled psychophysics to study how neuronal networks mediate flexible behavior in many systems, timescales and tasks.

Results

We compared evidence for and against two hypothesized attention mechanisms using neuronal responses collected while two rhesus monkeys performed the widely studied motion direction change detection task in Fig. 1c (refs. [5,9,20–22]), and then compared the results to recordings while monkeys performed a similar orientation change detection task[5] and a contrast discrimination task[6]. As in the two previously published datasets, the animals' performance in our new experiment was greatly affected (Fig. 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 (Fig. 1e, red points) and SC (blue points). MT and SC represent different stages of perceptual decision-making and therefore provide the opportunity to evaluate each hypothesized attention mechanism. MT contributes to motion perception[12,13]. SC is thought to play many roles in visually guided tasks including gaze control[14–16], decision-making[17–19] and attention[4].

Population recordings replicate previously observed effects of attention. The two predominant attention hypotheses make different predictions about how attention should affect MT and SC in our task. The first (information coding) hypothesis predicts that attention improves the motion direction information encoded in MT. The second (readout) hypothesis posits that attention changes the way that stimulus information is read out of MT to influence

Department of Neuroscience and Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, PA, USA. *e-mail: ruffd@pitt.edu

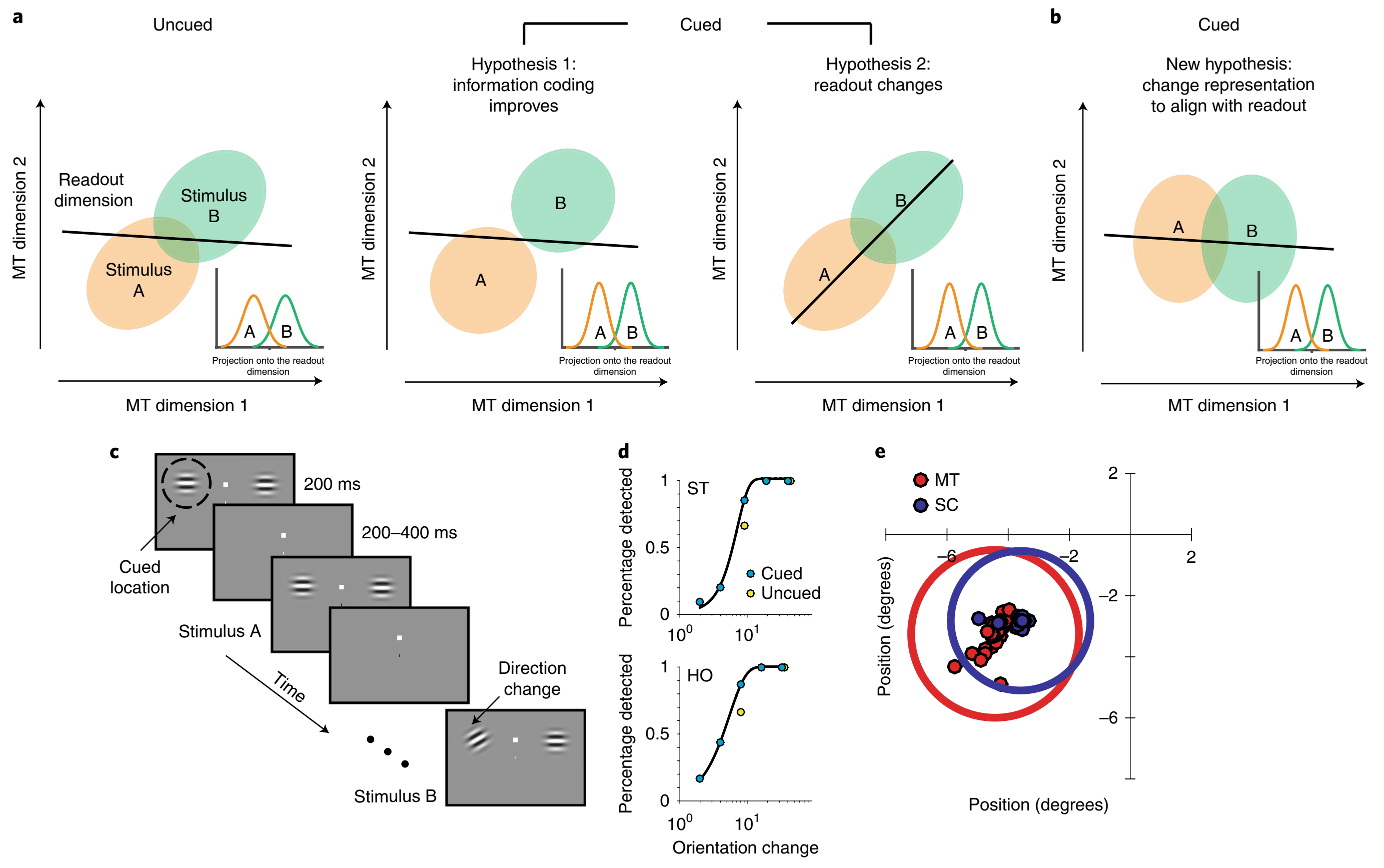

Fig. 1 | Hypotheses and methods. a, Schematics describing the predominant hypotheses about links between attention, visual cortical activity and behavior. The left plot depicts the population responses recorded from MT to two visual stimuli plotted along two dimensions in population response space (for example, the first two principal components; see Methods) and a readout dimension that represents the visual information 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 (for example, by separating the distributions of responses to the two stimuli). In this scenario, the distributions of projections along even a suboptimal readout axis may be more separable. Hypothesis 2 suggests that attention changes the way visual information is read out from MT such that projections of MT population responses to the two stimuli onto the readout dimension are more separable. **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 drifting Gabor stimuli before the change were identical on every trial within an experimental session and can be thought of as stimulus A, while the changed stimulus can be thought of as stimulus B 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 (cued 76.5% detected across sessions, uncued 51.8% detected; $n=15$ sessions, two-tailed Wilcoxon signed-rank test, $P=1.8\times10^{-4}$). **e**, Receptive field centers of recorded units from the same example session as in the top plot in **d**. The dots represent the receptive field center (red, MT; blue, SC). The circle represents the size and location of the median receptive field from each area.

downstream responses and ultimately behavior. Our strategy was to show that our data are consistent with those in past studies by replicating the results that have been used as evidence to support each hypothesis and then evaluate each hypothesis using analyses that leverage our simultaneous measurements from the animals' behavior and multi-neuron, multi-area recordings.

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 dataset, we replicated the observations that have been used as evidence in favor of each hypothesis. Consistent with previous studies evaluating the information coding hypothesis[2,3,23], we found that attention increased the trial-averaged responses of neurons in both MT and SC (Supplementary Fig. 1a,b) and 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[5,7,21] (quantified as the average spike count or noise correlation (r_{SC})[24]; Supplementary Fig. 1c). Consistent with studies evaluating the readout hypothesis, attention-related increases correlated variability between the two areas[9,10,25] (Supplementary Fig. 1c). These attention-related increases were weakly dependent on the visual responsivity of SC neurons (Supplementary Fig. 2).

The observed increase in correlations between areas suggests that attention-related effects are not simply due to global reductions in slow fluctuations, which has recently been hypothesized to explain attention-related correlation decreases within a single brain area[26,27] (Supplementary Fig. 3). On the face of it, this hypothesis does not seem to account for the spatially specific effects of spatial attention (for example, correlated variability increases in one hemisphere while decreasing in the other, even when neurons in the two hemispheres are simultaneously recorded[5]), meaning that reductions in the variability of global cognitive processes like arousal and motivation probably do not account for the

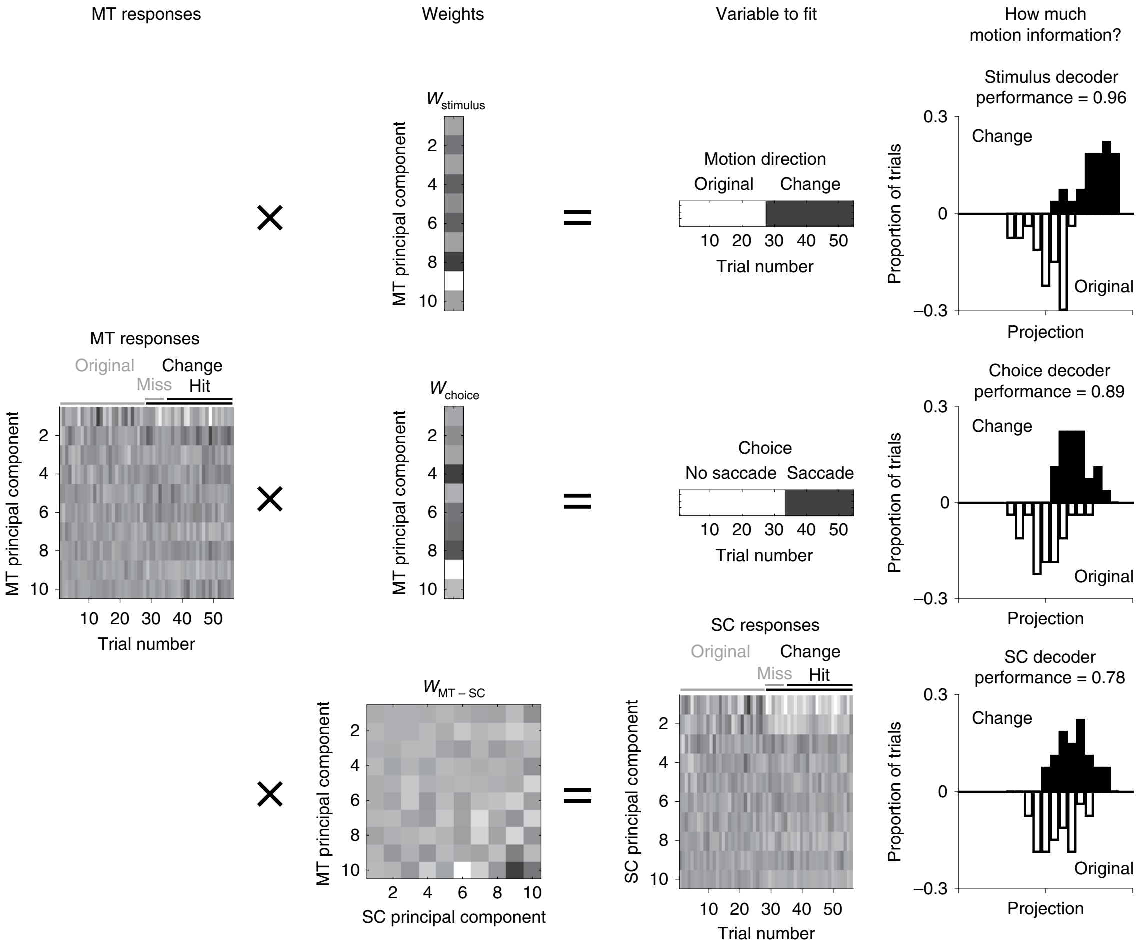

Fig. 2 | Schematic of our decoding procedure. We used linear regression to find the weights (second column) that best relate the first ten principal components of the population response in MT (left) to both the original and change stimuli (stimulus decoder; top row), the animal's choice in response to change stimuli (choice decoder; middle row) or the projections of the responses to the original stimulus 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) and responses to both the original and change stimuli. See Methods for detailed decoding and cross-validation procedures.

attention-related changes in visual cortex. In addition to the observation that attention has opposite effects on noise correlations between pairs of neurons in the same or different areas, we found that attention has opposite effects on the local dynamics of the population responses within MT or SC as it does on the interactions between the two areas (Supplementary Fig. 3c,d and Supplementary Fig. 1c). These results are in conflict with the idea that the attention-related decrease in covariability within each area is a by-product of a decrease in uncontrolled fluctuations in internal states because such a decrease should, presumably, be brain-wide.

Neuronal population decoding methods provide incomplete support for the information coding or readout hypotheses. 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. To this end, we determined whether attention affects the amount of stimulus information that can be decoded from the population of MT neurons using cross-validated linear decoders that are optimized to (1) dissociate between the original and changed stimuli ('stimulus' decoder in Fig. 2), (2) predict the animals' choices (whether or not they made an eye movement in response to change stimuli; 'choice' decoder), or (3) predict the activity of the population of SC neurons we recorded (using responses to the original stimulus; 'SC' decoder). These decoders were always constructed using data from trials with the intermediate change amount (see Fig. 1d).

The information coding hypothesis posits that attention improves the stimulus information that could be gleaned by an optimal stimulus decoder, but our data provided only weak support for this idea. Attention did not significantly affect the performance of an optimal decoder in our dataset, even when we used a decoder optimized separately for each attention condition (Fig. 3a, left bars). Recent theoretical work has demonstrated that high-dimensional decoders can ignore pairwise correlations that are orthogonal to the decoding axis and that correlations are more likely to be orthogonal to this axis in larger populations[28–30]. This suggests that the effects of attention on the stimulus information that can be decoded from small neuronal populations like the ones we recorded are probably even more minimal for larger populations, making it improbable that attention-related improvements in information coding account for the robust improvements in behavioral performance that we observed.

The readout hypothesis posits that attention changes the importance of the attended stimulus in guiding behavior by changing the way its representation is read out by the neurons involved in computing decisions. Therefore, this hypothesis posits that attention 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 (Fig. 3a, middle bars) or that is shared with the SC (Fig. 3a, right bars) than it did on the stimulus decoder. However, this difference could arise from either a weight change (Fig. 1a) or a change within MT that results in more stimulus-related visual information being projected onto a static readout dimension (Fig. 1b).

A new hypothesis: attention reshapes sensory activity so that it more effectively guides decisions. Our data do not support the hypothesis that attention changes weights relating MT to SC responses or behavior. Because the responses of MT neurons are correlated and because the behavioral readout is binary, the weights obtained by each decoder are nonunique, making it impossible to identify weight changes by analyzing the weights themselves[23,31]. However, we can infer their stability by measuring the stimulus information gleaned by each decoder using weights from the opposite attention condition from which they were calculated (see Methods). Both the choice and SC decoders gleaned more stimulus information from MT responses in the attended than unattended condition when we used the weights computed in the opposite attention condition (Fig. 3b,c). Together, these neuronal population analyses that use the animals' behavior and the activity of downstream neurons to assess the hypothesized attention mechanisms reveal that neither the information coding nor the readout hypothesis provide a satisfactory account of the large observed attention-related behavioral improvement.

Our observations suggest that in MT neurons recorded while monkeys are performing a change detection task, attention acts primarily by changing the visual information that is used to guide behavior using relatively fixed readout weights. To investigate the generality of these observations to different visual areas and different tasks, we tested these hypotheses using two additional datasets. In the first dataset, monkeys performed an orientation change detection task similar to the direction change detection task described in this study while we recorded from populations of V4 neurons[5]. Similar to our results in MT, we found that attention had larger effects on the stimulus information that is related to the animals' choices (choice decoder; Fig. 4a) than it did on the stimulus information that could be gleaned using an optimal (stimulus) decoder (Fig. 4b). As in our MT dataset (Fig. 3b), the results from this dataset suggest that attention typically reshapes V4 responses to align with relatively fixed readout mechanisms: decoding performance was typically better using the V4 responses from the cued condition and the choice decoder weights from the uncued condition (y axis) than using the V4 responses from the uncued condition and the choice decoder weights from the cued condition.

In the second new dataset, we searched for attention-related changes in information coding in V4 neurons while monkeys performed a discrimination task[6]. These data provide a particularly important test of the information coding hypothesis because unlike in the change detection task where attention has fairly uniform effects on V4 and MT neurons (increasing rates and decreasing noise correlations), we showed that in our discrimination task, attention can flexibly increase or decrease noise correlations in a way that is broadly consistent with improving information coding.

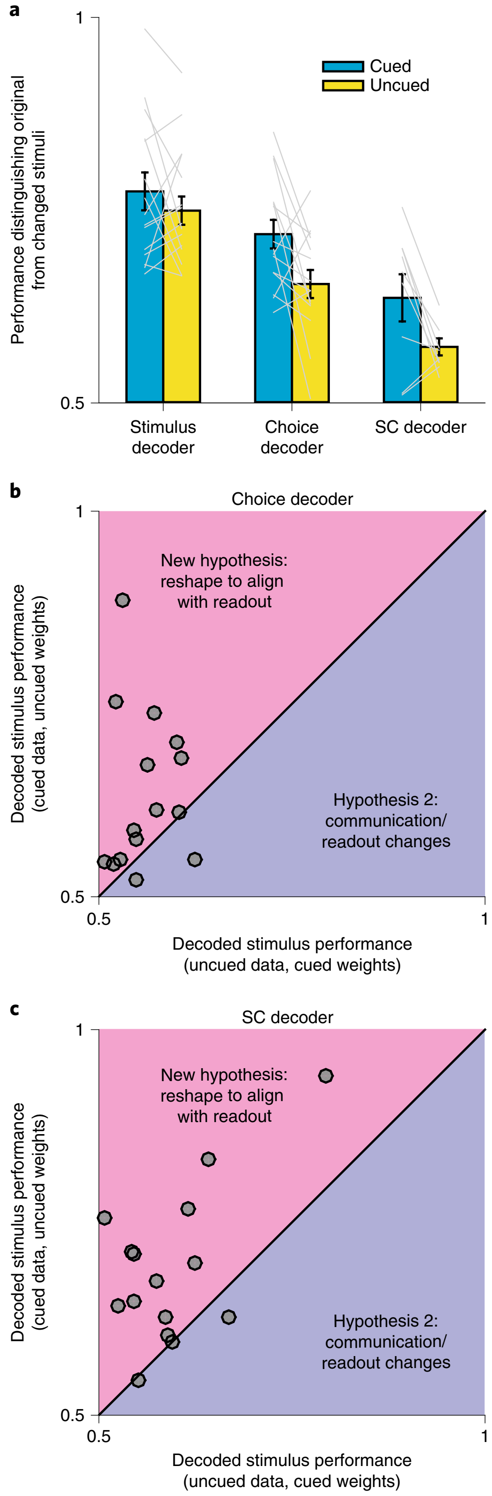

Fig. 3 | Effects of attention on the stimulus information that can be decoded from small populations of MT neurons. a, Ability of a cross-validated linear decoder to distinguish the original from changed stimuli (intermediate change amount) for each decoder. The error bars represent the s.e.m.; the gray lines are the individual sessions. The effect of attention was significant for the choice and SC decoders ($n = 15$ sessions; two-tailed paired t-tests, $P = 0.019$ and $P = 0.048$, respectively) but not for the stimulus decoder ($n = 15$ sessions; two-tailed paired t-test, $P = 0.42$). The effects of attention on the choice and SC decoders were greater than for the stimulus decoder ($n = 15$ sessions; two-tailed paired t-tests, $P = 0.023$ and $P = 0.030$, respectively), but not significantly different from each other ($n = 15$ sessions; two-tailed paired t-test, $P = 0.21$). **b**, Weight-swapping analysis demonstrates that decoding performance was typically better using MT responses from the cued condition and the choice decoder weights from the uncued condition (y axis) than using MT responses from the uncued condition and the choice decoder weights from the cued condition (x axis; $n = 15$ sessions; two-tailed paired t-test, $P = 0.005$). **c**, Same as **b**, using the weights from the SC decoder ($n = 15$ sessions; two-tailed paired t-test, $P = 0.012$).

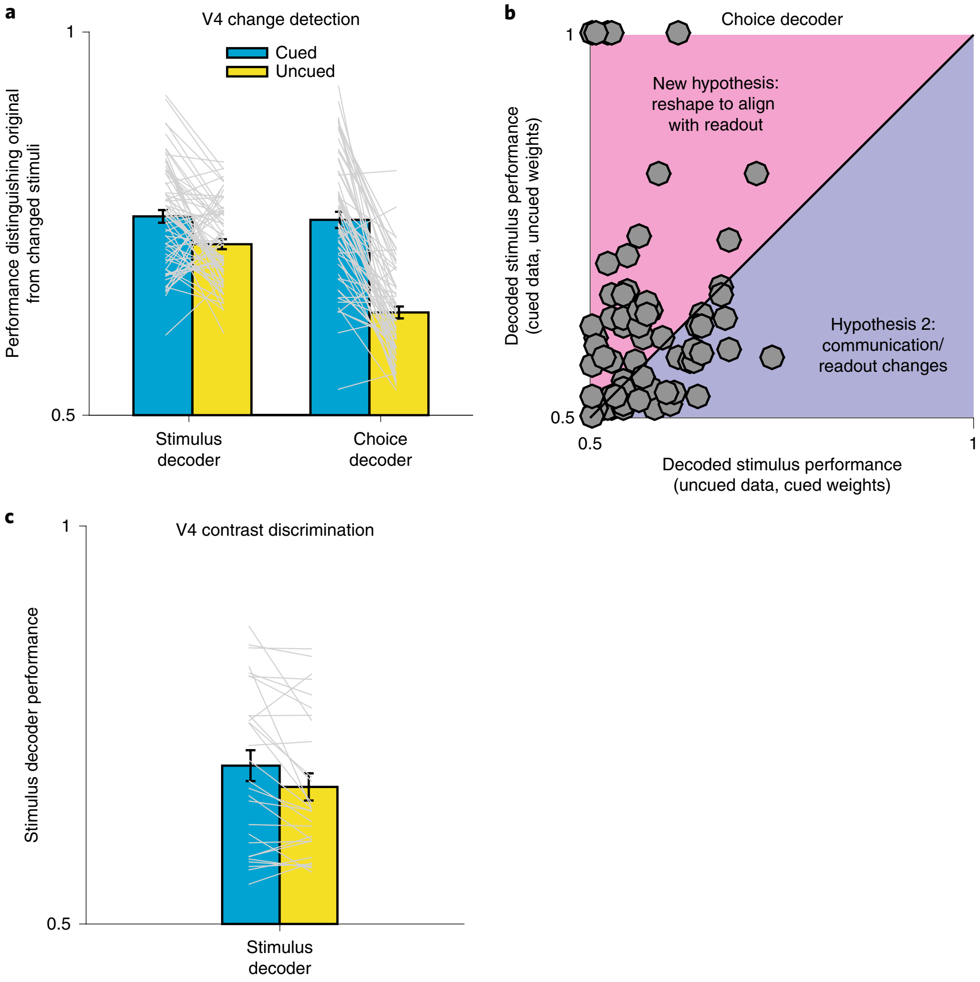

Fig. 4 | Similar attention-related effects on neuronal populations in two brain areas and two tasks. a, In a change detection task, the effects of attention on the stimulus information that can be decoded from small populations of V4 neurons is similar to that in MT. The plot shows the ability of a cross-validated linear decoder to distinguish the original from changed stimuli (intermediate change amount) for both the stimulus and choice decoders (no SC data were available). The error bars represent the s.e.m.; the gray lines are the individual hemisphere sessions (see Methods). Attention significantly affected the performance of both the stimulus and choice decoders ($n=98$ sessions; two-tailed paired t-test, $P=7.4\times10^{-4}$ and $P=1.1\times10^{-15}$, respectively), but the attention-related improvement in the choice decoder was greater than in the stimulus decoder ($n=98$ sessions; two-tailed paired t-test, $P=2.1\times10^{-8}$). **b**, Decoding performance was typically better using the V4 responses from the cued condition and the choice decoder weights from the uncued condition (y axis) than using the V4 responses from the uncued condition and the choice decoder weights from the cued condition (x axis; $n=98$ sessions; two-tailed paired t-test, $P=0.0029$; compare to Fig. 3b). **c**, The ability of a cross-validated linear decoder using V4 population responses to distinguish between stimulus configurations during a contrast discrimination task[6] reveals no significant effect of attention ($n=17$ sessions; two-tailed paired t-test, $P=0.31$). Plotting conventions as in **a**. Because of the details of the discrimination task, which did not include choices related to uncued stimuli, it was impossible to calculate a choice decoder using these data.

Despite these findings, the results of our decoding analyses were similar for the detection and discrimination tasks, meaning that we did not find strong evidence that attention improves the amount of stimulus information that can be optimally extracted from a population of visual neurons in either task (Fig. 4c). Together, these results provide evidence that in multiple visual areas and visually guided tasks, attention acts primarily to reshape population activity so that more stimulus information is used to guide behavior using relatively fixed decision mechanisms.

Our data support the hypothesis that attention reshapes the representation of attended stimuli to more effectively guide behavior (Fig. 1b). In this scenario, the critical changes are in visual cortex. However, this reshaping 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 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 often observed signatures of attention such as changes in firing rate gain[2,3,23] or pairwise noise correlations[5–7,9,20–22,32–37]. We investigated the possibility that these simple response changes can account for the attention-related improvement in the stimulus information decoded using both the choice and SC decoders in two stages. First, to verify the prediction of the weight-swapping analyses (Fig. 3b,c), we constructed a single choice decoder for both attention conditions (Fig. 5a) and determined that it captured the attention-related improvement in decoded stimulus information (compare the blue and yellow bars in Fig. 5b). Second, we used those same weights to decode stimulus information from population responses constructed using the mean rates from the uncued condition and the residuals from the cued condition (green bar). We found that simply using residuals from the cued condition, which incorporate

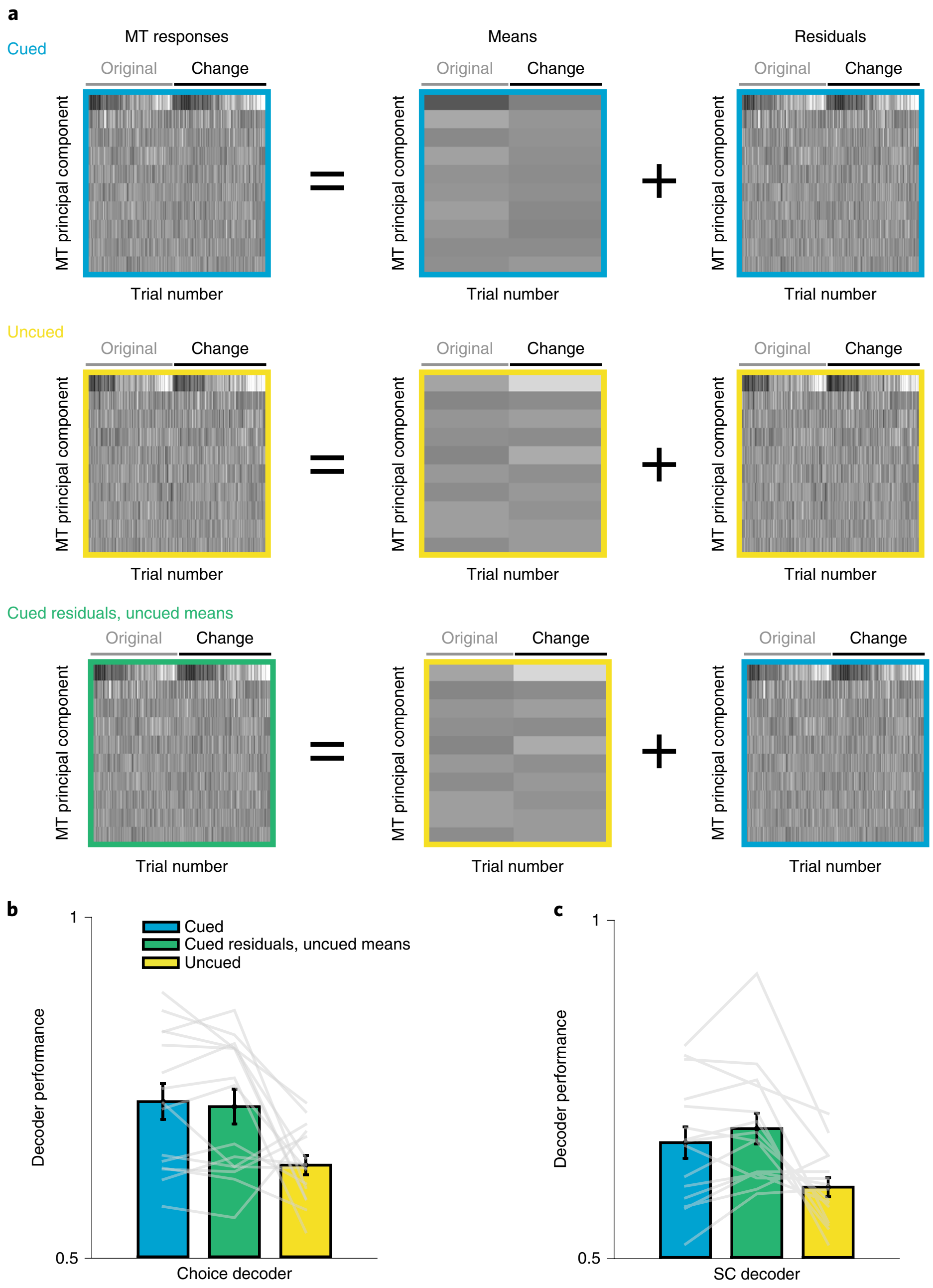

Fig. 5 | Effects of attention on the stimulus information that can be decoded from small populations of MT neurons is explained by changes in response variability. a, Schematic of our procedure to understand which attention-related changes could account for the improvement in the amount of stimulus information that could be gleaned using the choice decoder. We separated the first ten principal components of the MT population response (left) to the original and changed stimulus in both attention conditions into mean responses (scale adjusted to account for smaller value range) and residuals. We assessed the extent to which decoder performance was affected by attention-related changes in means and residuals by decoding stimulus information from MT responses on a separate set of trials in each attention condition and also using the residuals from the cued condition and the mean responses from the uncued condition (third row). See Methods for detailed decoding and cross-validation procedures. **b**, Using the procedure described in **a**, we found that reshaping the MT representation of the attended stimulus can be accomplished as a result of attention-related changes in response variability (for example, 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 ($n = 15$ sessions; two-tailed paired t-test, $P = 0.84$). The error bars represent the s.e.m.; the gray lines are the individual sessions. **c**, Same as **b** for the SC decoder. The amount of stimulus information that can be decoded using a single SC 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 ($n = 15$ sessions; two-tailed paired t-test, $P = 0.48$).

response variability that is private to each neuron and also shared between neurons, was enough to completely account for the attention-related improvement in decoded stimulus information in both the choice (Fig. 5b) and SC decoders (Fig. 5c). These common decoders captured the attention-related improvement in decoded stimulus information, and using residuals from the cued condition completely accounted for the attention-related improvement in decoded stimulus information.

Discussion

We used multi-neuron, multi-area recordings and psychophysics in detection and discrimination tasks to test two common previous hypotheses and a different hypothesis about the relationship between attention-related changes in perception and neuronal responses on the timescale of perceptual decisions. In contrast with the hypotheses motivating most of the extensive literature concerning the neuronal basis of attention, our data are most consistent with the hypothesis that attention reshapes population activity so that information about the attended stimulus is read out to guide behavior. Our conclusions are based on comparing the visual information that can be gleaned from decoders optimized for the stimulus, the animals' choices and the activity of groups of visuomotor neurons. These results support the idea that behavioral flexibility is mediated by reshaping the representation of visual stimuli rather than improvements in information coding, which may be impossible given the immense amount of sensory information encoded in the brains of even anesthetized animals[30] or in the responses of single neurons[13], or by changing the readout, which may be difficult to flexibly alter on the comparatively rapid timescale on which individuals can behaviorally shift attention[38].

The idea of reshaping sensory information to better align with static readout mechanisms at first sounds as though it would require much more exotic mechanisms than the other hypothesized attentional mechanisms. However, we showed that commonly observed effects of attention on neuronal response variability were sufficient to reshape the representation of attended stimuli so that they more effectively influence the activity of downstream neurons and behavior (Fig. 5b,c). Changing covariability may require a simpler mechanism than changing information coding or synaptic weights. We showed recently in a model that the covariability of a population of neurons can be readily changed by altering the balance of inhibition to excitation[39,40].

Although many studies are based on the implicit assumption that one or both of the information coding and communication hypotheses are true, several recent studies have failed to support the strongest versions of these hypotheses and our reshaping hypothesis unifies these results. Mante et al.[41] found the presence of both task-relevant and task-irrelevant information in the prefrontal cortex, suggesting that the task, or attention-related gating of information does not occur in earlier stages of processing, such as in visual cortex. This observation raises the question of why sensory responses are modulated if neurons near the end stages of processing in the prefrontal cortex still encode task-irrelevant information. To this point, Krauzlis et al.[42] suggested that attention-related changes in the sensory cortex may arise as a by-product of the process that interprets these signals. Further, a variety of experimental conditions that involve changing reward value[43] or saccade planning[44] result in changes in sensory responses that suggest a dependence on how the animal will use the sensory information. The reshaping hypothesis we propose in this study is consistent with all of these findings, suggesting that sensory responses are modulated by the task such that the relevant information affects behavior and the irrelevant information is retained, perhaps for future actions or memory. Our findings suggest that this reshaping is achieved by changes to correlated variability early on during visual processing, not by changing readout weights.

The idea that changing correlated variability better aligns sensory responses to a fixed readout is also consistent with our recent observation that in the change detection task, monkeys' choices are well aligned with the axis in population space that explains the most correlated noise[21]. 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[23]. 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.[45]).

While our results were broadly consistent across two tasks and two visual cortical areas, it is possible that attention uses different mechanisms in different tasks, brain areas or sensory modalities. In particular, it is possible that the mechanisms underlying change detection, which is an important component of natural vision, are different than other tasks or that the mechanisms differ by brain areas. Therefore, the observation that attention also does not change the amount of stimulus information that can be decoded from the visual cortex during a contrast discrimination task provides strong independent support for the generality of our findings. However, even if we happened on a special, albeit common, scenario using these two tasks, it is remarkable to observe a situation where the large attention-related change in behavioral performance can be accomplished without changing information coding or weights between areas. In contrast, theoretical models and machine learning techniques often accomplish flexibility in computation almost solely by changing weights[46–49]. Our results constitute an existence proof: an example of a situation where flexibility can be mediated by simple changes within the sensory cortex.

In the future, it will be interesting to use the same approach to determine whether similar mechanisms can account for behavioral changes associated with other cognitive processes (for example, task switching) that have been suggested to change the weights relating stimulus information to downstream neurons or behavior. Further, many neuropsychiatric disorders (including disorders of attention, autism spectrum disorder and schizophrenia) are thought to involve changes in the same computations thought to underlie attention[50]. 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 in this study.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0477-1.

Received: 23 January 2019; Accepted: 24 July 2019;
Published online: 2 September 2019

References

1. Carrasco, M. Visual attention: the past 25 years. *Vision Res.* **51**, 1484–1525 (2011).
2. Maunsell, J. H. R. Neuronal mechanisms of visual attention. *Annu. Rev. Vis. Sci.* **1**, 373–391 (2015).
3. Reynolds, J. H. & Chelazzi, L. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* **27**, 611–647 (2004).
4. Krauzlis, R. J., Lovejoy, L. P. & Zénon, A. Superior colliculus and visual spatial attention. *Annu. Rev. Neurosci.* **36**, 165–182 (2013).
5. Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
6. Ruff, D. A. & Cohen, M. R. Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* **17**, 1591–1597 (2014).
7. Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**, 879–888 (2009).
8. Fries, P. Rhythms for cognition: communication through coherence. *Neuron* **88**, 220–235 (2015).
9. Ruff, D. A. & Cohen, M. R. Attention increases spike count correlations between visual cortical areas. *J. Neurosci.* **36**, 7523–7534 (2016).
10. Ruff, D. A. & Cohen, M. R. A normalization model suggests that attention changes the weighting of inputs between visual areas. *Proc. Natl Acad. Sci. USA* **114**, E4085–E4094 (2017).

11. Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* **324**, 1207–1210 (2009).
12. Born, R. T. & Bradley, D. C. Structure and function of visual area MT. *Annu. Rev. Neurosci.* **28**, 157–189 (2005).
13. Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
14. Gandhi, N. J. & Katnani, H. A. Motor functions of the superior colliculus. *Annu. Rev. Neurosci.* **34**, 205–231 (2011).
15. Sparks, D. L. Conceptual issues related to the role of the superior colliculus in the control of gaze. *Curr. Opin. Neurobiol.* **9**, 698–707 (1999).
16. Wurtz, R. H. & Albano, J. E. Visual-motor function of the primate superior colliculus. *Annu. Rev. Neurosci.* **3**, 189–226 (1980).
17. Horwitz, G. D. & Newsome, W. T. Separate signals for target selection and movement specification in the superior colliculus. *Science* **284**, 1158–1161 (1999).
18. Crapse, T. B., Lau, H. & Basso, M. A. A role for the superior colliculus in decision criteria. *Neuron* **97**, 181–194.e6 (2018).
19. Horwitz, G. D. & Newsome, W. T. Target selection for saccadic eye movements: prelude activity in the superior colliculus during a direction-discrimination task. *J. Neurophysiol.* **86**, 2543–2558 (2001).
20. Cohen, M. R. & Maunsell, J. H. R. Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron* **70**, 1192–1204 (2011).
21. Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J. & Cohen, M. R. Learning and attention reveal a general relationship between population activity and behavior. *Science* **359**, 463–465 (2018).
22. Mayo, J. P. & Maunsell, J. H. Graded neuronal modulations related to visual spatial attention. *J. Neurosci.* **36**, 5353–5361 (2016).
23. Ruff, D. A., Ni, A. M. & Cohen, M. R. Cognition as a window into neuronal population space. *Annu. Rev. Neurosci.* **41**, 77–97 (2018).
24. Cohen, M. R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–819 (2011).
25. Pooresmaeili, A., Poort, J. & Roelfsema, P. R. Simultaneous selection by object-based attention in visual and frontal cortex. *Proc. Natl Acad. Sci. USA* **111**, 6467–6472 (2014).
26. Ecker, A. S., Denfield, G. H., Bethge, M. & Tolias, A. S. On the structure of neuronal population activity under fluctuations in attentional state. *J. Neurosci.* **36**, 1775–1789 (2016).
27. Denfield, G. H., Ecker, A. S., Shinn, T. J., Bethge, M. & Tolias, A. S. Attentional fluctuations induce shared variability in macaque primary visual cortex. *Nat. Commun.* **9**, 2654 (2018).
28. Moreno-Bote, R. et al. Information-limiting correlations. *Nat. Neurosci.* **17**, 1410–1417 (2014).
29. Kanitscheider, I., Coen-Cagli, R. & Pouget, A. Origin of information-limiting noise correlations. *Proc. Natl Acad. Sci. USA* **112**, E6973–6982 (2015).
30. Kohn, A., Coen-Cagli, R., Kanitscheider, I. & Pouget, A. Correlations and neuronal population information. *Annu. Rev. Neurosci.* **39**, 237–256 (2016).
31. Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical areas interact through a communication subspace. *Neuron* **102**, 249–259.e4 (2019).
32. Zénon, A. & Krauzlis, R. J. Attention deficits without cortical neuronal deficits. *Nature* **489**, 434–437 (2012).
33. Herrero, J. L., Gieselmann, M. A., Sanayei, M. & Thiele, A. Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron* **78**, 729–739 (2013).
34. Gregoriou, G. G., Rossi, A. F., Ungerleider, L. G. & Desimone, R. Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4. *Nat. Neurosci.* **17**, 1003–1011 (2014).
35. Luo, T. Z. & Maunsell, J. H. R. Neuronal modulations in visual cortex are associated with only one of multiple components of attention. *Neuron* **86**, 1182–1188 (2015).
36. Nandy, A. S., Nassi, J. J. & Reynolds, J. H. Laminar organization of attentional modulation in macaque visual area V4. *Neuron* **93**, 235–246 (2017).
37. Snyder, A. C., Morais, M. J. & Smith, M. A. Dynamics of excitatory and inhibitory networks are differentially altered by selective attention. *J. Neurophysiol.* **116**, 1807–1820 (2016).
38. Egeth, H. E. & Yantis, S. Visual attention: control, representation, and time course. *Annu. Rev. Psychol.* **48**, 269–297 (1997).
39. Huang, C. et al. Circuit models of low-dimensional shared variability in cortical networks. *Neuron* **101**, 337–348.e4 (2019).
40. Kanashiro, T., Ocker, G. K., Cohen, M. R. & Doiron, B. Attentional modulation of neuronal variability in circuit models of cortex. *eLife* **6**, e23978 (2017).
41. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
42. Krauzlis, R. J., Bollimunta, A., Arcizet, F. & Wang, L. Attention as an effect not a cause. *Trends Cogn. Sci.* **18**, 457–464 (2014).
43. Baruni, J. K., Lau, B. & Salzman, C. D. Reward expectation differentially modulates attentional behavior and activity in visual area V4. *Nat. Neurosci.* **18**, 1656–1663 (2015).
44. Steinmetz, N. A. & Moore, T. Eye movement preparation modulates neuronal responses in area V4 when dissociated from attentional demands. *Neuron* **83**, 496–506 (2014).
45. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
46. Wang, X. J. & Yang, G. R. A disinhibitory circuit motif and flexible information routing in the brain. *Curr. Opin. Neurobiol.* **49**, 75–83 (2018).
47. Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X. J. Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* **22**, 297–306 (2019).
48. Buonomano, D. V. & Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* **10**, 113–125 (2009).
49. Huang, G., Huang, G. B., Song, S. & You, K. Trends in extreme learning machines: a review. *Neural Netw.* **61**, 32–48 (2015).
50. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2012).

Acknowledgements

We are grateful to J. Maunsell for granting permission for the use of the V4 data, to K. McKracken for providing technical assistance and to F. Baqai, M. Getz, J. Hennig, C. Huang, A. Ni and J. Maunsell for comments on an earlier version of this manuscript. M.R.C. is supported by National Institutes of Health grant nos. R00EY020844 and R01EY022930 and by Core Grant P30 EY008098s, and by the McKnight, Whitehall, Sloan and Simons Foundations.

Author contributions

D.A.R. and M.R.C. 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.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0477-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.A.R.

Methods

Animals. Two adult male rhesus monkeys (*Macaca mulatta*, 8 and 9 kg and 8 and 6 years old, respectively) were the animals used to simultaneously record from MT and SC. 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 (v. 3)[51,52]) on a cathode-ray tube monitor (calibrated to linearize intensity; 1,024 × 768 pixels; 120 Hz refresh rate) placed 54 cm from each 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^{-1}), neuronal responses (30,000 samples s^{-1}) and the signal from a photodiode to align neuronal responses to stimulus presentation times (30,000 samples s^{-1}) using hardware from Ripple.

Behavioral task. As described previously[5], 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 full contrast, drifting 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 and 400 ms) until, at a random, unsignaled time, the direction of one of the stimuli changed from that of the preceding stimulus (Fig. 1c). The monkey received a liquid reward for making a saccade to the stimulus that changed within 450 ms of its onset. Attention was cued (using instruction trials before each block) in blocks of 50–100 trials, and randomly alternated between blocks where attention was cued to either the left or 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 direction change; Fig. 1d). To encourage fixation on longer trials, catch trials, where no stimulus changed direction and monkeys were rewarded for maintaining fixation, were randomly intermixed throughout each block and made up approximately 12% of total trials. Psychometric data were fitted with Weibull functions. Before recording commenced, the monkeys were extensively trained to have stable thresholds across a range of spatial locations (3–6 months). Because we recorded from several dozen neurons simultaneously, we could not optimize the stimuli for all neurons. We made sure to position one Gabor stimulus in the joint receptive field of the recorded neurons in both areas and we made an effort to set the properties of the size (approximately 3–6 degrees of visual angle), speed (approximately 3–12 degrees of visual angle per second) and direction of the stimuli so that they drove as many MT units as possible. The direction of all of the stimuli before the direction change (termed original stimulus) was constant throughout a recording session and this direction was typically either the median or mode of the distribution of MT preferred directions from that session. The range of direction changes differed from session to session, was selected based on the animals' training history and depended on stimulus properties such as eccentricity and size. A typical range of change amounts for both animals was 1–35 degrees in log-spaced steps.

Electrophysiological recordings. Using linear 24-channel moveable probes (Plexon), we simultaneously recorded extracellular activity from direction-selective neurons in MT and neurons in SC that responded either visually, before a saccade or both. Before beginning the experiment, we searched for neurons in both areas that had overlapping spatial receptive fields (Fig. 1e) as determined by mapping with both drifting gratings and a delayed saccade task. 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 SC for monkey HO; 36–53 units per session, mean 21 in MT, 22 in SC for monkey ST) collected from the right hemisphere using moveable, linear 24-channel V-probes (Plexon; interelectrode spacing in MT = 50µm, SC = 100µm). We presented visual stimuli and tracked eye position as described previously[9]. The data presented are from 6 d of recording for monkey HO and 9 d of recording for monkey ST. Each day consisted of multiple blocks of the attention task (Fig. 1c; mean 1,015 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 v3.3.5 (Plexon). We based our analyses on both single units and multiunit clusters and use the term 'unit' to refer to either. Neuronal analyses in Supplementary Figs. 1 and 2 used spike count responses between 50 and 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 Figs. 3 and 4 used shorter response windows. Responses to both original and changed stimuli were measured from 50 to 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. Using these shorter response windows did not qualitatively affect the measures of attention described in Supplementary Fig. 1. Attention still increased the firing rates of MT neurons (mean attention index = 0.034, median attention index = 0.034; $n = 306$ units, two-tailed Wilcoxon signed-rank test, $P = 1.2 \times 10^{-17}$) and SC neurons (mean attention index = 0.071, median attention index = 0.05; $n = 345$ units, two-tailed Wilcoxon signed-rank test, $P = 4.2 \times 10^{-44}$) and decreased noise correlations within MT ($n = 3{,}285$ pairs, two-tailed Wilcoxon signed-rank test, $P = 7.7 \times 10^{-17}$). 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 before 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 before the onset of the first stimulus or if its response during a 100 ms epoch before a saccade on hit (correct) trials to the contralateral side was 10% larger than that same baseline. Stimulus presentations during which a microsaccade was detected were excluded from the analyses[9,53].

For the firing rate analyses in Supplementary Fig. 1a,b, attention indices were calculated using the average spike counts on the (original) stimulus presentation before 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 $(\text{attend}_{\text{in}} - \text{attend}_{\text{out}})/(\text{attend}_{\text{in}} + \text{attend}_{\text{out}})$. For illustrative purposes, the significance of individual units was determined by a two-tailed 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} (ref. [24])) 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 Supplementary Fig. 1c, we included responses from stimulus presentations 2 through 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 (that is, stimulus presentations where the behavioral response on the subsequent stimulus presentation was not a saccade). We *z*-scored responses as a function of the stimulus presentation number in each trial and then pooled the data across stimulus presentations before calculating noise correlations. The results did not qualitatively change if we did not perform this *z*-score procedure. For Supplementary Fig. 1d, we included data from all stimulus presentations before 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 the correlation analyses. The data presented in Supplementary Fig. 1c consisted of 3,285 MT pairs, 3,948 SC pairs and 6,934 between-area pairs.

Decoding. We focused our decoding analyses (Figs. 2, 3 and 5) on trials where the third largest (middle) direction change occurred because changes of that magnitude occurred in both attention conditions. This approach also serves to linearize the problem by attempting to classify between one of two directions of motion. Therefore, we restricted our decoding approach to using linear methods. We performed the decoding analyses using responses from trials that were either hits (correct detection) or misses (maintained fixation after change in stimulus). All of the datasets contained at least ten trials in each attention condition and at least three hits and three misses in each condition. We did not include false alarms in the analyses because there were too few (and they were too inconsistent across recording sessions) to handle appropriately.

We used the decoding strategy schematized in Fig. 2. We began by constructing a matrix of MT responses for each attention condition: 'MT responses' (number of MT neurons by 2 × number of trials matrix of MT responses to the stimuli before the direction change and the changed stimulus on the relevant trials). The stimulus decoder was performed using two matrices for each attention condition: all of 'middle area temporal responses' (number of MT neurons by 2 × number of trials matrix of MT responses to the stimuli before the direction change and the changed stimulus on the relevant trials) and 'motion direction' (a 1 by 2 × number of trials vector of zeros for the stimulus before the change, referred to as 'original', and ones for the changed stimulus, referred to as 'change'). The choice decoder was performed using two matrices for each attention condition: the responses during change stimulus presentations from 'MT responses' (number of MT neurons by 1 × number of trials matrix of MT responses to the change stimulus on the relevant trials) and 'choice' (a 1 by 1 × number of trials vector of zeros for change stimulus presentations where 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'). The SC decoder was performed using two matrices for each attention condition: the responses during original stimulus presentations 'MT responses' (number of MT neurons by 1 × number of trials matrix of MT responses to the original stimulus on the relevant trials) and 'SC responses' (number of SC neurons by 1 × number of trials matrix of SC responses to the original stimuli on the relevant trials). We refer to this final decoder as 'SC decoder' but the weights are defined with no directionality. We have simply identified the weights that best relate the activity between the two areas. We used only responses to the original stimulus for the SC decoder because of the strong presaccadic responses present during the changed stimuli.

We cross-validated by holding out the two stimulus presentations from 'MT responses' (for the original and changed stimuli) from one trial at a time to perform the classification of motion direction. To reduce the number of weights we needed to fit and therefore improve our confidence in the weights we fitted, we performed principal component analysis on the MT and SC responses to find the first 10 principal components in each area. The choice of number of vectors did not qualitatively affect the results in the range of 4–15 vectors. 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 principal components plus a vector of ones to 'motion direction', 'choice' or projections along the first ten SC principal components in each attention condition.

We assessed the stimulus information in each decoder (Fig. 3) by multiplying the projections of MT responses to the original and changed stimuli from the held-out trial by the fitted weights, 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). The performance of the decoder is defined as the area under the receiver operating characteristic curve comparing the distributions of weighted average responses to each stimulus using the weights constructed for each decoder.

The critical aspect of the decoding analysis is that we ask how much stimulus information is contained in each different subset of MT activity. The stimulus (or optimal) decoder will perform best because it was designed specifically to ask this question. The choice and SC decoders identify different subspaces of MT activity and then ask how much stimulus information is contained in those subspaces. These decoders, by definition, will perform worse than the stimulus decoder, but they are asking the same question.

To assess the stability of the weights for each decoder in the two attention conditions, we assessed the stimulus information gleaned by each decoder using the sensory responses from one attention condition and the weights calculated from the other (Figs. 3 and 4). Because the responses of visual neurons are correlated and because our behavioral response is binary, the weights found with our linear decoding methods are nonunique[23,31]. Therefore, it is not informative to make direct comparisons of the weights across conditions. Instead, we borrowed the spirit of the analyses in a recent study[31] and compared the stimulus information that could be gleaned using each set of weights in each attention condition. In general, the choice and SC decoders performed better with weights computed from the same attention condition, even though we cross-validated these analyses. (This effect could be attributed to non-stationarities in the recordings or the monkey's behavior.) The critical comparison is the performance of the decoders using the sensory responses from one attention condition and the weights from the other (Figs. 3 and 4).

For the decoding analysis in Fig. 5, we took a similar approach to the previously described choice and SC decoders, except that we combined data from both the cued and uncued conditions to calculate the decoding weights. We then decomposed the population responses to each stimulus in each attention condition into mean responses and residuals ($R = M + S$, where R is the number of neurons by the 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.

The analyses of the V4 data from the change detection task (Fig. 4a,b) were carried out in the identical manner to MT data described earlier. This dataset consisted of multi-neuron recordings using Utah arrays placed in both hemispheres of V4 during 37 experimental sessions in two animals, the details of which are described in Cohen and Maunsell[5]. Data from each hemisphere was treated separately in the decoding analyses, so each session contributes two data points for each analysis (gray lines, Fig. 4a). The details of the contrast discrimination task used in Fig. 4c required a different form of the stimulus decoder. This dataset consisted of multi-neuron recordings using Utah arrays placed in both hemispheres of V4 during 17 experimental sessions in two animals. The details of this experiment have been described previously[6]. Briefly, two monkeys judged which of two stimuli in a pair was higher contrast by making a saccade to a target representing its choice. Attention toward one pair of stimuli or the other was changed in blocks. The stimulus decoder (Fig. 4c) compares performance using V4 responses to distinguish between a given stimulus configuration and its opposite configuration in the attended and unattended conditions. As in the other V4 dataset, data from each hemisphere were treated separately.

Statistics. Paired tests, either two-tailed *t*-tests or nonparametric Wilcoxon signed-rank tests, were employed for all statistical analyses. Where *t*-tests were used, the data distribution was assumed to be normal but this was not formally tested. No statistical methods were used to predetermine sample sizes; however, our sample sizes are similar to those reported in previous publications[6,9]. There was no way to perform data collection and analysis blind to the conditions of the experiments because our data were not grouped. Please see the Nature Research Reporting Summary for additional information.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The custom MATLAB code is available from the authors upon reasonable request.

References

51. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
52. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
53. Engbert, R. & Kliegl, R. Microsaccades uncover the orientation of covert attention. *Vision Res.* **43**, 1035–1045 (2003).

nature research

Corresponding author(s): Douglas Ruff

Last updated by author(s): Jul 18, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were recorded using Trellis software (Ripple, v. 1.7.2.6). Visual stimuli were presented using Matlab (v. 2012a) and Psychophysics toolbox (v. 3).
Data analysis	Custom code was written using MATLAB (v. 2018a). Spike sorting was performed using Plexon Offline Sorter (v 3.3.5).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes but our sample sizes are similar to those reported in previous publications (Cohen and Maunsell, 2009). In the newly described dataset, we recorded from 306 MT units and 345 SC units across 15 sessions from two monkeys.
Data exclusions	No data were excluded in the initial analyses.
Replication	Monkeys were trained and tested over a period of more than a year and produced consistent behavioral results. We replicated many of the previous observations of how attention affects the responses of MT and SC neurons and our results were reproducible across sessions.
Randomization	No experimental groupings were used in this study.
Blinding	Blinding is not relevant because no grouping was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Two male rhesus monkeys (Macaca mulatta) weighing 8-9 kg (8 and 6 years old, respectively) were used.
Wild animals	Wild animals were not used in this study.
Field-collected samples	Field-collected samples were not used in this study.
Ethics oversight	All animal procedures were approved by the Institutional Animal Care and Use Committees of the University of Pittsburgh and Carnegie Mellon University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.