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: ruffd@pitt.edu

Abstract – 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 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 novel 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.

Keywords – *attention, decoding, population analyses*

22 **Introduction –**

23 Each of the huge number of psychophysical and physiological studies of visual attention show
24 that attention profoundly affects subjects' perceptual abilities and also modulates the responses
25 of populations of neurons at every stage of visual and oculomotor processing[1-4], Despite these oft
26 replicated observations, whether any of the observed neuronal modulations can account for the
27 improvements in psychophysical performance remains unknown. Two, non-mutually exclusive,
28 hypotheses have dominated the literature (Figure 1A): that attention 1) improves visual
29 information coding[5-7], or 2) improves the efficiency with which visual information is read out by
30 the premotor neurons involved in decision-making [8-11]. The studies used to support these
31 hypotheses were limited by available data and analysis methods, which primarily involved the
32 responses of single neurons or pairs of simultaneously recorded neurons in the same brain area.
33 We evaluated these hypotheses using the responses of groups of simultaneously recorded
34 neurons in multiple stages of visuomotor processing, psychophysics, and data analysis methods
35 that leverage that unique combination. We recorded simultaneously from groups of neurons in
36 area MT, which encodes motion information [12,13] and the superior colliculus (SC), where
37 neuronal responses are either visual, oculomotor, or intermediate, contribute to gaze control [14-16]
38 and are involved in computing perceptual decisions [17-19]. When we analyzed the responses of
39 single neurons or pairs of neurons, we replicated previous observations, including the results
40 from two of our previous studies, which focus on visual area V4 in two different tasks with
41 spatial attention components: an orientation change detection task [5] and a contrast discrimination
42 task [6]. However, constraining our analyses of our MT data set or of both V4 data sets by the
43 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 (Figure 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 Figure 1C[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 data sets, the animals' performance in our new experiment 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[12,13]. The SC is thought to play many roles in visually guided tasks including gaze control [14-16], decision-making [17-19] and attention [4].

Figure 1

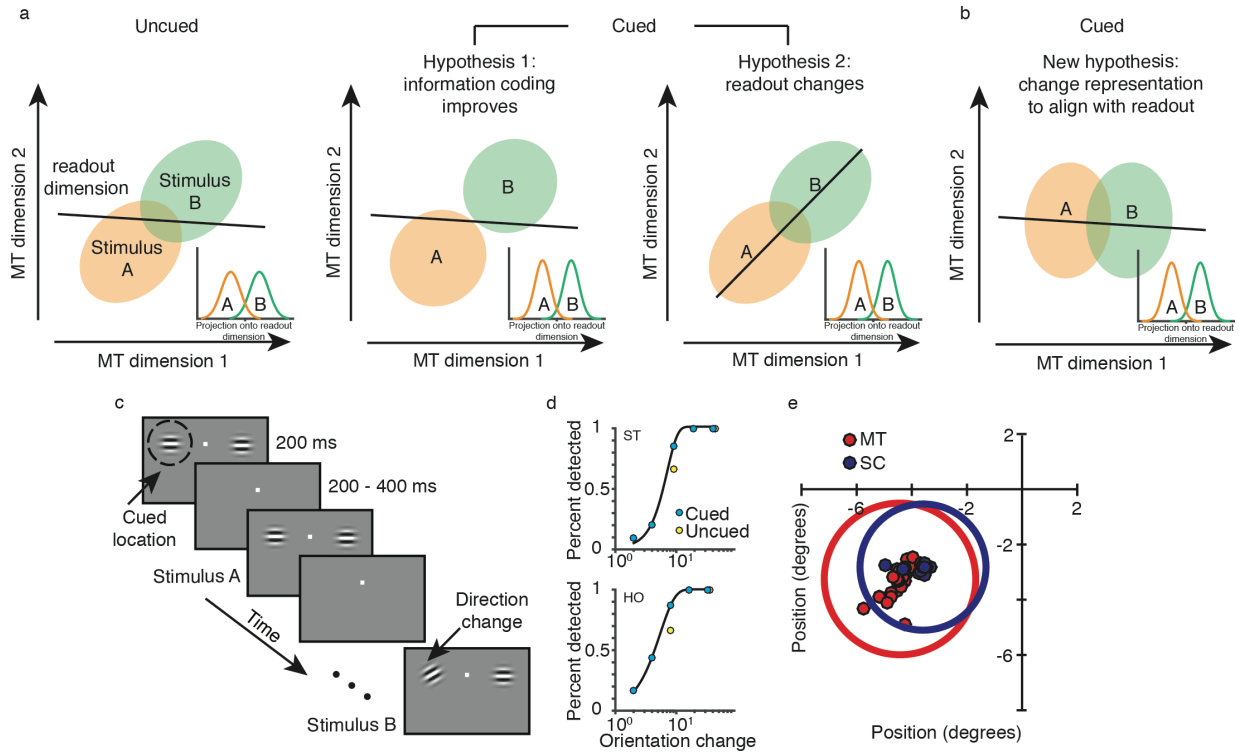

Figure 1. Hypotheses and methods. (A) Schematics describing predominant hypotheses about links between attention, visual cortical activity, and behavior. The left plot depicts MT population responses to two visual stimuli plotted along two dimensions in population response space (e.g. the first two principal 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 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 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 (cued 76.5% detected across sessions, uncued 51.8% detected; Wilcoxon signed-rank test, p<.001). (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.

Population recordings replicate previously observed effects of attention

The two 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 (readout) hypothesis posits that attention changes the way that stimulus information is read out of MT to influence 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 to evaluate each hypothesis using analyses that leverage our simultaneous measurements from the subjects' 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 data set, 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 the SC (Supplementary Figure 1A and 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

115 correlation, or r_{SC}[24]; Supplementary Figure 1C). Consistent with studies evaluating the readout
116 hypothesis, attention increases correlated variability between the two areas[9,10,25] (Supplementary
117 Figure 1C). This increase between areas suggests that attention-related effects are not simply due
118 to global reductions in slow fluctuations, which has recently been hypothesized to explain
119 attention-related correlation decreases within a single brain area[26] (Supplementary Figure 2 and
120 Supplementary Text). Further, this attention-related increase was weakly dependent on the visual
121 responsivity of SC neurons (Supplementary Figure 3).
122
123 *Neuronal population decoding methods provide incomplete support the information coding or*
124 *readout hypotheses*
125 We reasoned that analyzing the relationship between populations of simultaneously recorded
126 neurons in multiple brain areas with the animals' behavior would provide insight into the relative
127 importance of each hypothesized mechanism. To this end, we determined whether attention
128 affects the amount of stimulus information that can be decoded from the population of MT
129 neurons using cross-validated linear decoders that are optimized to a) dissociate between the
130 original and changed stimuli (Stimulus decoder in Figure 2), b) predict the animals' choices
131 (whether or not they made an eye movement in response to change stimuli; Choice decoder), or
132 c) predict the activity of the population of SC neurons we recorded (using responses to the
133 original stimulus; SC decoder). These decoders were always constructed using data from trials
134 with the intermediate change amount (see Figure 1D).

Figure 2

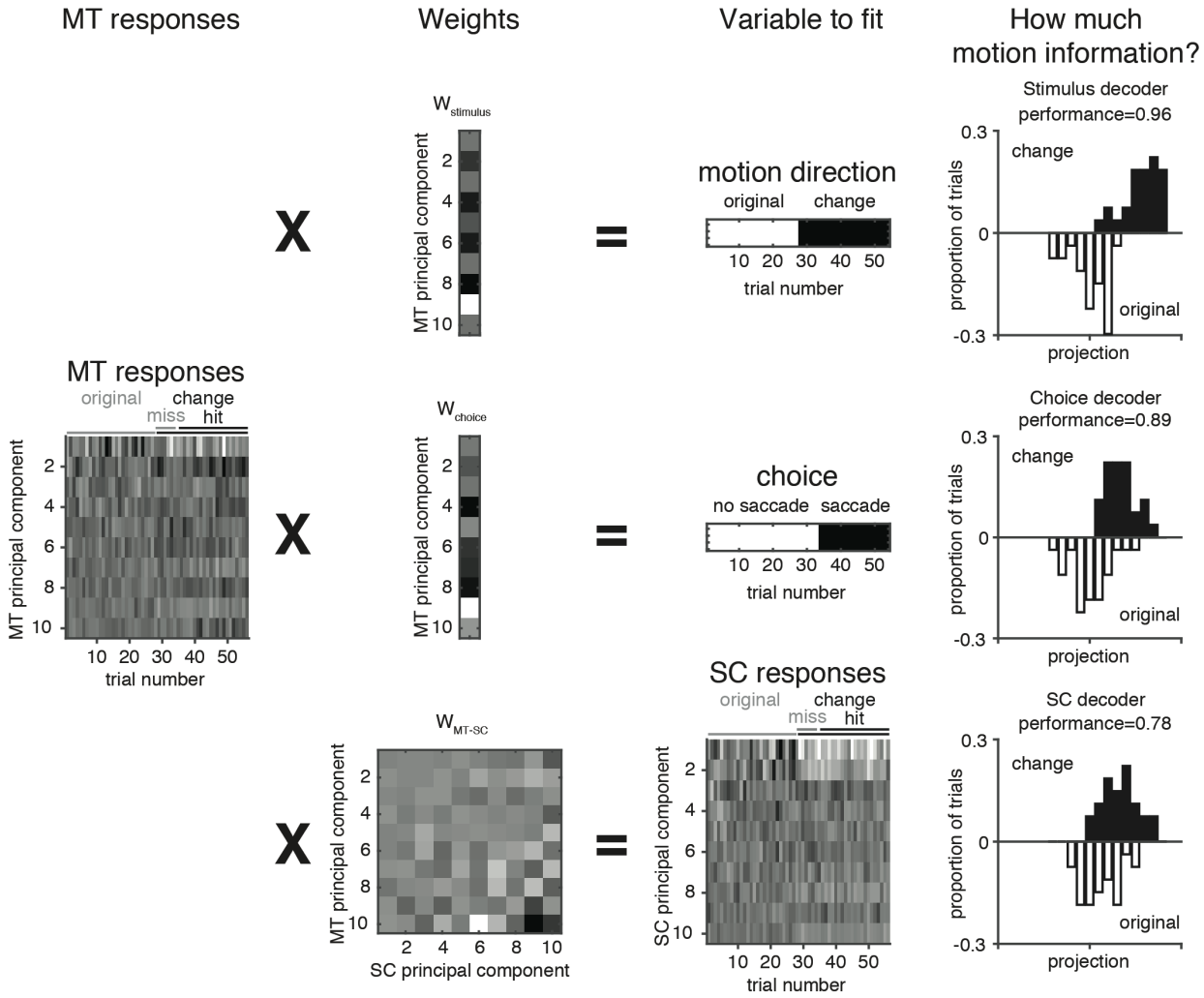

Figure 2 – Schematic of our decoding procedure. We used linear regression to find the weights (second column) that best relate the first ten principal components of the MT population's response (left) to both the original and change stimuli (Stimulus decoder; top row), the animal's choice in response 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.

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 data

set, even when we used a decoder optimized separately for each attention condition (Figure 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 increasingly likely to be orthogonal to this axis in larger populations [27-29]. 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 likely to be even more minimal for larger populations, making it seem unlikely 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 (Figure 3A, middle bars) or that is shared with the SC (Figure 3A, right bars) than it did on the Stimulus decoder ($p<.05$). However, this difference could arise from either a weight change (Figure 1A) or a change within MT that results in more stimulus-related visual information being projected onto a static readout dimension (Figure 1B).

Figure 3

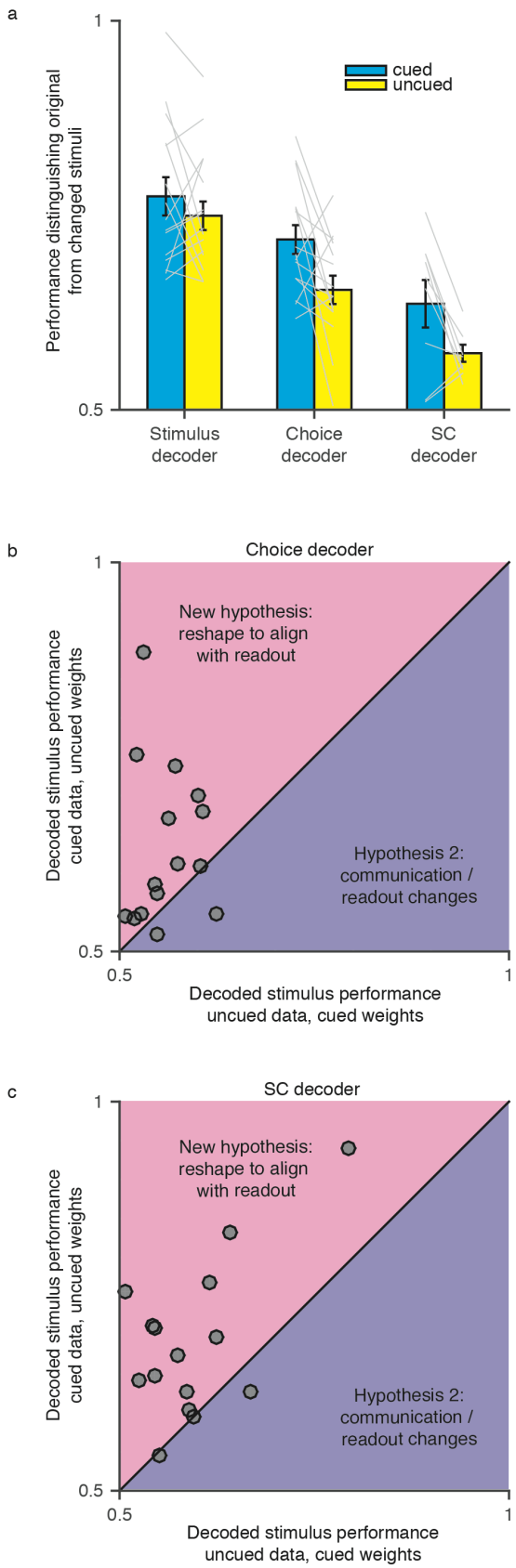

Figure 3. Effects of attention on the stimulus information that can be decoded from small populations of MT neurons. (A) Ability of a cross-validated linear decoder to distinguish the original from changed stimuli (intermediate change amount) for each decoder. Error bars represent SEM, gray lines are individual sessions. 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).

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 responses 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 non-unique, making it impossible to identify weight changes by analyzing the weights themselves [23,30]. However, we can infer their stability by measuring the stimulus information gleaned by

202 each decoder using weights from the opposite attention condition from which they were
203 calculated (see Methods). Both the Choice and SC decoders gleaned more stimulus information
204 from MT responses in the attended than unattended condition when we used the weights
205 computed in the opposite attention condition (Figures 3B and 3C). Together, these neuronal
206 population analyses that use the animals' behavior and the activity of downstream neurons to
207 assess the hypothesized attention mechanisms reveal that neither the information coding nor
208 readout hypothesis provide a satisfactory account of the large observed attention-related
209 behavioral improvement.
210
211 Our observations suggest that in MT neurons recorded while monkeys performing a change
212 detection task, attention acts primarily by changing the visual information that is used to guide
213 behavior using relatively fixed readout weights. To investigate the generality of these
214 observations to different visual areas and different tasks, we tested these hypotheses using two
215 additional datasets. In the first dataset, monkeys performed the same direction change detection
216 described here while we recorded from populations of V4 neurons[5]. Similar to our results in MT,
217 we found that attention had larger effects on the stimulus information that is related to the
218 animals' choices (Choice decoder; Figure 4A) than it did on the stimulus information that could
219 be gleaned using an optimal (Stimulus) decoder ($p<0.05$; Figure 4B). As in our MT data set
220 (Figure 3B), the results from this data set suggest that attention typically reshapes V4 responses
221 to align with relatively fixed readout mechanisms: decoding performance was typically better
222 using the V4 responses from the cued condition and the Choice decoder weights from the uncued
223 condition (y-axis) than using the V4 responses from the uncued condition and the Choice
224 decoder weights from the cued condition ($p<0.05$).

225 In the second new data set, we searched for attention-related changes in information coding in
226 V4 neurons while monkeys performed a discrimination task[6]. These data provide a particularly
227 important test of the information coding hypothesis because unlike in the change detection task
228 in which attention has fairly uniform effects on V4 and MT neurons (increasing rates and
229 decreasing noise correlations), we showed that in our discrimination task, attention can flexibly
230 increase or decrease noise correlations in a way that is broadly consistent with improving
231 information coding. Despite these findings, the results of our decoding analyses were similar for
232 the detection and discrimination tasks, meaning that we did not find strong evidence that
233 attention improves the amount of stimulus information that can be optimally extracted from a
234 population of visual neurons in either task (Figure 4C, p=.09). Together, these results provide
235 evidence that in multiple visual areas and visually-guided tasks, attention acts primarily to
236 reshape population activity so that more stimulus information is used to guide behavior using
237 relatively fixed decision mechanisms.

Figure 4

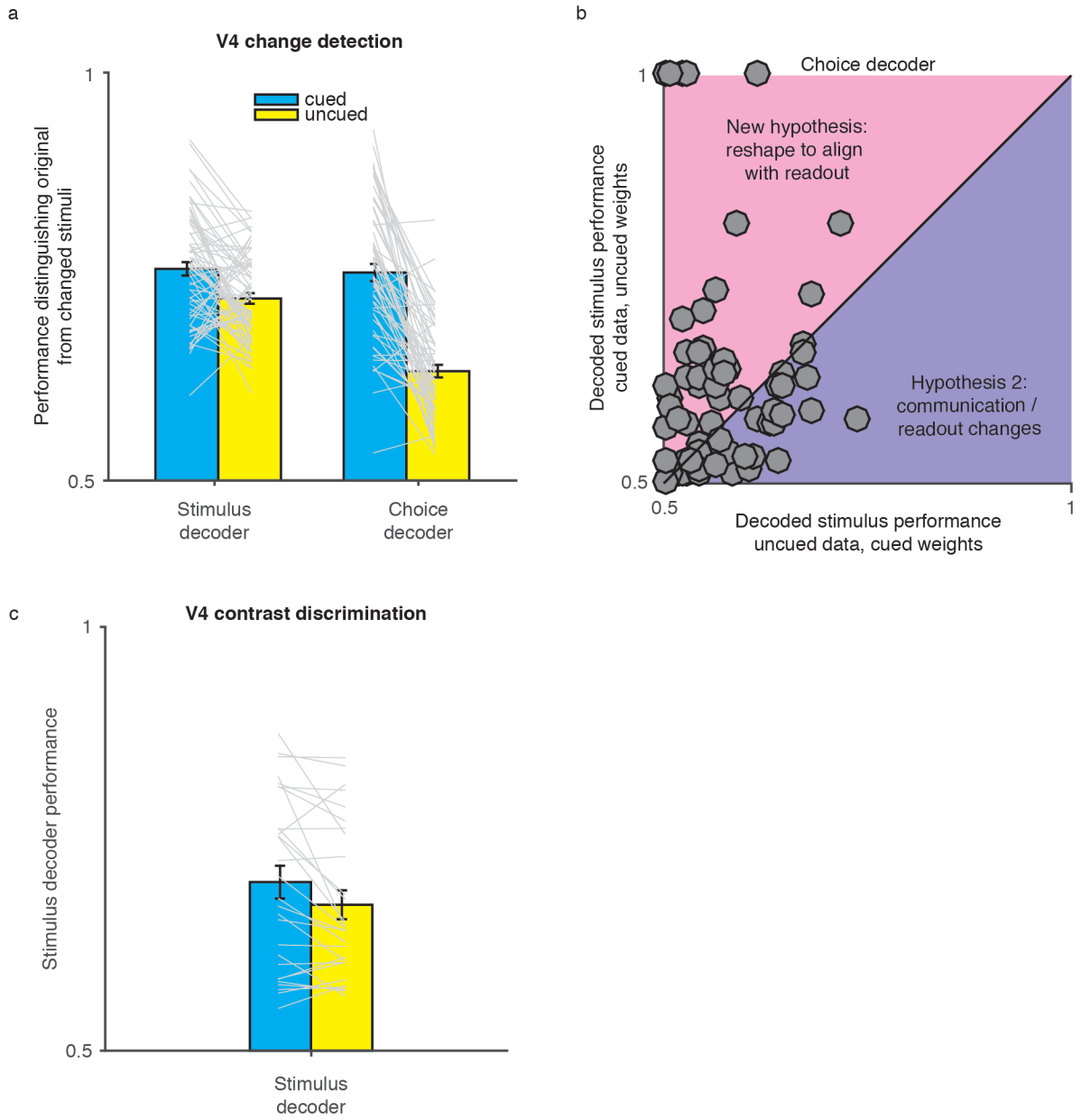

Figure 4. Similar attention-related effects on neuronal populations in two brain areas and two tasks. (A) In a change detection task, the effects of attention on the stimulus information that can be decoded from small populations of V4 neurons is similar to MT. The plot shows the ability of a cross-validated linear decoder to distinguish the original from changed stimuli (intermediate change amount) for both the Stimulus and Choice decoders (no SC data was available). Error bars represent SEM, gray lines are individual hemisphere-sessions (see methods). Attention significantly affected the performance of both the Stimulus and Choice decoders (t-test, $p<0.05$), but the attention-related improvement in the Choice decoder was greater than in the Stimulus decoder (paired t-tests, $p<0.05$). (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; paired t-test, $p<0.05$; compare to Figure 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 ($p=.09$). 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.

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 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 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 oft observed signatures of attention such as changes in firing rate gain[2,3,23] or pairwise noise correlations[5-7,9,20-22,31-37]. 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 (Figures 3B and 3C), we constructed a single Choice decoder for both attention conditions (Figure 5A) and determined that it captured the attention-related improvement in decoded stimulus information (compare the blue and yellow bars in Figure 5B; $p< .05$). 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 (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-

276 related improvement in decoded stimulus information in both the Choice (Figure 5B) and SC

277 decoders (Figure 5C). These common decoders captured the attention-related improvement in

278 decoded stimulus information (p< .05) and using residuals from the cued condition completely

279 accounted for the attention-related improvement in decoded stimulus information (p= .84 for the

280 Choice decoder and p=.48 for the SC decoder).

Figure 5

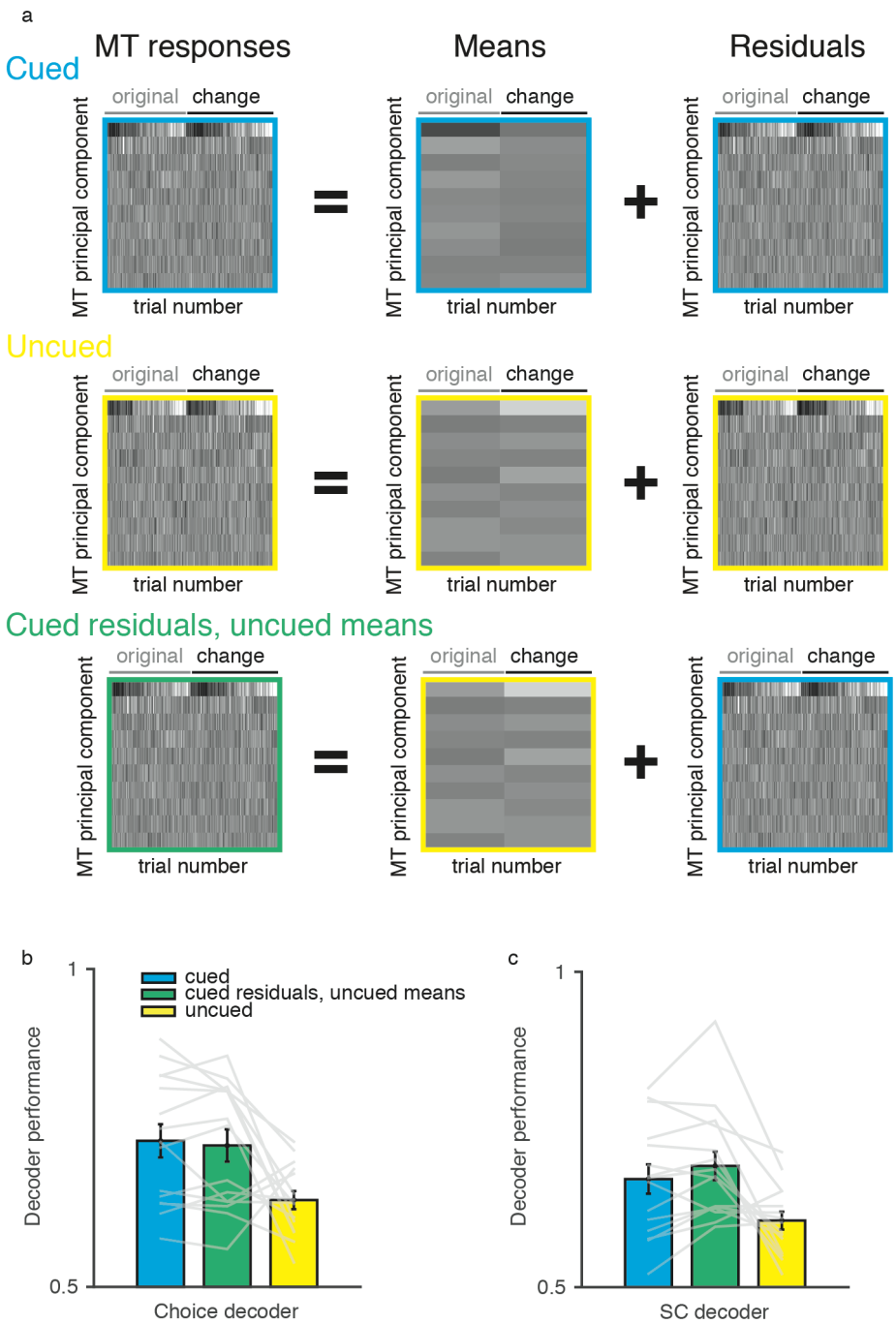

Figure 5. Effects of attention on the stimulus information that can be decoded from small populations of MT neurons is explained by changes in response variability. (A) Schematic of our procedure to understand which attention-related changes could account for the improvement in the amount of stimulus 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 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). Error bars represent SEM, gray lines are 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 (paired t-test, p=.48).

Discussion –

We used multi-neuron, multi-area recordings and psychophysics in detection and discrimination tasks to test two previous hypotheses and one novel hypothesis about the relationship between attention-related changes in perception and in 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 novel 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[29] or in the responses of single neurons[13]) or by changing read out, which may be comparatively difficult to flexibly alter on the approximately 150 ms timescale on which subjects can behaviorally shift attention[38-40].

The idea of reshaping sensory information to better align with static read out mechanisms seems like 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 (Figure 5B and 5C). 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 [41,42].

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 [43]).

While our results were broadly consistent across two tasks and two visual cortical areas, it remains 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

370 mechanisms differ by brain areas. Therefore, the observation that attention also does not change
371 the amount of stimulus information that can be decoded from visual cortex during a contrast
372 discrimination task provides strong independent support for the generality of our findings.
373 However, even if we happened upon a special, albeit common, scenario using these two tasks, it
374 is remarkable to observe a situation in which the large attention-related change in behavioral
375 performance can be accomplished without changing information coding or weights between
376 areas. In contrast, theoretical models and machine learning techniques accomplish flexibility in
377 computation almost solely by changing weights [44-47]. Our results constitute an existence proof: an
378 example of a situation in which flexibility can be mediated by simple changes within sensory
379 cortex.
380
381 In the future, it will be interesting to use the same approach to determine whether similar
382 mechanisms can account for behavioral changes associated with other cognitive processes (e.g.
383 task switching) that might seem more likely to change the weights relating stimulus information
384 to downstream neurons or behavior. Further, many neuropsychiatric disorders (including
385 disorders of attention, Autism, and schizophrenia) are thought to involve changes in the same
386 computations thought to underlie attention[48]. An exciting possibility is that these changes might
387 be identified and potential therapies evaluated in animal models using the combination of
388 behavioral evaluation and multi-neuron, multi-area recordings that we described here.
389

390 **Online Methods**
391 Methods Summary

The subjects of the simultaneously recorded MT and SC experiments 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.

Materials and Methods

The subjects of the simultaneously recorded MT and SC experiments 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 [49,50] on a CRT monitor (calibrated to linearize intensity; 1024x768 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[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 200-400 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 direction change, Figure 1D). Catch trials, where no stimulus changed direction and the monkey was rewarded for maintaining fixation, were randomly intermixed throughout each block and made up approximately 12% of total trials. Psychometric data were fit with Weibull functions. Before recording commenced, the monkeys were extensively trained to have stable thresholds

437 across a range of spatial locations (3-6 months). Because we recorded from several dozen
438 neurons simultaneously, we could not optimize the stimuli for all neurons. We made sure to
439 position one Gabor stimulus in the joint receptive field of the recorded neurons in both areas and
440 we made an effort to set the properties of the size (approximately 3-6 degrees of visual angle),
441 speed (approximately 3-12 degrees of visual angle per second) and direction of the stimuli so
442 that they drove as many MT units as possible. The direction of all of the stimuli prior to the
443 direction change (termed original stimulus) was constant throughout a recording session and this
444 direction was typically either the median or mode of the distribution of MT preferred directions
445 from that session. The range of direction changes differed from session to session, was selected
446 based on the animals' training history and depended on stimulus properties such as eccentricity
447 and size. A typical range of change amounts for both animals was 1-35 degrees in log-spaced
448 steps.
449
450 Electrophysiological Recordings

451 We simultaneously recorded extracellularly from single units and sorted multiunit clusters (the
452 term "unit" refers to either). The dataset consisted of a total of 306 responsive MT units and 345
453 responsive SC units total (36-58 units per session, mean 20 in MT, 24 in the SC for Monkey HO;
454 36-53 units per session, mean 21 in MT, 22 in SC for Monkey ST) in both MT and the SC in the
455 right hemisphere using moveable, linear 24-channel V-probes (Plexon; inter-electrode spacing in
456 MT = 50μm, SC = 100μm). We presented visual stimuli and tracked eye position as previously
457 described[9]. The data presented are from 6 days of recording for Monkey HO and 9 days of
458 recording for Monkey ST. Each day consisted of multiple blocks of the attention task (Figure 1C;

459 mean 1015 of trials for Monkey HO, 745 for Monkey ST) preceded by receptive field mapping
460 using a delayed saccade task and direction tuning during passive fixation.

461 Data Analysis

462 All spike sorting was done offline manually using Offline Sorter (version 3.3.5; Plexon). We
463 based our analyses on both single units and multiunit clusters and use the term "unit" to refer to
464 either. Neuronal analyses in Supplemental Figure 1 and 2 used spike count responses between
465 50-250 ms after stimulus onset to account for visual latencies in the two areas. To remove
466 response contamination from eye movements during change stimuli, data presented in the
467 decoding analyses in Figure 3 and 4 used shorter response windows. Responses to both original
468 and changed stimuli were measured from 50-185 ms after stimulus onset for monkey HO and 50-
469 220 ms for monkey ST. These times were selected based on the distribution of each animal's
470 reaction times with the goal of maximizing the number of trials that could be included in the
471 analyses. Trials with reaction times that began during those windows were excluded. Using these
472 shorter response windows did not qualitatively affect the measures of attention described in
473 Supplemental Figure 1. Attention still increased the firing rates of MT (mean attention index =
474 0.034, median attention index = 0.034), Wilcoxon signed rank test, $p< 10^{-18}$) and SC neurons
475 (mean attention index = 0.071, median attention index = 0.05), Wilcoxon signed rank test, $p< 10^{-}$
476 ^{44}) and decreased noise correlations within MT (Wilcoxon signed rank test, $p< 10^{-72}$). To
477 minimize the impact of adaptation on our results, we did not analyze the first stimulus
478 presentation in each trial. We only analyzed a recorded MT unit if its stimulus-driven firing rate
479 was 10% higher than its firing rate as measured in the 100 ms prior to the onset of the first
480 stimulus. We only analyzed a recorded SC unit if its stimulus-driven firing rate was 10% higher
481 than its firing rate as measured in the 100 ms prior to the onset of the first stimulus or if its

482 response during a 100 ms epoch prior to a saccade on hit (correct) trials to the contralateral side
483 was 10% larger than that same baseline. Stimulus presentations during which a microsaccade
484 was detected were excluded from analyses[9,51]).
485
486 For firing rate analyses in Supplementary Figure 1A and B, attention indices were calculated
487 using average spike counts on the (original) stimulus presentation prior to correct detections of
488 the intermediate change amount depending on whether attention was directed into or out of the
489 receptive fields of the recorded neurons using the formula $(attend_{in} - attend_{out})/(attend_{in} +$
490 $attend_{out})$. Significance of individual units was determined by a paired t-test ($p<0.05$).
491
492 Noise correlations
493
494 We defined the correlated variability of each pair of simultaneously recorded units (quantified as
495 spike count correlation or r_{SC}[24]) as the Pearson correlation coefficient between the responses of
496 the two units to repeated presentations of the same stimulus. This measure of r_{SC} represents noise
497 correlations rather than signal correlations because the responses used in this analysis were
498 always to an identical visual stimulus. For Supplementary Figure 1C, we included responses
499 from stimulus presentations 2 though 10 from trials that ended with either a hit, miss or correct
500 catch trial and that were immediately followed by the maintenance of fixation and continuation
501 of the trial (i.e., stimulus presentations where the behavioral response on the subsequent stimulus
502 presentation was not a saccade). We z-scored responses as a function of the stimulus presentation
503 number in each trial and then pooled data across stimulus presentations before calculating noise
504 correlations. Results did not qualitatively change if we did not perform this z-score procedure.

For Supplementary 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.

Decoding

We focused our decoding analyses (Figures 2, 3 and 5) on trials in which 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 have 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 stimulus) only from data sets for which there were at least 10 trials in each attention condition and at least three hits and three misses in each condition. These criteria led us to exclude one data set. 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 Figure 2. We began by constructing a matrix of MT responses for each attention condition: 'MT responses' (a # MT neurons by 2*# 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 'MT responses' (a # MT neurons by 2*# 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*# 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' (a # MT neurons by 1*# trials matrix of MT responses to the change stimulus on the relevant trials) and 'choice' (a 1 by 1*# trials vector of zeros for change 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'). The SC decoder was performed using two matrices for each attention condition: the responses during original stimulus presentations 'MT responses' (a # MT neurons by 1*# trials matrix of MT responses to the original stimulus on the relevant trials) and 'SC responses' (a # SC neurons by 1*# trials matrix of SC responses to the original stimuli on the relevant trials). We refer to this final decoder as an '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 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 did fit, we performed PCA on the MT and SC responses to find the first 10 PCs 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 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 3) 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). 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 conditions and the weights calculated from the other (Figures 3 and 4). Because the responses of visual neurons are non-unique and because our behavioral response is binary, the

weights found with our linear decoding methods are non-unique[23,30]. It is therefore not informative to make direct comparisons of the weights across conditions. Instead, we borrowed the spirit of the analyses in a recent study[30] 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 sensory responses from one attention condition and weights from the other (Figures 3 and 4).

For the decoding analysis in Figure 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 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 by 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 (Figure 4A and 4B) were identical manner to the MT data described above. This dataset consisted of multineuron recordings using

Utah arrays placed in both hemispheres of V4 during 37 experimental sessions in two animals, the details of which are described in[5]. Data from each hemisphere was treated separately in the decoding analyses, so each session contributes two data points for each analysis (gray lines in Figure 4A). The details of the contrast discrimination task used in Figure 4C required a different form of the Stimulus decoder. This dataset consisted of multineuron recordings using Utah arrays placed in both hemispheres of V4 during 17 experimental sessions in two animals. The details of this experiment have been previously described[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 (Figure 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 data set, data from each hemisphere was treated separately.

Code availability

Custom Matlab code is available upon reasonable request to the authors.

Acknowledgements
We are grateful to John Maunsell for granting permission for the use of the V4 data, to Karen McKracken for providing technical assistance, and to Faisal Baqai, Matt Getz, Jay Hennig, Chengcheng Huang, Amy Ni and John Maunsell for 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.

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

Competing interests

The authors declare no competing interests

References:

1 Carrasco, M. Visual attention: the past 25 years. *Vision Res* **51**, 1484-1525, doi:10.1016/j.visres.2011.04.012 (2011).

2 Maunsell, J. H. R. Neuronal Mechanisms of Visual Attention. *Annu Rev Vis Sci* **1**, 373-391, doi:10.1146/annurev-vision-082114-035431 (2015).

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

4 Krauzlis, R. J., Lovejoy, L. P. & Zénon, A. Superior Colliculus and Visual Spatial Attention. *Annual Review of Neuroscience* **36**, 165-182, doi:10.1146/annurev-neuro-062012-170249.Superior (2013).

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

6 Ruff, D. A. & Cohen, M. R. Attention can either increase or decrease spike count correlations in visual cortex. *Nature Neuroscience* **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, doi:10.1016/j.neuron.2015.09.034 (2015).

9 Ruff, D. A. & Cohen, M. R. Attention Increases Spike Count Correlations between Visual Cortical Areas. *Journal of Neuroscience* **36**, 7523-7534, doi:10.1523/JNEUROSCI.0610-16.2016 (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 U S A* **114**, E4085-E4094, doi:10.1073/pnas.1619857114 (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. *Annual review of neuroscience* **28**, 157-189, doi:10.1146/annurev.neuro.26.041002.131052 (2005).

13 Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annual review of neuroscience* **21**, 227-277, doi:10.1146/annurev.neuro.21.1.227 (1998).

14 Gandhi, N. J. & Katnani, H. A. motor functions of the superior colliculus. *Annual Review of Neuroscience* **34**, 205-231, doi:10.1016/j.micinf.2011.07.011.Innate (2011).

15 Sparks, D. L. Conceptual issues related to the role of the superior colliculus in the control of gaze. *Current opinion in neurobiology* **9**, 698-707, doi:10.1016/S0959-4388(99)00039-2 (1999).

16 Wurtz, R. H. & Albano, J. E. Visual-motor function of the primate superior colliculus. *Annual review of neuroscience* **3**, 189-226, doi:10.1146/annurev.ne.03.030180.001201 (1980).

17 Horwitz, G. D. & Newsome, W. T. Separate Signals for Target Selection and Movement Specification in the Superior Colliculus. *Science* **284**, 1158-1161, doi:10.1126/science.284.5417.1158 (1999).

18 Crapse, T. B., Lau, H. & Basso, M. A. A Role for the Superior Colliculus in Decision Criteria. *Neuron* **97**, 181-194 e186, doi:10.1016/j.neuron.2017.12.006 (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. *Journal of neurophysiology* **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, doi:10.1523/JNEUROSCI.0192-16.2016 (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. *Nature Neuroscience* **14**, 811-819 (2011).
25 Pooresmaeili, A., Poort, J. & Roelfsema, P. R. Simultaneous selection by object-based attention in visual and frontal cortex. *Proceedings of the National Academy of Sciences* **111**, 6467-6472, doi:10.1073/pnas.1316181111 (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, doi:10.1523/JNEUROSCI.2044-15.2016 (2016).
27 Moreno-Bote, R. *et al.* Information-limiting correlations. *Nature Neuroscience* **17**, 1410-1417 (2014).
28 Kanitscheider, I., Coen-Cagli, R. & Pouget, A. Origin of information-limiting noise correlations. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6973-6982, doi:10.1073/pnas.1508738112 (2015).
29 Kohn, A., Coen-cagli, R., Kanitscheider, I. & Pouget, A. Correlations and Neuronal Population Information. *Annual Review of Neuroscience* **39**, 237-256, doi:10.1146/annurev-neuro-070815-013851 (2016).
30 Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical Areas Interact through a Communication Subspace. *Neuron*, doi:10.1016/j.neuron.2019.01.026 (2019).
31 Zénon, A. & Krauzlis, R. Attention deficits without cortical neuronal deficits. *Nature* **489**, 434-437 (2012).
32 Herrero, J. L., Gieselmann, M., Sanayei, M. & Thiele, A. Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron* **78**, 729-739 (2013).
33 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. *Nature neuroscience* **17**, 1003-1011, doi:10.1038/nn.3742 (2014).
34 Luo, Thomas Z. & Maunsell, John H. R. Neuronal Modulations in Visual Cortex Are Associated with Only One of Multiple Components of Attention. *Neuron* **86**, 1182-1188, doi:10.1016/j.neuron.2015.05.007 (2015).

35 Luo, T. Z. & Maunsell, J. H. R. Attentional Changes in Either Criterion or Sensitivity Are Associated with Robust Modulations in Lateral Prefrontal Cortex. *Neuron* **97**, 1382-1393 e1387, doi:10.1016/j.neuron.2018.02.007 (2018).
36 Nandy, Anirvan S., Nassi, Jonathan J. & Reynolds, John H. Laminar Organization of Attentional Modulation in Macaque Visual Area V4. *Neuron* **93**, 235-246, doi:10.1016/j.neuron.2016.11.029 (2016).
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, doi:10.1152/jn.00343.2016 (2016).
38 Bisley, J. W. & Goldberg, M. E. Neuronal activity in the lateral intraparietal area and spatial attention. *Science* **299**, 81-86 (2003).
39 Herrington, T. M. & Assad, J. A. Neural activity in the middle temporal area and lateral intraparietal area during endogenously cued shifts of attention. *The Journal of neuroscience* **29**, 14160-14176, doi:10.1523/JNEUROSCI.1916-09.2009 (2009).
40 Egeth, H. E. & Yantis, S. Visual attention: Control, representation, and time course. *Annual Review of Psychology* **48**, 269-297 (1997).
41 Huang, C. *et al.* Circuit Models of Low-Dimensional Shared Variability in Cortical Networks. *Neuron* **101**, 337-348, doi:10.1016/j.neuron.2018.11.034 (2019).
42 Kanashiro, T., Ocker, G. K., Cohen, M. R. & Doiron, B. Attentional modulation of neuronal variability in circuit models of cortex. *Elife* **6**, doi:10.7554/eLife.23978 (2017).
43 Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nature Reviews Neuroscience* **7**, 358-366 (2006).
44 Wang, X. J. & Yang, G. R. A disinhibitory circuit motif and flexible information routing in the brain. *Curr Opin Neurobiol* **49**, 75-83, doi:10.1016/j.conb.2018.01.002 (2018).
45 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*, doi:10.1038/s41593-018-0310-2 (2019).
46 Buonomano, D. V. & Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nat Rev Neurosci* **10**, 113-125, doi:10.1038/nrn2558 (2009).
47 Huang, G., Huang, G. B., Song, S. & You, K. Trends in extreme learning machines: a review. *Neural Netw* **61**, 32-48, doi:10.1016/j.neunet.2014.10.001 (2015).
48 Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nature Reviews Neuroscience* **13**, 51-62, doi:10.1038/nrn3136 (2012).
49 Brainard, D. H. The Psychophysics Toolbox. *Spatial vision* **10**, 433-436 (1997).
50 Pelli, D. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision* **10**, 437-442 (1997).
51 Engbert, R. & Kliegl, R. Microsaccades uncover the orientation of covert attention. *Vision Research* **43**, 1035-1045, doi:10.1016/S0042-6989(03)00084-1 (2003).