A general decoding strategy explains the relationship between behavior and correlated variability

Amy M. Ni*[1,2], Chengcheng Huang[1,2,3], Brent Doiron[2,3], and Marlene R. Cohen[1,2]

[1]Department of Neuroscience, University of Pittsburgh, Pittsburgh, PA 15260, USA
[2]Center for the Neural Basis of Cognition, Pittsburgh, PA 15260, USA
[3]Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA

*Correspondence: amn75@pitt.edu

ABSTRACT

Increases in perceptual performance correspond to decreases in the correlated variability of sensory neuron responses. No sensory information decoding mechanism has yet explained this relationship. We hypothesize that when observers must respond to a stimulus change of any magnitude, decoders prioritize *generality*: a single set of neuronal weights to decode any stimulus response. Our mechanistic circuit model supports that a general decoding strategy explains the inverse relationship between perceptual performance and V4 correlated variability observed in two rhesus monkeys performing a visual attention task. Further, based on the recorded V4 population responses, a monkey's decoding mechanism was more closely matched the more broad the range of stimulus changes used to compute a sensory information decoder. These results support that observers use a general sensory information decoding strategy based on a single set of decoding weights, capable of decoding neuronal responses to the wide variety of stimuli encountered in natural vision.

INTRODUCTION

Many studies have demonstrated that increases in perceptual performance correspond to decreases in the correlated variability of the responses of sensory neurons to repeated presentations of the same stimulus (Cohen & Maunsell, 2009; 2011; Gregoriou et al., 2014; Gu et al., 2011; Herrero et al., 2013; Luo & Maunsell, 2015; Mayo & Maunsell, 2016; Mitchell et al., 2009; Nandy et al., 2017; Ni et al., 2018; Ruff & Cohen, 2014a; 2014b; 2016; 2019; Verhoef & Maunsell, 2017; Yan et al., 2014; Zénon & Krauzlis, 2012). We recently found that the axis in neuronal population space that explains the most correlated variability (which is often quantified as noise correlations or spike count correlations; Cohen & Kohn, 2011; Nirenberg & Latham, 2003) explains virtually all of the choice-predictive signals in visual area V4 (Ni et al., 2018).

These observations comprise a paradox. The shared variability of population activity in visual cortex occupies a low-dimensional subset of the full neuronal population space (Ecker et al., 2014; Goris et al., 2014; Huang et al., 2019; Kanashiro et al., 2017; Lin et al., 2015; Rabinowitz et al., 2015; Semedo et al., 2019; Williamson et al., 2016). Yet, recent theoretical work shows that neuronal population decoders that extract the maximum amount of sensory information for the specific task at hand can easily ignore correlated noise that is restricted to a small number of dimensions, particularly if that noise does not corrupt the dimensions of neuronal population space that are most informative about the stimulus (Kanitscheider et al., 2015b; Moreno-Bote et al., 2014; for review, see Kohn et al., 2016).

Here, we test a hypothesis that addresses this paradox: Even in the context of a simple, well-learned laboratory task, downstream decoders of population activity use a *general* decoding strategy: one set of neuronal population decoding weights to extract sensory information about any visual stimulus. If an observer's decoder were designed to decode a wide variety of stimuli, their perceptual performance might be inextricably linked to correlated variability, which depends on neuronal tuning similarity for many stimulus features (Cohen & Kohn, 2011).

We tested this idea using a laboratory version of a real-life scenario: The observer must report that a stimulus changed, regardless of the magnitude of the change. For example, an observer might need to report when a door opens but not by how much, or when a light turns on but not its brightness. We selected this basic case because: 1) in natural environments, it is often the case that an observer cares *if* a stimulus changes as opposed to *how much* it changes, and 2) many of the studies that found a relationship between behavioral performance and correlated variability used a laboratory version of this scenario (i.e., a change-detection task; Cohen & Maunsell, 2009; 2011; Herrero et al., 2013; Luo & Maunsell, 2015; Mayo & Maunsell, 2016; Nandy et al., 2017; Ni et al., 2018; Ruff & Cohen, 2016; 2019; Verhoef & Maunsell, 2017; Yan et al., 2014; Zénon & Krauzlis, 2012).

We hypothesize that in this common scenario, the observer uses a general decoding strategy: one set of neuronal weights to decode sensory neuron population responses to any stimulus change. With this strategy, a downstream brain area would not need to change how it weights the influence of a given sensory neuron based on the specific stimulus change detection required. While greater perceptual precision may be achieved using a specific decoding strategy that uses a different set of neuronal weights to decode each stimulus change, a general decoding strategy may prioritize flexibility in the face of the rapidly fluctuating stimulus conditions that may be encountered in the natural world.

RESULTS

A behavioral framework for studying the general decoder hypothesis

We designed a behavioral task with two main components that allowed us to test the hypothesis that observers use a general decoder when tasked with responding to a stimulus change of any size. First, two rhesus monkeys performed a change-detection task with multiple potential stimulus changes (**Fig. 1a**; different aspects of these data were presented previously, Ni et al., 2018). Two Gabor stimuli of the same orientation flashed on and off until, at a random time, the orientation of one of the stimuli changed. The changed orientation was randomly selected from five options (**Fig. 1b**). The monkey could not predict which orientation change was to be detected on any given trial and was rewarded for responding to any orientation change.

Second, we made a manipulation designed to create a larger dynamic range of perceptual performance. We modulated perceptual performance by manipulating visual attention within the task (**Fig. 1a**), using a classic Posner cueing paradigm (Posner, 1980). We recorded from a population of V4 neurons (**Fig. 1c**) to measure correlated variability changes due to this attention manipulation. Cued trials were collected for all five change amounts and uncued trials were collected mainly for the median change amount (**Fig. 1b**). Our attention analyses focused on this median change amount, for which we had both cued and uncued trials.

Figure 1

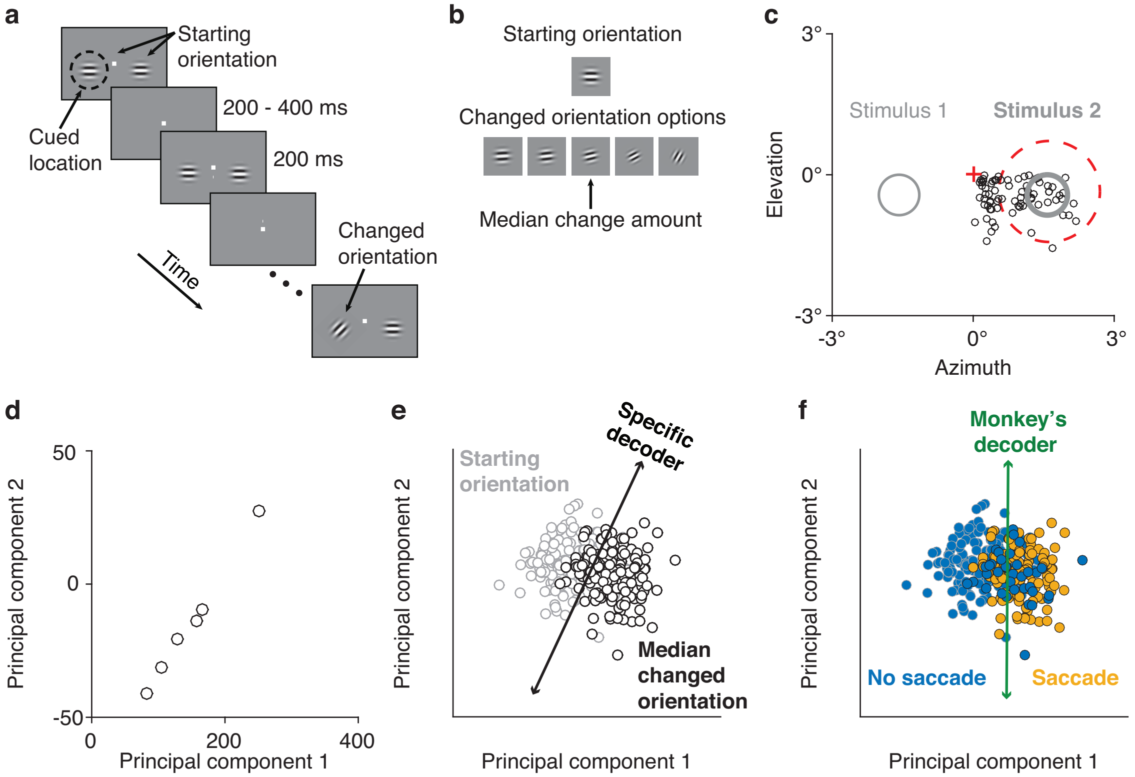

Figure 1. Electrophysiological data collection and decoders. (**a**) Visual change-detection task with cued attention. After the monkey fixated the central spot, two Gabor stimuli synchronously flashed on (200 ms) and off (randomized 200-400 ms period) at the starting orientation until, at a random time, the orientation of one stimulus changed. To manipulate attention, the monkey was

cued in blocks of 125 trials as to which of the two stimuli would change in 80% of the trials in the block, with the change occurring at the uncued location in the other 20%. (**b**) A cued changed orientation was randomly assigned per trial from five potential orientations. An uncued changed orientation was randomly either the median (20 trials) or largest change amount (5 trials). To compare cued to uncued changes, median orientation change trials were analyzed. (**c**) The activity of a neuronal population in V4 was simultaneously recorded using microelectrode arrays. Plotted for Monkey 1: the location of Stimulus 2 (thick gray circle) relative to fixation (red cross) overlapped the receptive field (RF) centers of the recorded units (black circles). A representative RF size is illustrated (red dashed circle). Only orientation changes at the RF location were analyzed. Stimulus 1 was located in the opposite hemifield (thin gray circle). (**d**) Example session plot of the first versus second principal component (PC) of the V4 population responses to each of the six orientations presented in the session. Though the brain may use nonlinear decoding methods, the neuronal population representations of the small range of orientations tested per session were reasonably approximated by a line; thus, linear methods were sufficient to capture decoder performance. See **Fig. 2, 3** for model analyses of the full range of orientations. (**e**) Schematic of specific decoder. Neuronal weights were determined using linear regression to best differentiate the V4 neuronal population responses (first and second PCs shown for illustrative purposes) to the median changed orientation from the responses to the starting orientation presented immediately before it. (**f**) Schematic of monkey's decoder. Neuronal weights were determined for the same neuronal responses as in (**e**), but weights were instead optimized to best differentiate the V4 responses when the monkey made a saccade (indicating it detected the orientation change) from when the monkey did not choose to make a saccade.

Strategy for testing the general decoder hypothesis

We hypothesized that the monkey's behavioral choices on this task reflected a general decoding strategy. We set out to test this in three steps.

1) Use electrophysiological recordings to compare how attention affects the amount of sensory information extracted from a neuronal population about a specific stimulus change when using different decoding strategies (**Fig. 1d-f**). Prediction: The effect of attention on the *monkey's choice decoder* (**Fig. 1f**) will not be matched by the effect of attention on a *specific decoder* that maximizes the amount of extracted sensory information for the specific stimulus change (**Fig. 1e**), with far larger attentional effects with the monkey's decoder.
2) Use a circuit model of attention to generate a large data set with an experimentally unfeasible number of stimulus conditions with which to compare the electrophysiological *monkey's decoder* (**Fig. 1f**) to a modeled ideal *general* or *specific decoder*. Predictions: 1) the modeled specific decoder will be similar to the physiological specific decoder (which would validate the model), and 2) the effects of attention on the monkey's decoder will more closely match the modeled general than specific decoder.
3) Use the collected electrophysiological responses to five different stimulus changes to compare increasingly more-general decoders to the monkey's decoder. Prediction: the more general the decoder, the more its performance will be correlated with that of the monkey's decoder.

Testing decoder hypotheses using a mechanistic circuit model

When designing our behavioral task, we made the decision to limit the number of orientation changes in the interest of using the limited number of trials to obtain repeated trials of the same conditions. However, testing the hypothesis that monkeys employ a general decoding strategy would benefit from the ability to calculate a general decoder of all orientations.

We therefore modeled responses to all possible orientations by extending our previously published excitatory/inhibitory cortical network model of attention (Huang et al., 2019). We extended the three-layer model of V1 and V4 neuronal populations (Huang et al., 2019; 2020) to mimic realistic orientation tuning and organization in the V1 layer (**Fig. 2a**). We calculated the effects of attention (**Fig. 2b, c**) on a modeled specific decoder and on a modeled general decoder that used the same set of neuronal weights to estimate all orientations. The model well captured our recorded attentional changes in V4 firing rates (**Fig. 2d**), correlated variability (**Fig. 2e**), and covariance eigenspectrum (**Fig. 2f**). The model allowed us to test larger modeled ranges of those values than those we recorded (**Fig. 2d-f**).

Figure 2

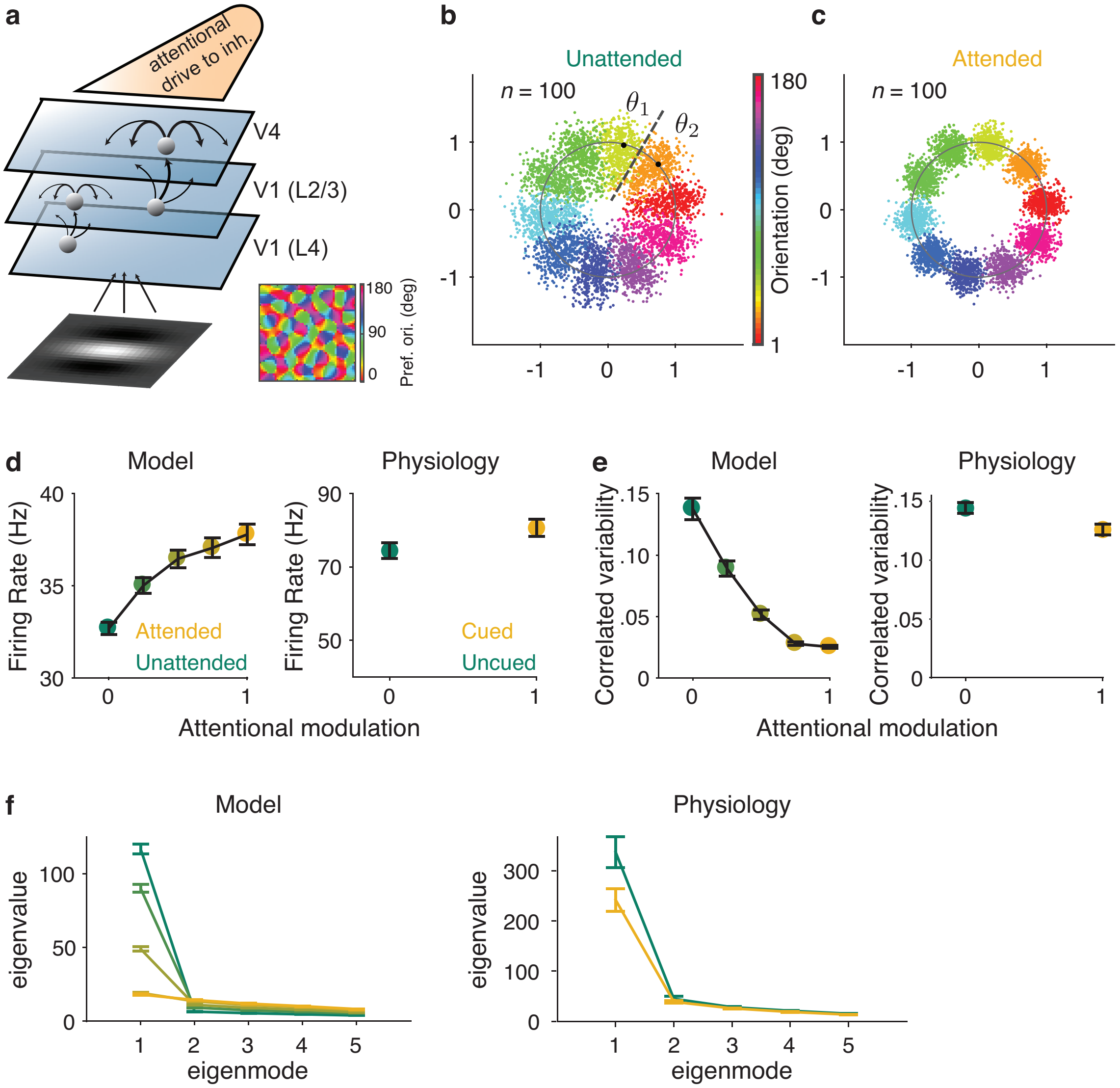

Figure 2. Mechanistic circuit model of attention effects. (**a**) Schematic of an excitatory and inhibitory neuronal network model of attention (Huang et al., 2019) that extends the three-layer, spatially ordered network to include the orientation tuning and organization of V1. The network models the hierarchical connectivity between layer 4 of V1, layers 2 and 3 of V1, and V4. In this model, attention depolarizes the inhibitory neurons in V4 and increases the feedforward projection strength from layers 2 and 3 of V1 to V4. (**b, c**) To compute a general decoder optimized for all orientations, we first mapped the *n*-dimensional neuronal activity of our model to a 2-dimensional space (a ring). Each dot represents the neuronal activity of the simulated population on a single trial and each color represents the trials for a given orientation. The fluctuations of the neurons that are proportional to their firing rates are mapped to the radial direction. These fluctuations are more elongated in the (**b**) unattended state than in the (**c**) attended state. (**d-f**) Comparisons of the modeled versus electrophysiologically recorded effects of attention on V4 population activity: (**d**) firing rates of excitatory neurons increased, (**e**) correlated variability decreased, and (**f**) as illustrated with the first five largest eigenvalues of the shared component of the spike count covariance matrix from the V4 neurons, attention largely reduced the eigenvalue of the first mode. Attentional state denoted by marker color for model (yellow: most attended; green: least attended) and electrophysiological data (yellow: cued; green: uncued). For model: 30 samplings of $n = 50$ neurons. Monkey 1 data illustrated for electrophysiological data: $n = 46$ days of recorded data. SEM error bars.

The monkey's strategy was most closely matched to the general decoder

First, we compared the recorded attentional effects on the specific versus monkey's decoders (**Fig. 3a**). Manipulating attention affected the performance of each decoder differently: The performance of the specific decoder was little affected by attention, while that of the monkey's decoder was strongly affected by attention.

The lack of attentional effect on the specific decoder (**Fig. 3a**) prompted us to compare the electrophysiological data to the modeled data. First, we compared the attentional effects on the modeled specific decoder (**Fig. 3b**) to those on the physiological specific decoder (**Fig. 3a**). The performance of the modeled specific decoder was similarly little affected by attention.

Thus, we tested the general decoder hypothesis by comparing the attentional effects on the modeled general decoder (**Fig. 3b**) to those on the monkey's decoder (**Fig. 3a**). The performance of the general decoder was similarly strongly affected by attention. In sum, the monkey's decoding strategy was most qualitatively matched to the general decoder.

We next tested the crux of our hypothesis: that a general decoding strategy underlies the oft-reported relationship between behavioral performance and correlated variability (for review, see Ruff et al., 2018). In the physiological data, the performance of the monkey's decoder was more strongly related to correlated variability than the performance of the specific decoder (**Fig. 3c**). We found that the performance of the modeled general decoder was also more strongly related to correlated variability than the performance of the modeled specific decoder (**Fig. 3d**). To summarize the model's findings, the general decoder matched both the large effect of attention on the monkey's choice decoder (**Fig. 3a, b**) and the relationship between the monkey's choices and correlated variability (**Fig. 3c, d**).

Finally, we used the physiological responses collected for a limited number of orientation changes to test increasingly more-general decoders to the monkey's decoder. The more general the decoder, the more its performance matched that of the monkey's decoder (**Fig. 3e**).

Figure 3

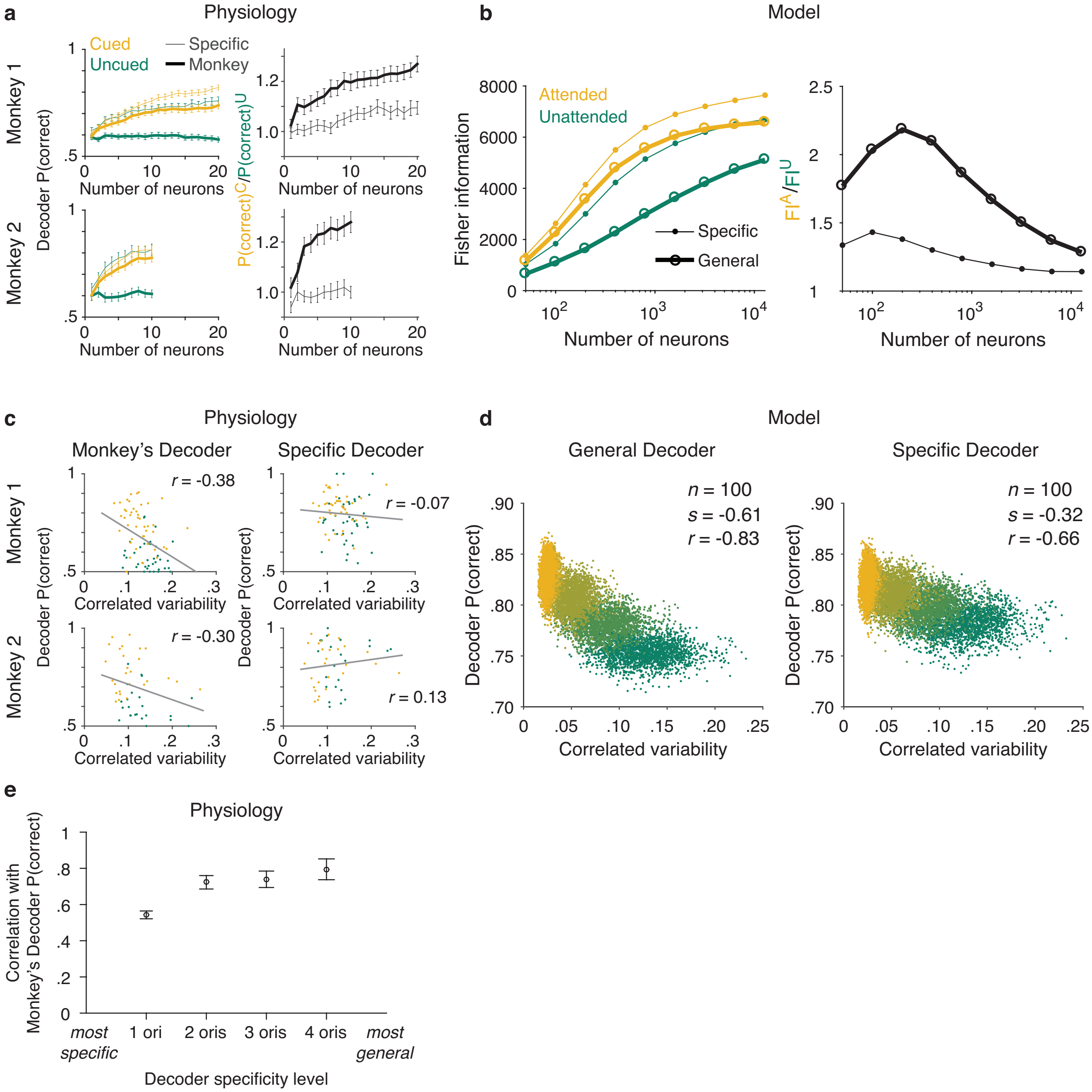

Figure 3. The monkey's strategy was most closely matched to the general decoder. (**a**) Physiological data for Monkey 1 and Monkey 2: the effect of attention on decoder performance was larger for the monkey's decoder than for the specific decoder. Left plots: decoder performance (y-axis; leave-one-out cross-validated proportion of correctly identified orientation: starting vs. median changed orientation) for each neuronal population size (x-axis) is plotted for the specific (thin lines) and monkey's (thick lines) decoders in the cued (yellow) and uncued (green) attention conditions. Right plots: the ratio of the decoder performance in the cued versus uncued conditions is plotted for each neuronal population size. SEM error bars (Monkey 1: *n* = 46 days; Monkey 2: *n* = 28 days). (**b**) Modeled data: the effect of attention on decoder performance was larger for the general decoder than for the specific decoder. The specific

decoder used weights based on the *n*-dimensional discrimination of two orientations to test the decoder's ability to discriminate those two orientations. The general decoder used weights based on all of the orientations in the ring (**Fig. 2b, c**) but, like the specific decoder, was also tested on the 2-dimensional discrimination of the two orientations. Left plot: the inverse of the variance of the estimation of theta (y-axis; equivalent to linear Fisher information for the specific decoder) for each neuronal population size (x-axis) is plotted for the specific decoder (small markers; **Eq. 1**, see **Methods**) and for the general decoder (large markers; **Eq. 3**, see **Methods**) in the attended (yellow) and unattended (green) conditions. Right plot: the ratio of Fisher information in the attended versus unattended conditions is plotted for each neuronal population size. (**c**) Physiological data for Monkeys 1 and 2: the performance of the monkey's decoder was more related to mean correlated variability (left plots; gray lines of best fit; Monkey 1 Pearson's correlation coefficient: $n = 86$, or 44 days with two attention conditions plotted per day and two data points excluded – see **Methods**, $r = -0.38$, $p = 5.9 \times 10^{-4}$; Monkey 2: $n = 54$, or 27 days with two attention conditions plotted per day, $r = -0.30$, $p = 0.03$) than that of the specific decoder (right plots; Monkey 1 Pearson's correlation coefficient: $r = -0.07$, $p = 0.53$; Monkey 2: $r = 0.13$, $p = 0.36$). For both monkeys, the correlation coefficients associated with the two decoders were significantly different from each other (Williams' procedure; Monkey 1: $t = 3.7$, $p = 2.3 \times 10^{-4}$; Monkey 2: $t = 3.2$, $p = 1.4 \times 10^{-3}$). (**d**) Modeled data: the performance of the general decoder was more related to mean correlated variability (left plot) than that of the specific decoder (right plot; number of neurons fixed at 100 and attentional state denoted by marker color, yellow: most attended, green: least attended). The model allowed comparisons to a wider range of correlated variability values (also see **Fig. 2e**), likely explaining the statistically significant relationship between correlated variability and performance of the specific decoder observed for the modeled specific decoder only (right plot), and not for the physiological specific decoder (**Fig. 3c**, right plots). (**e**) Physiological data from both monkeys combined: the more general the decoder (x-axis; number of orientation changes used to determine the sensory information decoder, with the decoder that best differentiated the V4 responses to the starting orientation from those to one changed orientation on the far left, and the decoder that best differentiated V4 responses to the starting orientation from those to four different changed orientations on the far right), the more correlated its performance to the performance of the monkey's decoder (y-axis). SEM error bars (see **Methods** for *n* values).

DISCUSSION

Our results suggest that the relationship between behavior and correlated variability is explained by our hypothesis that observers use a general strategy for decoding arbitrary stimulus changes. Our modeled general decoder explained both the effect of attention on the monkey's choice decoder and the relationship between the monkey's choice decoder and correlated variability. Further, based on the electrophysiological data we found that the more general the decoder (the more orientation change amounts used to determine the decoder weights) the more its performance was correlated with that of the monkey's decoder. Together, these results support the hypothesis that observers use a general decoding strategy in scenarios that require flexibility to changing stimulus conditions.

Our study also demonstrates the utility of combining electrophysiological and circuit modeling approaches to studying neural coding. Our model mimicked the correlated variability and effects of attention in our physiological data. Using a circuit model allowed us to perform a very large number of trials for many different orientations, allowing us to test a true general decoder for orientation. The model also allowed us to test large neuronal population sizes available to the decoder (**Fig. 3b**). Finally, the model allowed us to test a much wider range of correlated variability values than those collected in our electrophysiological data (**Fig. 2e**), which is important for making inferences about the large number of neurons that are likely involved in any behavioral process. Our physiological dataset supported the model's results by allowing us to address a specific hypothesis: the more general the stimulus information decoder, the more its performance should match that of the monkey's decoder (**Fig. 3e**).

A general decoding strategy in the face of unpredictable stimuli

We tested the general decoder strategy in the context of a change-detection task because this type of task was used in many of the studies that reported a relationship between perceptual performance and correlated variability (Cohen & Maunsell, 2009; 2011; Herrero et al., 2013; Luo & Maunsell, 2015; Mayo & Maunsell, 2016; Nandy et al., 2017; Ni et al., 2018; Ruff & Cohen, 2016; 2019; Verhoef & Maunsell, 2017; Yan et al., 2014; Zénon & Krauzlis, 2012).

However, a general decoding strategy may explain observations in studies that use a variety of behavioral and stimulus conditions. Studies using a variety of tasks have also demonstrated a relationship between perceptual performance and correlated variability. These tasks include heading (Gu et al., 2011), orientation (Gregoriou et al., 2014), and contrast (Ruff & Cohen, 2014a; 2014b) discrimination tasks, in which the observer must respond to only stimulus value or compare stimulus values. Interestingly, some studies of discrimination tasks suggest that the relationship between perceptual performance and correlated variability cannot be explained by a specific decoding strategy that maximizes the amount of sensory information extracted for the task (Clery et al., 2017; Gu et al., 2011).

On the other hand, other studies of perceptual performance have found that observers can achieve high levels of perceptual precision under certain circumstances (Burgess et al., 1981; Kersten, 1987). Such studies suggest that decoding strategies that maximize the amount of extracted sensory information might be used in certain situations. Further tests of decoding strategies in a variety of stimulus conditions and behavioral contexts will be necessary to determine when sensory information decoding prioritizes accuracy, flexibility, or other behavioral advantages.

General decoders of all features would be inextricably linked to correlated variability

Our results address a paradox in the literature. The idea that a specific decoding strategy, in which different sets of neuronal weights are used to decode different stimulus changes, cannot easily explain the relationship between behavioral performance and correlated variability is supported by electrophysiological (Clery et al., 2017; Haefner et al., 2013; Jin et al., 2019; Ni et al., 2018; Ruff & Cohen, 2019; for review, see Ruff et al., 2018) and theoretical evidence (Abbott & Dayan, 1999; Averbeck et al., 2006; Kanitscheider et al., 2015b; Moreno-Bote et al., 2014; for review, see Kohn et al., 2016). Correlated variability is restricted to a small number of dimensions (Ecker et al., 2014; Goris et al., 2014; Huang et al., 2019; Kanashiro et al., 2017; Lin et al., 2015; Rabinowitz et al., 2015; Semedo et al., 2019; Williamson et al., 2016). Specific decoders of neuronal population activity can easily ignore changes along one or few dimensions (Kohn et al., 2016; Moreno-Bote et al., 2014). In other words, correlated variability changes in one dimension are easy to ignore: Observers should simply use one of the many other possible combinations of neuronal responses to guide their perceptual performance.

The general decoder hypothesis offers a resolution to this paradox. A fully general decoder of stimuli that vary along many feature dimensions would be one whose neuronal weights depend on the tuning properties of the neurons to all stimulus features to which they are selective. For example, two V4 neurons may both prefer vertical orientations. But, if they also share a color tuning preference for red, a large response from both neurons might indicate vertical orientation, the color red, or a combination of both features. A fully general decoder would need to resolve this discrepancy by choosing weights for these and other neurons that take not only their tuning for orientation but also their tuning for color into account.

Therefore, the weights of a fully general decoder would depend on the tuning of all neurons to all of the stimulus features to which they are selective. A large number of studies have shown that correlated variability also depends on tuning similarity for all stimulus features (for review, see Cohen & Kohn, 2011). The implication is that the decoding weights for a fully general decoder would depend on exactly the same properties as correlated variability.

The hypothesis that such a truly general decoder explains the relationship between perceptual performance and correlated variability is suggested by our finding that the modeled general decoder for orientation was more strongly related to correlated variability than the modeled specific decoder (**Fig. 3d**). However, direct tests of this idea would be needed to determine if this decoding strategy is used in the face of multiple changing stimulus features. Further, such tests would need to consider alternative hypotheses for how sensory information is decoded when observers observe multiple aspects of a stimulus (Berkes et al., 2009; Deneve, 2012; Lorteije et al., 2015).

In conclusion, the findings of this study support the usefulness of a framework that relates sensory information decoding to behavior (for review, see Panzeri et al., 2017). By first determining the decoder that guided each monkey's behavioral choices, we were able to compare the monkey's decoder to modeled specific and general decoders to test our hypothesis. These results demonstrate that constraining analyses of neuronal data by behavior can provide important insights into the neurobiological mechanisms underlying perception and cognition.

METHODS

Electrophysiological recordings. The subjects were two adult male rhesus monkeys (*Macaca mulatta*, 8 and 10 kg). All animal procedures were approved by the Institutional Animal Care and Use Committees of the University of Pittsburgh and Carnegie Mellon University. Different aspects of these data were presented previously (Ni et al., 2018). We recorded extracellularly from single units and sorted multiunit clusters (the term "unit" refers to either; see Ni et al., 2018) in V4 of the left hemisphere using chronically implanted 96-channel microelectrode arrays (Blackrock Microsystems) with 1 mm long electrodes. We performed all spiking sorting manually using Plexon's Offline Sorter (version 3.3.5, Plexon).

We only included a recorded unit if its stimulus-driven firing rate was both greater than 10 Hz and significantly higher than the baseline firing rate (baseline calculated as the firing rate in the 100 ms window immediately prior to to the onset of the first stimulus per trial; two-sided Wilcoxon signed rank test: $p < 10^{-10}$). The population size of simultaneously recorded units was 8-45 units (mean 39) per day for Monkey 1 and 7-31 units (mean 19) per day for Monkey 2.

Behavioral task. The monkeys performed a change-detection task (**Fig. 1a**; Cohen & Maunsell, 2009) with multiple orientation change options (**Fig. 1b**) and cued attention (Posner, 1980) while we recorded electrophysiological data. We presented visual stimuli on a CRT monitor (calibrated to linearize intensity; 1,024 × 768 pixels; 120 Hz refresh rate) placed 52 cm from the monkey, using custom software written in MATLAB (Psychophysics Toolbox; Brainard, 1997; Pelli, 1997). We monitored each monkey's eye position using an infrared eye tracker (Eyelink 1000; SR Research) and recorded eye position, 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 Ripple hardware.

A trial began when a monkey fixed its gaze on a small, central spot on the video display while two peripheral Gabor stimuli (one overlapping the RFs of the recorded neurons, the other in the opposite visual hemifield; **Fig. 1c**) synchronously flashed on (for 200 ms) and off (for a randomized period between 200-400 ms) at the same starting orientation until at a random, unsignaled time the orientation of one of the stimuli changed. The monkey received a liquid reward for making a saccade to the changed stimulus within 400 ms of its onset.

Attention was cued in blocks of trials, with each block preceded by 10 instruction trials that cued one of the two stimulus locations by only presenting stimuli at that location. Each block consisted of approximately 125 orientation-change trials. In each block, the orientation change occurred at the cued location in 80% of the change trials and at the uncued location in 20% of the change trials. Catch trials were intermixed, in which no orientation change occurred within the maximum of 12 stimulus presentations. In catch trials, the monkeys were rewarded for maintaining fixation. Trial blocks with attention cued to the left hemifield location or to the right hemifield location were presented in alternating order within a recording day.

The changed orientation at the cued location was randomly selected per trial from one of five changed orientations (with the constraint of required average numbers of presentations per changed orientation per block; **Fig. 1b**) such that the monkeys could not predict which orientation change amount was to be detected on any given trial. The changed orientation at the uncued location was randomly either the median (20 trials per block) or the largest orientation change amount (5 trials per block). Uncued changes were collected mainly for the median change amount to maximize the number of uncued trials collected for one change amount. All analyses of the effects of attention analyzed the cued versus uncued median change amounts.

The size, location, and spatial frequency of the Gabor stimuli were fixed across all recordings. These parameters were set to maximize the neuronal responses and were determined using a receptive field mapping task prior to recording the data presented here. The orientation of all stimuli before the orientation change (the starting orientation; **Fig. 1a, b**) was identical within each day of recording but changed by 15° between days. The five changed orientation options (**Fig. 1b**) also changed between days, to maintain the task at approximately the same level of difficulty across days. If they changed within a day (across different trial blocks), again to maintain a consistent level of task difficulty, they were binned for analysis based on their log distribution.

Electrophysiological data analysis. The data presented are from 46 days of recording for Monkey 1 and 28 days of recording for Monkey 2. Instruction trials were not included in any analyses. Only trials in which the orientation changes occurred at the RF location (**Fig. 1c**) and catch trials were analyzed (see below for specific inclusions per analysis). The first stimulus presentation of each trial was excluded from all analyses to minimize temporal non-stationarities due to adaptation.

Firing rates (**Fig. 2d**), correlated variability (**Fig. 2e, 3c**), and covariance eigenspectrum analyses (**Fig. 2f**) were calculated based on cued orientation-change trials on which the monkey correctly detected the change and on catch trials. From these trials, only the starting orientation stimulus presentations were included in the analyses. The firing rate per stimulus presentation was based on the spike count response between 60-260 ms after stimulus onset to account for V4 latency. These analyses were performed per recording day (such that all stimuli analyzed together were identical). Data were presented as the mean per day (**Fig. 3c**) or across days (**Fig. 2d-f**) per attention condition (cued or uncued).

We defined the correlated variability of each pair of simultaneously recorded units (quantified as noise correlation or spike count correlation; Cohen & Kohn, 2011) as the Pearson's correlation coefficient between the firing rates of the two units in response to repeated presentations of the same stimulus. This measure of correlated variability represents correlations in noise rather than in signal because the visual stimulus was always the same.

For **Fig. 3c**, we compared the Pearson's correlation between the performance of the monkey's decoder and the mean correlated variability per day to the Pearson's correlation between the performance of the specific decoder and correlated variability using Williams' procedure for comparing correlated correlation coefficients (Howell, 2007).

For Monkey 1, two outlier points (uncued trials for each of two days) with correlated variability values greater than 0.35 were excluded from analysis based on the Tukey method (see **Fig. 3c** for the range of included correlated variability values for Monkey 1). For **Fig. 3c**, with the excluded points included, the Pearson's correlation coefficients were qualitatively unchanged: for the monkey's decoder, $n = 88$, or 44 days (see below for data included in decoder analyses) with two attention conditions plotted per day, $r = -0.34$, $p = 1.7 \times 10^{-3}$; for the specific decoder, $r = -0.22$, $p = 0.05$.

V4 population specific decoder. The specific decoder based on the electrophysiologically recorded V4 neuronal population data (**Fig. 3a, c**; Ni et al., 2018; Ruff & Cohen, 2019) was determined per monkey as illustrated in **Fig. 1e** (first and second principal components shown for illustrative purposes only – analyses based on neuronal population firing rates as described below). To avoid artifacts in neuronal firing rates due to eye movements in response to the changed orientation, all V4 population decoder analyses were based on neuronal firing rates during an abbreviated time window: 60-130 ms after stimulus onset.

Neuronal weights were determined using linear regression to best differentiate the population responses to the median changed orientation from the responses to the starting orientation presented immediately before it. The weights were calculated per day and per attention condition based on two matrices: 1) a matrix of firing rate responses with dimensions # V4 neurons x # analyzed stimulus presentations (each median changed orientation stimulus and each starting orientation stimulus presented immediately before it), and 2) a matrix of stimulus orientations with dimensions 1 x # analyzed stimulus presentations (with values of one for median changed orientations and values of zero for starting orientations). The matrix of stimulus orientations was used to categorize each column of stimulus presentation responses.

Decoder performance was quantified as the leave-one-out cross-validated proportion of correctly identified orientations (median changed orientation or starting orientation). For **Fig. 3a**, decoder performance was analyzed per number of neurons (x-axis). Per neuronal population size, the most responsive neurons (ranked by evoked response: stimulus-evoked firing rate minus baseline firing rate) were analyzed. For **Fig. 3c & e**, decoder performance was illustrated for a set number of neurons (Monkey 1: 20 units, Monkey 2: 10 units). The number of neurons analyzed for these plots was selected to maximize the number of included neurons and recording days (Monkey 1: *n* = 44 days, two days with 8 and 19 recorded units excluded; Monkey 2: *n* = 27 days, one day with 7 recorded units excluded).

V4 population monkey's decoder. As illustrated in **Fig. 1f**, the V4 population responses to the same set of stimuli (each median changed orientation stimulus and each starting orientation stimulus presented immediately before it) used to determine the specific decoder were used to determine the monkey's decoder. The monkey's decoder differed only in its classification of those stimuli. Neuronal weights were determined using linear regression to best differentiate the population responses when the monkey made a saccade indicating it detected the orientation change from those when the monkey did not make a saccade (both correctly in response to the starting orientation and incorrectly when the monkey missed the changed orientation). Of the two matrices used to calculate the decoder weights, the matrix of firing rate responses was identical to that used for the specific decoder, and only the second matrix differed: a matrix of monkey choices with dimensions 1 x # analyzed stimulus presentations (with values of one when the monkey made a saccade and of zero when the monkey did not make a saccade). The matrix of monkey's choices was used to categorize each column of stimulus presentation responses.

The performance of the monkey's decoder was quantified exactly as that of the specific decoder. Thus, while the specific and monkey's decoders used different weights, their performance was tested on the same task of correctly identifying stimulus orientation (median changed orientation or starting orientation).

V4 population general decoders. For **Fig. 3e**, we calculated increasingly more-general decoders to compare their performance to that of the monkey's decoder. Only cued orientation-change trials were included, as uncued change trials were collected mainly for one orientation change amount only. The data from both monkeys were illustrated together in **Fig. 3e**.

For the analysis presented in **Fig. 3e**, we avoided the relationship that would be inherent between decoders that were based on the same stimulus presentations by basing only the weights for the monkey's decoder on the median orientation-change trials. Therefore, while the weights of the monkey's decoder were calculated as described above (under **V4 population monkey's decoder**), the weights of all of the other decoders in this analysis were based on trials other than the median orientation-change trials. All of the decoders in this analysis were tasked with identifying stimulus orientation on the same set of stimuli: each second largest orientation change stimulus and each starting orientation stimulus presented immediately before it.

The neuronal weights for the most specific to the most general decoders (**Fig. 3e**, x-axis) were determined using linear regression to best differentiate the population responses to changed orientation stimuli from the responses to the starting orientation presented immediately before them. The weights for the most specific decoder (**Fig. 3e**, '1 ori') best differentiated neuronal responses to the starting orientation from those to the second largest changed orientation (n = 2 decoders; 1 per monkey). This was the '1 ori' decoder because it differentiated responses to the starting orientation from those to one changed orientation.

The '2 oris' decoders best differentiated neuronal responses to the starting orientation from those to two different changed orientations. Each '2 ori' decoder was based on two changed orientations out of the four possibilities: the first, second, fourth, and fifth (max) largest changed orientations (n = 12 decoders; 6 per monkey). As stated above, the median changed-orientation trials were not used to calculate any decoder weights besides the monkey's decoder.

Each '3 oris' decoder was based on three changed orientations out of the four possibilities (n = 8 decoders; 4 per monkey). The '4 oris' decoder was based on all four changed orientations (n = 2 decoders; 1 per monkey).

Data availability. Electrophysiological data analyzed in this manuscript are available at https://pitt.box.com/v/NiRuffAlbertsSymmondsCohen2017.

Code availability. Computer code for all simulations and analysis of the resulting data will be available at https://github.com/hcc11/.

Network model description. The network model is similar to the one in Huang et al. (2019). Briefly, the network consists of three modeled stages: 1) layer (L) 4 neurons of V1, 2) L2/3 neurons of V1, and 3) L2/3 neurons of V4 (**Fig. 2a**). Neurons from each area are arranged on a uniform grid covering a unit square $\Gamma = [-0.5, 0.5] \times [-0.5, 0.5]$. The L4 neurons of V1 are modeled as a population of excitatory neurons, the spikes of which are taken as inhomogeneous Poisson processes with rates determined as below. The L2/3 of V1 and V4 populations are recurrently coupled networks with excitatory and inhibitory neurons. Each neuron is modeled as an exponential integrate-and-fire (EIF) neuron. The connection probability between neurons decays with distance. The network model captures many attention-mediated changes on neuronal responses, such as the reduction of correlated variability within each visual area, increase in correlated variability between visual areas, and the quenching of the low-dimensional shared variability by attention. The network parameters are the same as those used in Huang et al. (2019) except the following. The feedforward projection width from V1(L2/3) to V4 is $\alpha^{(3)}_{\text{ffwd}} = 0.05$. The feedforward strength from V1(L2/3) to V4 is $[J^3_{\text{eF}}, J^3_{\text{iF}}] = \gamma[1, 0.4]$. From the most unattended state to the most attended state (attentional modulation scale from 0 to 1), γ varies from 20 to 23 mV, and the depolarizing current to the inhibitory neurons in V4, μ_i, varies from 0 to 0.5 mV/ms (**Fig. 2**, **Fig. 3b,d**).

The model differs from the previous model (Huang et al., 2019) in the following ways. We modeled the V1(L4) neurons as orientation selective filters with static nonlinearity and Poisson spike generation (Kanitscheider et al., 2015b). The firing rate of each neuron i is $r_i(\theta, t) = [F_i \times \tilde{I}(\theta, t)]_+$, where F_i is a Gabor filter and $\tilde{I}(\theta, t)$ is a Gabor image corrupted by independent noise following the Ornstein-Uhlenbeck process,

$$\tilde{I}(\theta, t) = I(\theta) + \eta(t) \quad \text{and} \quad \tau_n d\eta_i = -\eta_i dt + \sigma_n dW,$$

with $\tau_n = 40$ ms and $\sigma_n = 3.5$. The Gabor filters were normalized such that the mean firing rate of V1(L4) neurons was 10 Hz. Spike trains of V1(L4) neurons were generated as inhomogeneous Poisson processes with rate $r_i(\theta, t)$. The Gabor image is defined on Γ with 25×25 pixels with spatial Gaussian envelope width $\sigma = 0.2$, spatial wavelength $\lambda = 0.6$ and phase $\phi = 0$ (Kanitscheider et al., 2015b, Supp Eq. 6). The Gabor filters of V1(L4) neurons had the same σ, λ and ϕ as the image (Kanitscheider et al., 2015b, Supp Eq. 5). The orientation θ was normalized between 0 and 1. The orientation preference map of L4 neurons in V1 was generated using the formula from Kaschube et al. (2010, Supp Eq. 20) with average column spacing $\Lambda = 0.2$.

Each network simulation was 20 sec long consisting of alternating OFF (300 ms) and ON (200 ms) intervals. During OFF intervals, spike trains of Layer 1 neurons were independent Poisson process with rate $r_X = 5$ Hz. An image with a randomly selected orientation was presented during ON intervals. Spike counts during the ON intervals were used to compute the performance of different decoders and correlated variability. The first spike count in each simulation was excluded. For each parameter condition, the connectivity matrices were fixed for all simulations. The initial states of each neuron's membrane potential were randomized in each simulation. All simulations were performed on the CNBC Cluster in the University of Pittsburgh. All simulations were written in a combination of C and Matlab (Matlab R 2015a, Mathworks). The differential equations of the neuron model were solved using the forward Euler method with time step 0.01 ms.

Network model specific decoder. Let $\mathbf{r}$ be a vector of spike counts from all neurons on a single trial, $\mathbf{f}$ be the tuning curve function, and Σ be the covariance matrix. Consider a fine discrimination task of two orientations $\theta^+ = \theta_0 + d\theta$ and $\theta^- = \theta_0 + d\theta$. The specific decoder is a local linear estimator:

$$\hat{\theta} = \theta_0 + \mathbf{w}^T(\mathbf{r} - \frac{\mathbf{f}(\theta^+) + \mathbf{f}(\theta^-)}{2}).$$

The optimal weight to minimize the mean squared error over all trials, $E = \langle|\hat{\theta} - \theta|^2\rangle$, is

$$\mathbf{w}^s_{\text{opt}} = \frac{\Sigma^{-1}\mathbf{f}'}{\mathbf{f}'\Sigma^{-1}\mathbf{f}'}.$$

The linear Fisher information is equivalent to the inverse of the variance of the optimal specific decoder:

$$I = \frac{1}{\text{Var}(\hat{\theta}_{opt}|\theta^i)} = \mathbf{f}'\Sigma^{-1}\mathbf{f}'.$$

The linear Fisher information is estimated with bias-correction (**Fig. 3b**) (Kanitscheider et al., 2015a):

$$\hat{I} = \frac{(\mathbf{f}^+ - \mathbf{f}^-)^T}{d\theta}\left(\frac{\Sigma^+ + \Sigma^-}{2}\right)^{-1}\frac{(\mathbf{f}^+ - \mathbf{f}^-)}{d\theta}\left(\frac{2N_{\text{tr}} - N - 3}{2N_{\text{tr}} - 2}\right) - \frac{2N}{N_{\text{tr}}d\theta^2}, \tag{1}$$

where $\mathbf{f}^i$ and Σ^i are the empirical mean and covariance, respectively, for θ^i, $i \in \{+, -\}$. The number of neurons sampled is N, and the number of trials for each θ^i is N_{tr}. In simulations, we used $\theta_0 = 0.5$ and $d\theta = 0.01$. There were 58,500 spike counts in total for θ^+ and θ^-.

Network model general decoder. The general decoder is a complex linear estimator $\hat{z} = \mathbf{w^T r}$ (Shamir & Sompolinsky, 2006) where $\mathbf{w}$ is fixed for all θ. The estimator $\hat{z}$ maps the population activity $\mathbf{r}$ in response to all orientations to a circle ($z = e^{i\theta}$ in complex domain). The estimation of orientation is $\hat{\theta} = \arg(\hat{z})$. The optimal weight $\mathbf{w}^g_{\text{opt}}$ that minimizes the mean squared error, $E(\mathbf{w}) = \langle|\hat{z} - z|^2\rangle_{\theta,\mathbf{r}}$, averaged over all θ and trials of $\mathbf{r}$, is

$$\mathbf{w}^g_{\text{opt}} = \langle\Sigma(\theta) + \mathbf{f}\mathbf{f}^T\rangle_\theta^{-1}\langle\mathbf{f}e^{i\theta}\rangle_\theta, \tag{2}$$

The mean squared error of the optimal weight is

$$E(\mathbf{w}^g_{\text{opt}}) = 1 - (\langle\mathbf{f}e^{i\theta}\rangle_\theta)^*\langle\Sigma(\theta) + \mathbf{f}\mathbf{f}^T\rangle_\theta^{-1}(\langle\mathbf{f}e^{i\theta}\rangle_\theta),$$

where $*$ denotes the conjugate transpose. Hence, the estimation error of $\hat{z}$ depends on both the covariance matrix, Σ, and tuning similarity, $\mathbf{f}\mathbf{f}^T$. The performance of the general decoder is measured as $I_{\text{g}} = 1/\text{Var}(\hat{\theta})$ (**Fig. 3b**). The estimation of I_{g} is

$$\hat{I}_{\text{g}} = \frac{1}{\text{Var}(\arg((\mathbf{w}^g_{\text{opt}})^T\mathbf{r}) - \theta)}\frac{N_{\text{tr}} - N - 2}{N_{\text{tr}} - 1}, \tag{3}$$

where N_{tr} is the total number of trials for all θ's. In simulations, we used 50 θ's uniformly spaced between 0 and 1. There were 117,000 trials in total for all θ's.

Dependence of network model decoders' performance on correlated variability (Fig. 3d). We trained specific and general decoders on the same spike count dataset ($\mathbf{r}$) in response to pairs of orientations, θ_1 and θ_2 (with difference $\Delta\theta = 0.04$). The specific decoder was trained on the N-dimensional space of neural responses, using support vector machine model with two-fold cross-validation to linearly classify $\mathbf{r}$ for the two orientations. The general decoder first maps $\mathbf{r}$ to a two-dimensional plane $\hat{z} = (\mathbf{w}^{g}_{\text{opt}})^T\mathbf{r}$ using the optimal weight $\mathbf{w}^{g}_{\text{opt}}$ (**Eq. 2**) computed with the spike counts of all orientations. Then a two-dimensional support vector machine model with two-fold cross-validation was trained to linearly classify $\hat{z}$ for θ_1 and θ_2. The correlated variability was computed from the spike counts data for θ_1 of each pair. There were 200 sampling of $N = 100$ excitatory neurons from the V4 network, and 10 orientation pairs varying between 0 and 1. There were on average 2,340 trials for each θ.

Factor analysis for network model. Let $x \in \mathbb{R}^{n\times 1}$ be the spike counts from n simultaneously recorded neurons. Factor analysis assumes that x is a multi-variable Gaussian process:

$$x \sim \mathcal{N}(\mu, LL^T + \Psi)$$

where $\mu \in \mathbb{R}^{n\times 1}$ is the mean spike counts, $L \in \mathbb{R}^{n\times m}$ is the loading matrix of the m latent variables and $\Psi \in \mathbb{R}^{n\times 1}$ is a diagonal matrix of independent variances for each neuron (Cunningham & Yu, 2014). We chose $m = 5$ and compute the eigenvalues of LL^T, λ_i $(i = 1, 2, \ldots, m)$, ranked in descending order. Spike counts were collected using 200 ms window. There were on average 2,340 trials per attentional condition.

REFERENCES

Abbott, L. F. & Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural Comput* .**11,** 91-101 (1999).

Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7,** 358-366 (2006).

Berkes, P., Turner, R. E. & Sahani, M. A structured model of video reproduces primary visual cortical organisation. *PLoS Comput. Biol.* **5**, e1000495 (2009).

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

Burgess, A. E., Wagner, R. F., Jennings, R. J. & Barlow, H. B. Efficiency of human visual signal discrimination. *Science* **214**, 93-94 (1981).

Clery, S., Cumming, B. G. & Nienborg, H. Decision-Related Activity in Macaque V2 for Fine Disparity Discrimination Is Not Compatible with Optimal Linear Readout. *J. Neurosci.* **37,** 715-725 (2017).

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

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

Cohen, M. R. & Maunsell, J. H. R. Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron* **70,** 1192-1204 (2011).

Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17,** 1500-1509 (2014).

Deneve, S. Making decisions with unknown sensory reliability. *Front. Neurosci.* **6**, 75 (2012).

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., Smirnakis, S. M., Bethge, M. & Tolias, A. S. State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82,** 235-248 (2014).

Goris, R. L., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nat. Neurosci.* **17,** 858-865 (2014).

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).

Gu, Y., Liu, S., Fetsch, C. R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G. C. & Angelaki, D. E. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71,** 750-761 (2011).

Haefner, R. M., Gerwinn, S., Macke, J. H. & Bethge, M. Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nat. Neurosci.* **16**, 235-242 (2013).

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).

Howell, D. C. *Statistical Methods for Psychology* (Thomson Wadsworth, Belmont, CA, 2007).

Huang, C., Pouget, A. & Doiron, B. Internally generated population activity in cortical networks hinders information transmission. *bioRxiv* doi: 10.1101/2020.02.03.932723 (2020).

Huang, C., Ruff, D. A., Pyle, R., Rosenbaum, R., Cohen, M. R. & Doiron, B. Circuit Models of Low-Dimensional Shared Variability in Cortical Networks. *Neuron* **101,** 337-348 e334 (2019).

Jin, M., Beck, J. M. & Glickfeld, L. L. Neuronal Adaptation Reveals a Suboptimal Decoding of Orientation Tuned Populations in the Mouse Visual Cortex. *J. Neurosci.* **39,** 3867-3881 (2019).

Kanashiro, T., Ocker, G. K., Cohen, M. R. & Doiron, B. Attentional modulation of neuronal variability in circuit models of cortex. *Elife* **6,** e23978 (2017).

Kanitscheider, I., Coen-Cagli, R., Kohn, A. & Pouget, A. Measuring Fisher information accurately in correlated neural populations. *PLoS Comput. Biol.* **11,** e1004218 (2015a).

Kanitscheider, I., Coen-Cagli, R. & Pouget, A. Origin of information-limiting noise correlations. *Proc. Natl. Acad. Sci. U S A* **112,** E6973-E6982 (2015b).

Kaschube, M., Schnabel, M., Lowel, S., Coppola, D. M., White, L. E. & Wolf, F. Universality in the evolution of orientation columns in the visual cortex. *Science* **330,** 1113-1116 (2010).

Kersten, D. Statistical efficiency for the detection of visual noise. *Vision Res.* **27**, 1029-1040 (1987).

Kohn, A., Coen-Cagli, R., Kanitscheider, I. & Pouget, A. Correlations and Neuronal Population Information. *Annu. Rev. Neurosci*. **39,** 237-256 (2016).

Lin, I. C., Okun, M., Carandini, M. & Harris, K. D. The Nature of Shared Cortical Variability. *Neuron* **87,** 644-656 (2015).

Lorteije, J. A. M. *et al.* The Formation of Hierarchical Decisions in the Visual Cortex. *Neuron* **87**, 1344-1356 (2015).

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).

Mayo, J. P. & Maunsell, J. H. R. Graded Neuronal Modulations Related to Visual Spatial Attention. *J. Neurosci.* **36,** 5353-5361 (2016).

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

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

Nandy, A. S., Nassi, J. J. & Reynolds, J. H. Laminar Organization of Attentional Modulation in Macaque Visual Area V4. *Neuron* **93,** 235-246 (2017).

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).

Nirenberg, S. & Latham, P. E. Decoding neuronal spike trains: how important are correlations? *Proc. Natl. Acad. Sci. U S A* **100**, 7348-7353 (2003).

Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E. & Fellin, T. Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* **93**, 491-507 (2017).

Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10,** 437-442 (1997).

Posner, M. I. Orienting of attention. *Q. J. Exp. Psychol.* **32,** 3-25 (1980).

Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the shared gain of V4 populations. *Elife* **4,** e08998 (2015).

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

Ruff, D. A. & Cohen, M. R. Global cognitive factors modulate correlated response variability between V4 neurons. *J. Neurosci.* **34,** 16408-16416 (2014b).

Ruff, D. A. & Cohen, M. R. Stimulus Dependence of Correlated Variability across Cortical Areas. *J. Neurosci.* **36,** 7546-7556 (2016).

Ruff, D. A. & Cohen, M. R. Simultaneous multi-area recordings suggest that attention improves performance by reshaping stimulus representations. *Nat. Neurosci.* **22**, 1669-1676 (2019).

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

Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical Areas Interact through a Communication Subspace. *Neuron* **102,** 249-259 e244 (2019).

Shamir, M. & Sompolinsky, H. Implications of neuronal diversity on population coding. *Neural Comput.* **18,** 1951-1986 (2006).

Verhoef, B. E. & Maunsell, J. H. R. Attention-related changes in correlated neuronal activity arise from normalization mechanisms. *Nat. Neurosci.* **20,** 969-977 (2017).

Williamson, R. C., Cowley, B. R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M. A. & Yu, B. M. Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models. *PLoS Comput. Biol.* **12,** e1005141 (2016).

Yan, Y., Rasch, M. J., Chen, M., Xiang, X., Huang, M., Wu, S. & Li, W. Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nat. Neurosci.* **17,** 1380-1387 (2014).

Zenon, A. & Krauzlis, R. J. Attention deficits without cortical neuronal deficits. *Nature* **489,** 434-437 (2012).

Author contributions. A.M.N., C.H., B.D., and M.R.C. designed the project; A.M.N. collected and analyzed the electrophysiological data; C.H. performed the model simulations and analyzed the model data; M.R.C. supervised the project; A.M.N., C.H., B.D., and M.R.C. contributed to writing the manuscript.

Competing interests. The authors declare no competing financial interests.