NEUROSCIENCE

Neuronal correlates of selective attention and effort in visual area V4 are invariant of motivational context

Supriya Ghosh* and John H. R. Maunsell

Task demands can differentially engage two fundamental attention components: selectivity (spatial bias) and effort (total nonselective attentional intensity). The relative contributions and interactions of these components in modulating neuronal signals remain unknown. We recorded V4 neurons while monkeys' spatially selective attention and effort were independently controlled by adjusting either task difficulty or reward size at two locations. Neurons were robustly modulated by either selective attention or effort. Notably, increasing overall effort to improve performance at a distant site reduced neuronal responses even when performance was unchanged for receptive field stimuli. This interaction between attentional selectivity and effort was evident in single-trial spiking and can be explained by divisive normalization of spatially distributed behavioral performance at the single-neuron level. Changing motivation using task difficulty or reward produced indistinguishable effects. These results provide a cellular-level mechanism of how attention components integrate to modulate sensory processing in different motivational contexts.

INTRODUCTION

Attention plays an essential role in motivating human behavior and cognition by selectively enhancing the processing of relevant sensory information. Goal-directed attention supporting for cognitively demanding tasks is often driven by external incentives. Many cortical and subcortical brain areas change their activity when attention shifts (*1*–*4*). They are also sensitive to the size of the reward that motivates those shifts (*5*–*11*). Although reward expectation and attention have been described as conceptually distinct cognitive constructs, it remains challenging to distinguish these factors owing to their covariance and the high similarity of their effects on neuronal responses [for review, see (*12*)]. Attentional levels can also be elevated owing to internal desire to complete a task without any apparent changes in external incentives, such as increased cognitive demand as a result of increased task difficulty (*13*). For example, professional athletes or musicians address challenging situations with increased effort so as to maintain a given performance level. The contributions of different sources of motivation to regulation of sensory processing in cortex and overall perceptual behavior remain elusive.

To adapt to varying environmental and stimulus contexts, subjects can shift their attention between spatially localized targets or selective stimulus features to spatially global targets or nonselective features. Many studies have characterized the neuronal modulations associated with selective attention by assaying how performance improves for attended spatial locations or stimulus features relative to distant locations or unrelated features. When a monkey's attention is selectively directed toward the location of a neuron's receptive field (RF), improvement in perceptual performance in that region is typically accompanied by increased spike rates (*4*, *14*) and reduced individual response variance and pairwise spike-count correlations (*1*, *15*). Although experimental studies most often treat attention as all or none, it has another fundamental aspect, effort (*16*)—how intensely attention is focused independent of selectivity. Attentional effort can be considered as an objective measure of total cognitive engagement in a goal-directed attention-demanding task (*17*). Attention-related modulations of neurons in area V4 in primate visual cortex have been examined using a variety of visual detection tasks (*2*, *3*, *18*, *19*). Some of these studies show that V4 neuronal activity is enhanced as a result of higher cognitive effort in response to increased task demand (*13*, *20*). It remains relatively unknown how selective attention and attentional effort integrate in the brain to improve sensory perception and performance, and how motivational contexts influence these processes.

To address these questions, we trained monkeys to do an attention-demanding visual task that allowed us to independently control the monkey's attentional selectivity and effort in either of two different motivational contexts. We either varied task difficulty while reward size was kept fixed or varied reward size for a fixed task difficulty. Using simultaneous electrophysiological recordings from populations of V4 neurons and computational models, we found that attentional selectivity and effort independently modulate neuronal spiking. Single-trial spike trains encode multiplexed signals of attentional selectivity and effort with comparable strengths in a way that is independent of how the animal was motivated to allocate its attention. Further, the effects of attentional selectivity and effort interact to determine the resultant influence of attention on neuronal spiking. A spatially tuned normalization model of attention can account for this interaction. Thus, we provided a detailed account of how fundamental components of attention interact at the level of V4 spikes. By extending the spectrum of attention-related cognitive representations in V4, the results clarify how individual neurons contribute to higher-order cognition.

RESULTS

Independent control of attentional selectivity and effort by varying task difficulty

We trained two rhesus monkeys to distribute their visual spatial attention between two stimuli in the left and right hemifields while doing an orientation change detection task (Fig. 1A). The animal held its gaze on a central fixation spot throughout each trial. After a randomly varying period of fixation, two Gabor sample stimuli

Department of Neurobiology and Neuroscience Institute, The University of Chicago, Chicago, IL 60637, USA.
*Corresponding author. Email: sghosh5@uchicago.edu

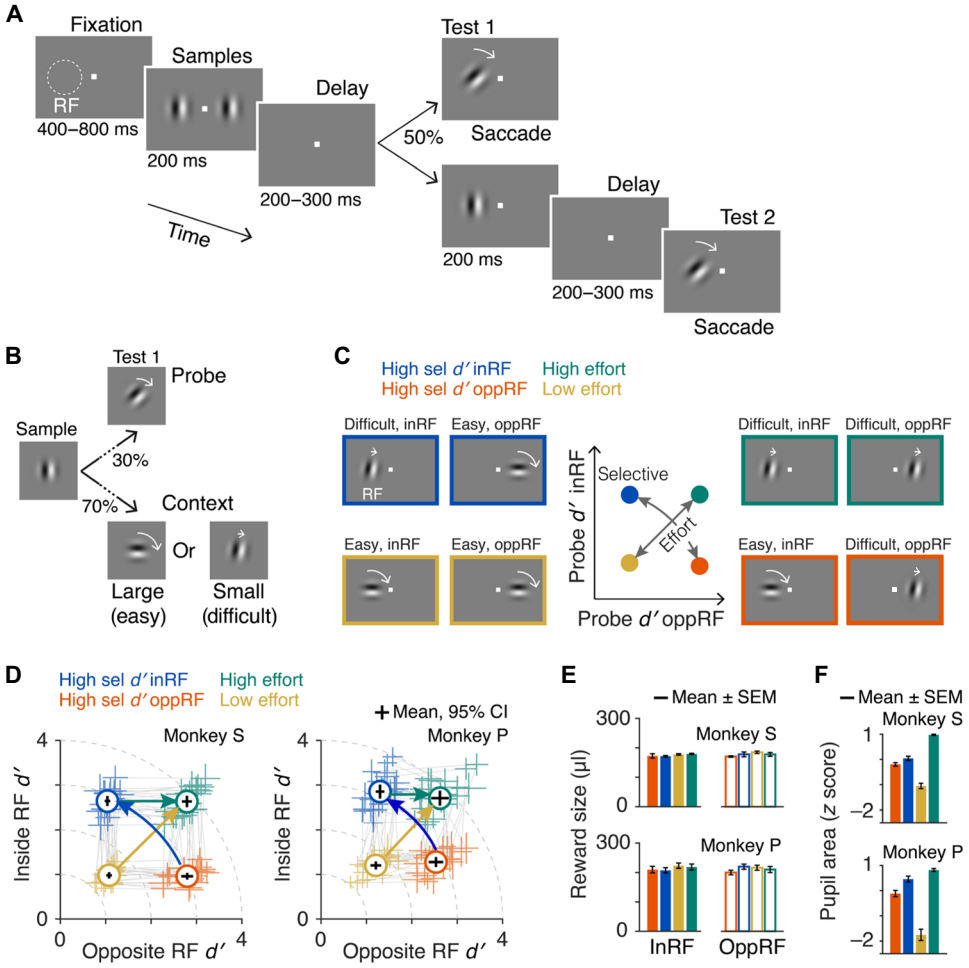

Fig. 1. Independent control of attentional selectivity and effort using differential task demands. (**A**) Visual spatial attention task. Monkeys were required to fixate, attend to sample stimuli (Gabors) presented in both hemifields [inside and opposite side of recorded neurons' receptive field (RF)], and report an orientation change that occurred in one of the two test intervals by making a saccade to the stimulus location. (**B**) Task demands. Orientation changes on ~30% of the nonmatched trials are intermediate (probe) and on the rest ~70% of the nonmatched trials are either large (easy context) or small (difficult context). (**C**) Middle: Distribution of task difficulties across two locations in opposite hemifields for four task conditions: high selective attention either inside the RF (blue) or opposite RF location (orange) when attentional effort remains fixed, and low or high nonselective attentional effort (yellow and green). Left and right: Four attention conditions consisted of different combinations of these task contexts (easy and difficult) at the two stimuli locations. (**D**) Attention operating characteristic (AOC) curve, indicating behavioral sensitivity (*d′*) on individual sessions and their average (circles) for test stimuli inside and opposite side RF during the four attention conditions (monkey S, 20 sessions; monkey P, 22 sessions). Gray lines connect attention conditions within a session. Dashed lines, iso-effort lines. Error bars, 95% confidence intervals (CI). (**E**) Session averaged reward sizes across different attention conditions for individual monkeys. (**F**) Session averaged pupil area (*z* scored) during the sample stimulus period. Error bars, ±SEM.

appeared for 200 ms. This was followed by a delay of 200 to 300 ms, after which a single Gabor test stimulus appeared at one of the two sample locations (selected pseudo-randomly). If the orientation of the test stimulus differed from the orientation of the sample stimulus that had appeared in that location (a target), then the monkey had to rapidly make a saccade to the stimulus to earn a juice reward. On a random 50% of the trials, the orientation of test stimulus was unchanged (a nontarget), and the monkey was required to maintain fixation. In that case, a second test stimulus that always had a different orientation was presented after a short delay, and the monkey needed to saccade to this target stimulus to earn a reward.

To control the animal's attention, in each block of trials, we set the orientation change of the first test stimulus at each location to be either easy to detect (~80°) or difficult to detect (~18°) (Fig. 1B). In each block, the size of the orientation change at the two locations was set independently, providing four possible combinations (Fig. 1C). We measured the behavioral consequences of different combinations of difficulty by presenting an orientation change of intermediate difficulty (30°, probe) on a randomly selected fraction of all trials (~30%). These probe trials allowed us to directly compare behavioral sensitivity (*d′*) at both locations across all four block types. Figure 1D plots the average *d′*s for the left and right stimulus locations on probe trials for the two monkeys separately, with different colors representing the four different combinations of difficulty. Cross marks represent average *d′*s from individual sessions, and gray lines join the four means from within individual sessions. The changes in behavioral performance document that the animals responded to task difficulty by adapting their allocation of attention. Behavioral *d′* for the probe orientation change on each side was substantially higher when most orientation changes were difficult to detect, and lower when most

changes were easy to detect, with approximately symmetrical d's at both locations during most individual sessions.

Spatial selectivity of attention was quantified by selectivity index that measured the relative behavioral d' at the RF location compared with the opposite location (Materials and Methods). Attentional effort was measured by overall absolute behavioral d's in the two locations (Materials and Methods). The ~4-fold difference in orientation change [median easy change, 80°; interquartile range (IQR), 80° to 90°; median difficult change, 18°; IQR, 16° to 18°] strongly motivated animals to adjust their behavioral d', whether the interblock changes on the two sides were in opposite directions (blue arrows, Fig. 1D) or in the same direction (gold arrows, Fig. 1D). In both cases, behavioral d' changed by ~2-fold (selectivity indices in the opposite direction: monkey S, for low RF d' mean −0.58, SEM 0.01; for high RF d' mean 0.52, SEM 0.01, $P < 10^{-12}$; monkey P, for low RF d' mean −0.41, SEM 0.02; for high RF d' mean 0.46, SEM 0.01, $P < 10^{-13}$; attentional effort in the same direction: monkey S, for nonselective low d' mean 1.46, SEM 0.04; for nonselective high d' mean 3.84, SEM 0.08, $P < 10^{-9}$; monkey P, for nonselective low d' mean 1.71, SEM 0.07; for nonselective high d' mean 3.78, SEM 0.11, $P < 10^{-13}$; table S1). These changes in allocation of attention were driven by changes in task difficulty alone. Although reward sizes for correct responses were varied somewhat from trial to trial (see below), the average was kept equal on both sides across all task difficulty configurations (Fig. 1E).

Nonluminance-mediated task-evoked increases in pupil size are commonly considered an index of arousal or attentional engagement and are sensitive to task demands across species (*21*–*23*). Consistent with this, pupil area during the sample stimuli increased progressively with the increase in attentional effort [$F_{3,76} = 62.71$, $P < 10^{-19}$ for monkey S; $F_{3,84} = 88.19$, $P < 10^{-25}$ for monkey P, analysis of variance (ANOVA); Fig. 1F]. Pupil area was greatest when discriminations were difficult on both sides, and smallest when they were easy on both sides. Both animals had a small spatial bias. In each, we recorded from the hemisphere representing the hemifield where the animal's performance was better. A small spatial bias in attention between the left and right stimuli in the high-selectivity condition can be seen in the average pupil areas (blue versus orange, Fig. 1F) and in a slightly higher d' for selective attention inside the RF for monkey P (blue versus orange, Fig. 1D). This small bias would have acted to reduce the neuronal modulations we describe below related to attention opposite to the RF relative to modulations when performance was unbiased.

Relative neuronal modulation of V4 with attentional selectivity and effort

We recorded 1194 single units and small multiunit clusters (single unit, 385; multiunit, 809) during 42 recording sessions from the two monkeys (monkey S, 20 sessions and 714 units; monkey P, 22 sessions and 480 units) using 96-channel multielectrode arrays chronically implanted in V4 in the superficial prelunate gyrus. Neurons typically responded more strongly to the sample stimuli during the trial blocks when the monkey's behavioral d' at the RF location was high. This increased spiking response was seen whether the behavioral d' differences involved different selectivity for the RF versus other location (orange versus blue) or a change in attentional effort with no change in selectivity (gold versus green). The spike responses did not depend exclusively on d' in the RF location. Neuronal responses during high nonselective behavioral d' (green, Fig. 2, A and B) were reduced

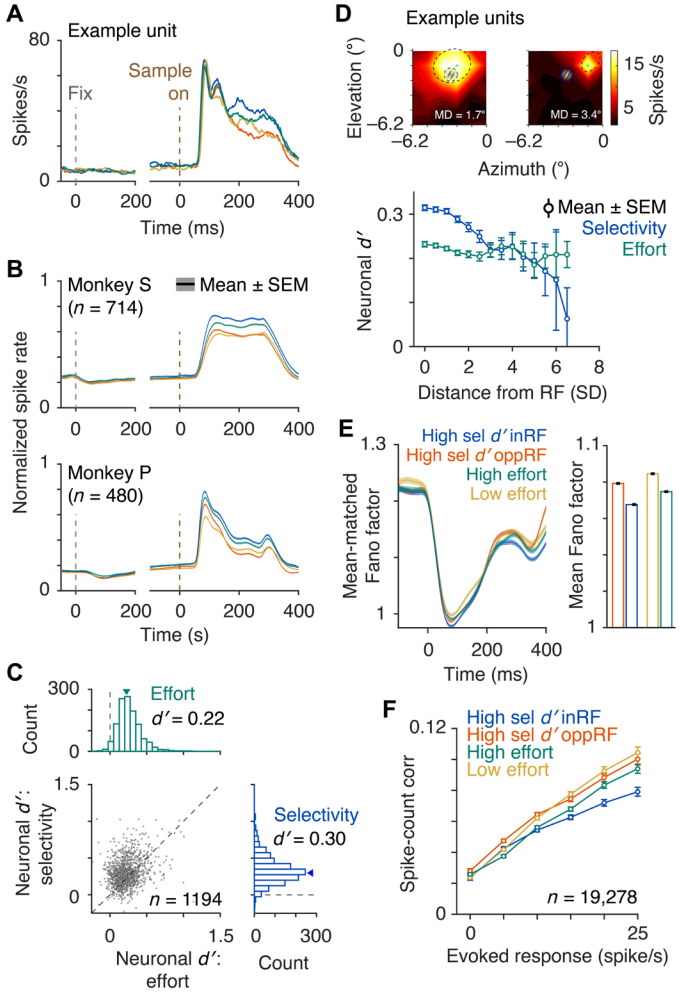

Fig. 2. Neuronal modulation with changes in attentional selectivity and effort. (**A**) Peri-stimulus time histograms (PSTHs) of spike rates of correct trials in different attention conditions for an example neuron in V4. Single-trial spike counts were binned at 2 ms, smoothed with σ = 15 ms half-Gaussian, and then aligned at the onset of sample stimulus. (**B**) Population PSTHs for monkey S (top) and monkey P (bottom). For population average, spike rates of each neuron were normalized to its peak response within 60 to 260 ms from sample stimulus onset (monkey S, $n = 714$; monkey P, $n = 480$). (**C**) Distribution of neuronal d' for attentional selectivity and effort of all units from both monkey S and P ($n = 1194$). (**D**) Top: RF locations of two example units relative to the Gabor stimulus. Mahalanobis distance (MD) measures the standardized distance between neurons' RF and the Gabor stimulus in the RF location in units of SD. Bottom: Distribution of neuronal d' as a function of RF-Gabor distance. (**E**) Left: Mean-matched Fano factor. Right: Mean-matched Fano factor averaged over 60 to 260 ms from the sample onset. Error bars, ±SEM. (**F**) Pairwise correlations between spike counts of simultaneously recorded neurons over 60 to 260 ms from sample onset ($n = 19,278$ pairs, all units) and binned according to their evoked responses (geometric mean of baseline subtracted spike counts). Error bars, ±SEM.

compared with responses with identical RF d' and low d' at the distant location (blue, Fig. 2, A and B).

To quantify neuronal modulation by attentional selectivity and effort, we computed a neuronal d' as the difference of z-scored firing rates (60 to 260 ms from sample onset) between high and low attention states. The mean firing rate modulation was significantly greater for attentional selectivity compared with nonselective effort

(neuronal d' for selective attention, mean ± SEM = 0.30 ± 0.01; for nonselective attentional effort, mean ± SEM = 0.22 ± 0.05, $P < 10^{-31}$, $n = 1194$, t test; Fig. 2C). Single neurons and multiunit clusters separately showed similar spike modulation by attentional selectivity and nonselective effort (mean ± SEM for single units: selective attention, 0.29 ± 0.01; effort, 0.23 ± 0.01; $P < 10^{-7}$, t test; mean ± SEM for multiunits: selective attention, 0.30 ± 0.01; effort, 0.22 ± 0.01; $P < 10^{-26}$, t test). These attentional effects were significant for the monkeys individually (monkey S, $P < 10^{-27}$; monkey P, $P < 10^{-8}$). High levels of effort might involve changes in global arousal that could elevate ongoing neuronal activity compared with high selective attention. In that case, neuronal modulation of ongoing neuronal activity of visually nonresponsive neurons would be higher in high effort condition compared with high selective attention. To examine this, we measured neuronal modulations of presample (−200 to 0 ms) spike counts in visually nonresponsive units for selective attention and effort. There were weak presample modulations in visually nonresponsive neurons for both selective attention and effort compared with the visually responsive neurons (neuronal d' for selective attention: mean ± SEM = 0.02 ± 0.01, $P < 0.05$; for effort: mean ± SEM = 0.02 ± 0.01, $P < 0.05$, $n = 198$; fig. S2). This suggests that the modulations of ongoing neuronal activity might be associated with absolute level of effort (or arousal) irrespective of attentional selectivity. Weak modulation of prestimulus spiking has previously been reported with selective visual attention (*24*–*26*).

Next, we looked at whether the lower spike rates with attentional shift from high selective to high effort altered sensory information encoding, which is essential for intact perception and behavioral performance. First, we measured neuronal d' for discrimination of orientations of sample stimulus in the RF location during high selective attention and high effort conditions (fig. S5). We found no difference in sensory discriminability of V4 neurons. We also tested whether the sensory information encoding at the network level (population coding) differed between the two high attention conditions. Spike peri-stimulus time histograms (PSTHs) of simultaneously recorded units in every session were decomposed into demixed principal components to measure decoding accuracy of sample orientations (fig. S6) (*27*). There was no difference in decoding accuracy between the high selective attention and high effort conditions to the extent stimulus orientation was represented by V4 spike trains. Thus, an apparent disconnect between the mean spike rates and the behavioral d' does not necessarily mean that the sensory representations are impaired in the high effort conditions.

Changes in attentional effort with this task design do not rule out all forms of spatial selection because the animals might have attended to locations other than the two stimulus locations tested. If so, V4 neurons could have been modulated by the spatially selective shifting of attention from those other locations to the two stimulus locations. We examined the broader spatial distribution of attention by measuring the correlation between firing rate modulation and the proximity of a V4 neuron's RF and the attended Gabor stimulus (Fig. 2D) using Mahalanobis distance to measure proximity. As expected, neuronal d' dropped substantially with increasing RF distance from the stimulus center when animals shifted their spatially selective attention (Spearman correlation coefficient, $\rho = -0.18$, $P < 10^{-8}$; Fig. 2D). In contrast, d' for the same neurons varied little with RF distance when animals were encouraged to adjust their attentional effort (Spearman, $\rho = -0.05$, $P = 0.11$; Fig. 2D), supporting the absence of spatially selective attention in this manipulation. Correlation between the RF distance and neuronal d' for attentional selectivity was significantly higher compared with attentional effort ($P = 0.002$, z test).

Because we recorded from the same fixed multielectrode arrays over many sessions, it is possible that some units were sampled in more than one session. We investigated the effect of potential resampling by analyzing a subsample that included only one unit from each electrode across all recording sessions ($n = 85$ for monkey S and $n = 80$ for monkey P). For this conservative set of unequivocally unique units, both the high selective and nonselective effort increased spike rates, and the modulation was stronger for selective attention than nonselective effort (mean ± SEM neuronal d' for monkey S: selective attention, 0.25 ± 0.01; nonselective effort, 0.19 ± 0.01; $P < 10^{-3}$; for monkey P: selective attention, 0.31 ± 0.02; nonselective effort, 0.25 ± 0.02; $P = 0.02$; t test; fig. S3) by amounts that were indistinguishable from the whole population. Thus, the results cannot be attributed to multiple sampling that might have occurred from units with uncharacteristic properties.

In addition to spike rate modulation, we tested the relative effects of attentional selectivity and nonselective effort on signal to noise of individual V4 units by measuring mean-matched Fano factor (the ratio of the variance of the spike counts to the mean). Fano factors during the sample stimulus period were significantly reduced by increased attentional selectivity ($F_{1,232} = 73.27$, $P < 10^{-14}$; ANOVA; Fig. 2E) as well as effort ($F_{1,232} = 310.92$, $P < 10^{-43}$, ANOVA). Further, the effect of attentional selectivity on Fano factor was reduced (smaller decrease in Fano factor) for an identical improvement in behavioral d' in the RF location during high effort compared with low effort (selectivity-by-effort interaction, $F_{1,232} = 4.91$, $P = 0.027$, ANOVA). Similar to the signal to noise, pairwise spike count correlations of simultaneously recorded units were also reduced with higher attentional selectivity and effort (selectivity, $F_{1,19277} = 7.75$, $P < 0.01$; effort, $F_{1,19277} = 40.91$, $P < 10^{-9}$; selectivity-by-effort interaction, $F_{1,19277} = 21.55$, $P < 10^{-5}$, ANOVA; Fig. 2F). These results suggest that although the neuronal modulation by selective attention is stronger than the modulation by effort, they share many similarities, and both contribute appreciably to attention-related modulations.

Independent control of attentional selectivity and effort using differential reward size as the external motivator

Subjects are motivated in many different ways to allocate their attention. So far in our task, changes in task difficulty motivated monkeys to spatially redistribute their attention to match task demands. We next tested whether the encoding of attention components in V4 neurons depends on how animals are motivated to attend. For this, we instructed the same monkeys to shift their spatial attention by varying reward sizes between the two locations (Fig. 3, A and B). The size of the orientation change was kept constant and challenging on both sides throughout the session. Consequently, no probe orientation changes were needed or presented. The reward manipulation sessions were conducted on different days that were interleaved with the task difficulty sessions described in previous sections. The trial distributions and the probability of an orientation change in the two locations in opposite hemifields were the same. Thus, the allocation of spatial attention across the hemifields was primarily motivated by the reward distributions. Figure 3C plots behavioral d' on the first test stimuli on the left and right sides for all four reward schedules. Behavioral d's on both locations were symmetrical during most individual sessions.

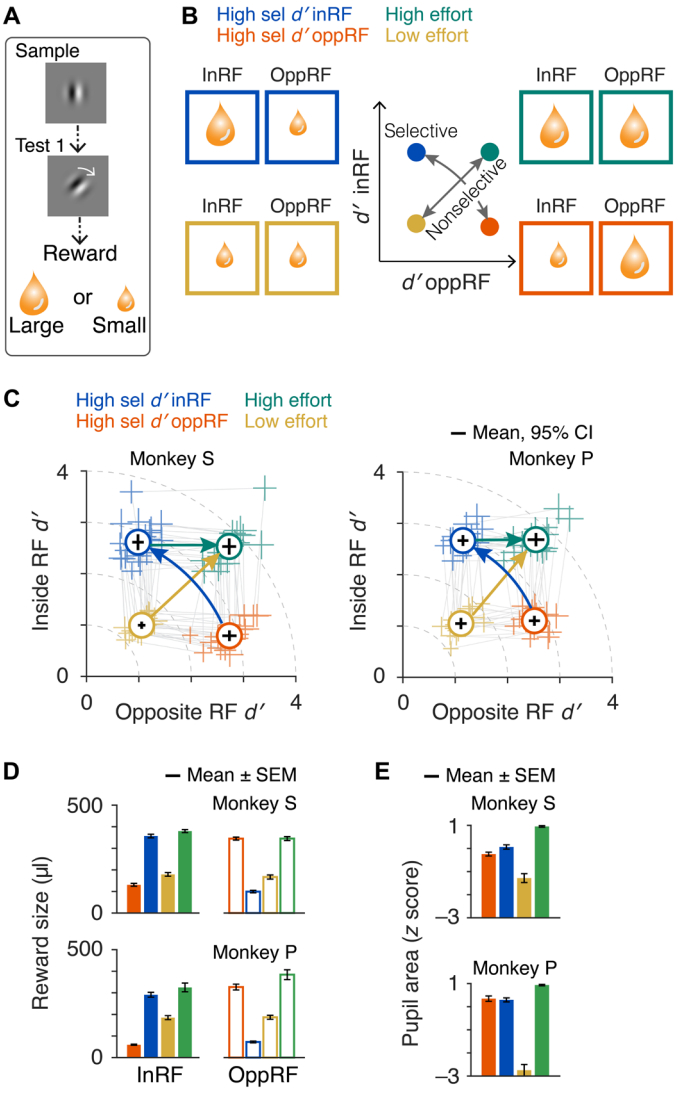

Fig. 3. Independent control of attentional selectivity and effort using differential reward size. (**A**) Single orientation change of the nonmatched trials remains fixed. Reward size for a correct choice is either small or large. (**B**) Four attention conditions consisted of different combinations of reward sizes (small and large rewards) at the two stimuli locations. (**C**) AOC curve, indicating behavioral sensitivity (*d*′, circles) on individual sessions and their average (solid markers) for test stimuli inside and opposite side RF (monkey S, 20 sessions; monkey P, 16 sessions). Dashed colored lines indicate average *d*′ in each hemifield. Lines connect two reward conditions within a session. Error bars, 95% confidence intervals. (**D**) Session averaged reward sizes across the four conditions for individual monkeys. (**E**) Session averaged pupil area (*z* scored) during sample stimulus periods. Error bars, ±SEM.

The ~2.5-fold increase in reward size (median small, 136 µl; IQR, 87 to 177 µl; median large, 340 µl; IQR, 305 to 373 µl; Fig. 3D) strongly motivated animals to adjust their behavioral *d*′, whether the direction of change on the two sides was opposite (blue arrows, Fig. 3C) or the same (gold arrows, Fig. 3C). In both cases, behavioral *d*′ changed by ~2-fold (opposite direction: selectivity indices in RF, monkey S, for low *d*′ mean −0.64, SEM 0.02; for high *d*′ mean 0.54, SEM 0.02, $P < 10^{-11}$; monkey P, low *d*′ mean − 0.47, SEM 0.02; high *d*′ mean 0.48, SEM 0.02, $P < 10^{-11}$; same direction: effort indices, monkey S, for low nonselective effort mean 1.45, SEM = 0.05; for high effort mean 3.72, SEM 0.1, $P < 10^{-9}$; monkey P, for low effort mean 1.53, SEM 0.07; for high effort mean 3.70, SEM 0.09, $P < 10^{-10}$; table S2). These changes in attention allocation were produced by changes in reward size alone.

As with changes in task difficulty, high nonselective attentional effort mediated by reward size was also associated with increased pupil area during the sample stimuli ($F_{3,76} = 54.12$, $P < 10^{-18}$ for monkey S; $F_{3,60} = 160.47$, $P < 10^{-27}$ for monkey P; ANOVA; Fig. 3E).

Neuronal modulation of attentional selectivity and effort by differential reward size was indistinguishable from task difficulty–mediated modulation

We recorded a total of 1331 single units and small multiunit clusters (single unit, 419; multiunit, 912) in V4 during 36 reward manipulation recording sessions from the two monkeys (monkey S, 20 sessions and 850 units; monkey P, 16 sessions and 481 units). Spike response modulation was similar to the effects observed when attention was controlled using task difficulty (Fig. 4, A to C). Population PSTHs for correctly completed trials increased with high selective attention inside the neuron's RF (orange and blue traces, Fig. 4, A and B). Spiking activity also increased for higher attentional effort but was relatively smaller compared with the modulation because of increased selective attention (yellow and green traces, Fig. 4, A and B). Neuronal *d*′ for attentional selectivity was higher than effort (mean ± SEM for selectivity, 0.37 ± 0.006; effort, 0.24 ± 0.004; $P < 10^{-72}$, *t* test), and they were significantly correlated ($\rho = 0.25$, $P < 10^{-19}$, Spearman correlation coefficient; Fig. 4C). Compared with the effort, neuronal *d*′ for attentional selectivity dropped more strongly with the RF-sample stimulus proximity [selective attention, $\rho = -0.21$ ($P < 10^{-13}$); effort, $\rho = -0.09$ ($P = 0.0003$); $P = 0.005$, *z* test; Fig. 4D]. In addition to spiking, neurons' mean-matched Fano factor and pairwise spike count correlations reduced with higher attentional selectivity and effort (Fano factor: selectivity, $F_{1,224} = 21.64$, $P < 10^{-5}$; effort, $F_{1,224} = 66.19$, $P < 10^{-13}$; pairwise correlations: selectivity, $F_{1,26233} = 226.5$, $P < 10^{-50}$; effort, $F_{1,26233} = 33.36$, $P < 10^{-8}$, ANOVA; Fig. 4, E and F). Similar to the task difficulty, a significant interaction was also detected between the selectivity and effort on the Fano factor in the reward context ($F_{1,224} = 7.57$, $P < 0.01$, ANOVA) as well as pairwise spike count correlations ($F_{1,26233} = 6.4$, $P < 0.05$, ANOVA).

Encoding of attentional selectivity and effort within single-trial spike train

Spike trains of V4 neurons provide dynamic signals about many task-relevant variables (*28*). We next compared the relative contributions of attentional selectivity and effort on the within-trial instantaneous spiking of individual V4 neurons in varying difficulty and varying reward contexts using a generalized linear encoding model (Fig. 5A) (*28*). The probability of an observed spike count within a small time window (50 ms) was modeled as an exponential function of a weighted linear combination of task variables: attentional selectivity (ratio of *d*′s at the RF location over the oppRF location), attentional effort (radial distance from the inRF *d*′–oppRF *d*′ from the origin), selectivity-by-effort interaction, orientation of the sample stimulus inside the neuron's RF, and the direction of the eventual response saccade. The probability of a spike was constructed to follow a negative binomial distribution (Materials and Methods). Most neurons were well fit with this model [difficulty context, 1169/1194 (98%), $P < 0.05$; reward context, 1311/1331 (98%), $P < 0.05$; *F* test]. Figure 5B shows PSTHs of observed and model fitted spike

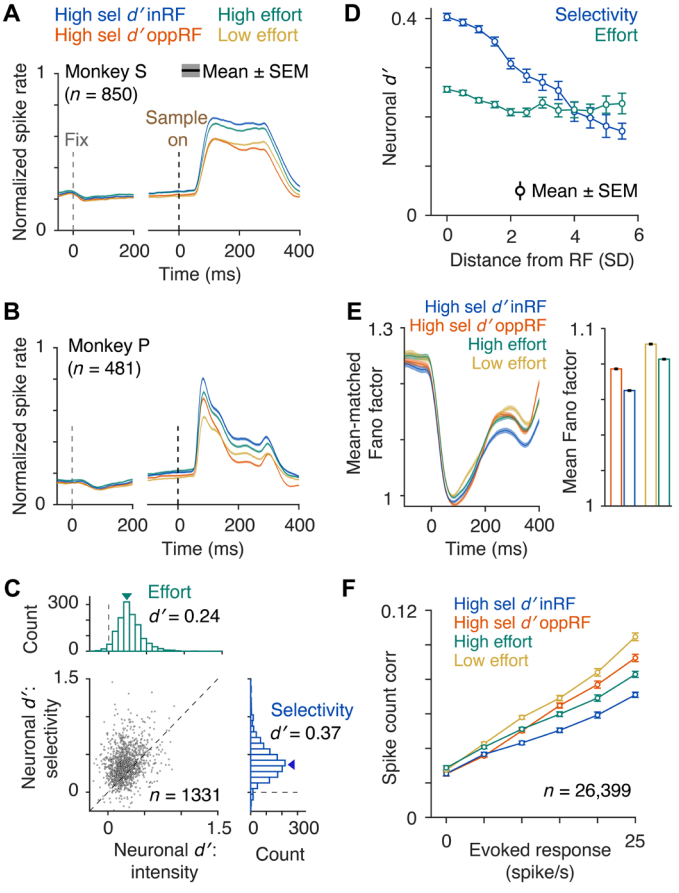

Fig. 4. Neuronal modulation with changes in attention components controlled by differential reward size. (**A** and **B**) Population PSTHs of spike rates of correct trials in different attention conditions for monkey S (A) and monkey P (B). Single-trial spike counts were binned at 2 ms and smoothed with σ = 15 ms half-Gaussian. Spike rates of each neuron were normalized to its peak response within 60 to 260 ms from sample stimulus onset (monkey S, $n = 850$; monkey P, $n = 481$). (**C**) Distribution of neuronal d' for attentional selectivity and effort of all units from both monkeys ($n = 1331$). (**D**) Distribution of neuronal d' as a function of Mahalanobis distance between neurons' RF and Gabor stimulus. (**E**) Left: Mean-matched Fano factor. Right: Mean-matched Fano factor averaged over 60 to 260 ms from the sample onset. Error bars, ±SEM. (**F**) Pairwise correlations between spike counts of simultaneously recorded neurons ($n = 26,399$ pairs, all units) and binned according to their evoked responses (geometric mean). Error bars, ±SEM.

counts in different attention conditions for trials in cross-validation test datasets for an example neuron in a varying task difficulty session. Figure 5C illustrates fitted model components of attention for the same example neuron as in Fig. 5B. The effective influence of distinct attention components on spike counts is expressed as multiplicative gains (exponentiated fitted coefficients) at a representative time during the sample stimulus (140 ms). The product of these gain components results in the predicted rate for a single trial (combined gain, bottom row). For an identical increase in behavioral d' in the RF location, the increase in spike counts will be higher when the opposite RF d' is small (blue arrow, Fig. 5C) compared to large (green arrow, Fig. 5C).

We next ask whether attentional selectivity, effort, and their interaction were encoded by distinct set of V4 neurons and whether task context affected these population representations. Many units showed significant effects for several or all variables: attentional selectivity, effort, and selectivity-by-effort interaction [difficulty context: selectivity and effort, 586/919 (64%), $P < 10^{-10}$; selectivity and selectivity by effort, 604/904 (67%), $P < 10^{-10}$; effort and selectivity by effort, 621/772 (80%), $P < 10^{-10}$; selectivity, effort, and selectivity by effort, 542/934 (58%), $P < 10^{-10}$; reward context: selectivity and effort, 709/1038 (68%), $P < 10^{-10}$; selectivity and selectivity by effort, 713/1033 (69%), $P < 10^{-10}$; effort and selectivity by effort, 710/871 (81%), $P < 10^{-10}$; selectivity, effort, and selectivity by effort, 642/1047 (61%), $P < 10^{-10}$; chi-square test; Fig. 5, D and F]. Further, these distributions were largely unchanged across the task difficulty and reward contexts ($P = 0.97$, chi-square test). The distributions of positive and negative coefficients also remained indifferent between the two contexts (proportion of positive coeffects in difficulty context: selectivity, 86.7%; effort, 52.2%; and selectivity-by-effort interaction, 41.8%; reward context: selectivity, 88.3%; effort, 56.5%; and selectivity-by-effort interaction, 43.4%; $P = 0.47$, chi-square test). This mixed representation indicates multiplexed encoding of attention components by the same V4 unit.

We next compared the relative contributions of the cognitive and task variables on spike responses for individual units as measured by predictor importance (PI; normalized magnitude of fitted coefficients; Fig. 5, E and G; Materials and Methods). Following the onset of sample stimulus, stimulus orientation had a dominant contribution to the spike counts in both task contexts. This is expected because V4 neurons have robust visual responses, and most are orientation selective. Attentional selectivity, effort, and selectivity-by-effort interaction remained strong predictors of spike trains from the start of the trial and increased immediately after the stimulus onset. Saccade direction contributed negligibly to V4 activity. Contributions of attention components were indifferent to the task contexts. Both animals, single units and multiunits, showed similar results (figs. S8 to S11). Units in reward sessions were also fitted with an alternate model that had an additional reward history parameter to test contribution of reward information on spike trains (fig. S12). Reward information had negligible contribution on spike trains of V4 units, as reported in a previous study (*28*). Collectively, these results suggest that individual V4 neurons independently carry multiplexed information about selective attention and attentional effort in single-trial spike trains relative to other sensory and task variables. These two attention components are integrated independently of the way the animal is motivated to allocate their attention. We also found no differences in spike rate latency between the two task contexts ($F_{1,2379} = 1.98$, $P = 0.16$, ANOVA; fig. S14; Materials and Methods). Further, it is unlikely that the neurons showing attention-related modulation in the two task contexts come from two distinct populations (monkey S: fraction of units with highest probability, 8.8%; probability, 0.44; monkey P: fraction of units with highest probability, 9.5%; probability, 0.46; Materials and Methods).

A normalization model of attention can account for the interactions between attentional selectivity and effort

The decrease in V4 responses with the increase in behavioral d' in the opposite hemifield (blue to green, and yellow to orange, Figs. 2D and 3C) might seem unexpected, both because behavioral performance at the RF location does not change and because the overall behavioral performance is better when the animal allocates high attention to both locations. The rate of firing was reduced more when stimulus orientation matched with the neuron's preferred orientation compared

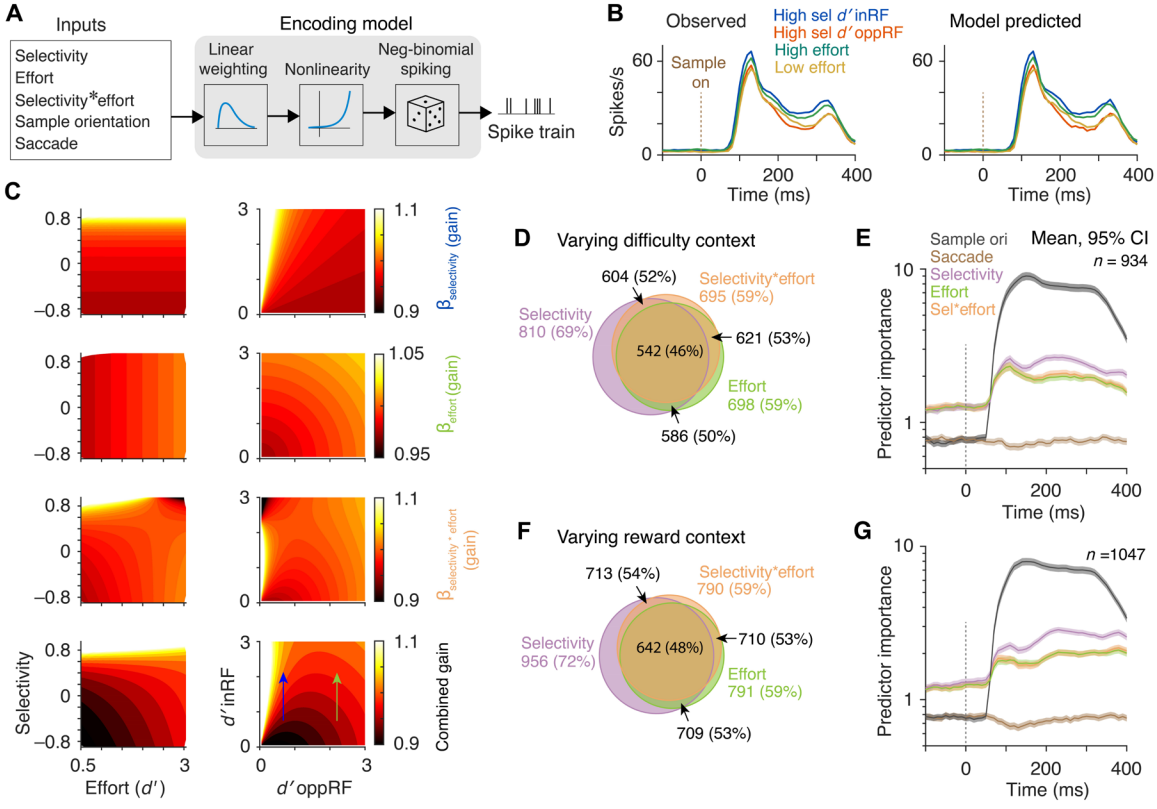

Fig. 5. Single-trial encoding of attention components in different motivation contexts. (**A**) Generalized linear encoding model. A neuron's spike count over 50 ms sliding window (10 ms shift) is modeled as an exponential function of a linear combination of weighted (β coefficient) experimental variables: stimulus orientation, saccade, attentional selectivity, attentional effort, and their selectivity-by-effort interaction. (**B**) Model predicted (left) and observed (right) PSTHs from cross-validation test dataset for an example neuron in a task difficulty session. (**C**) Distributions of GLM fitted coefficients (exponentiated gain) for attention components at a representative time 140 ms as a function of attentional selectivity and effort (left column) and behavioral *d*′s oppRF and inRF (right column) for the same example neuron in (B). These components multiplicatively influence spike counts. (**D** and **E**) Varying task difficulty context. Proportion of neurons that are modulated ($P < 0.05$) by attentional selectivity, effort, and selectivity-by-effort interaction estimated from the model (D). Comparing predictor importance (PI) that measures contributions of different predictor variables estimated by absolute standardized predictor coefficient values of all well fitted neurons (E). Error bars, 95% confidence intervals (bootstrap, $n = 10^4$). (**F** and **G**) Varying reward context. Same as (D) and (E) when attention was controlled by differential reward sizes. Error bars, 95% confidence intervals.

with the nonpreferred orientation (fig. S15). The orientation similarity between the sample stimuli in the two locations (inRF and oppRF) did not affect the spike rate reduction. This suppressive effect of high nonselective effort on spiking can be understood in the context of spike response normalization. To gain a mechanistic understanding of the observed interactions between attentional selectivity and nonselective effort seen in V4 responses, we tested whether a simple extension of a sensory normalization model (*29*) with spatially distributed behavioral *d*′ can account for these effects on spike responses. Mean stimulus–evoked spike counts were expressed as $r = (d'_{in} * E_{in,G} + d'_{opp} * E_{opp,G})/(d'_{in} * S_{in,G} + d'_{opp} * S_{opp,G} + \sigma)$, where d'_i represents behavioral *d*′ at location *i* in or opposite the RF hemifield; $E_{i,G}$ and $S_{i,G}$ represent excitation and suppression, respectively, at location *i* due to the Gabor stimulus ($G = 1$) or the background ($G = 0$); and σ is baseline suppression (Fig. 6A and Materials and Methods). Model parameters $E_{i,G}$, $S_{i,G}$, and σ were fit for each unit with the trial-averaged spike counts over 200 ms during the prestimulus fixation, sample, and test interval periods on training datasets. Performance of the full model (model with *d*′) was measured on the fourfold cross-validation test datasets subsampled across different stimulus configurations and attention conditions and compared with two alternate models that lacked any behavioral *d*′ factors: model without *d*′, and model without *d*′ and background display (Materials and Methods).

The heatmap in Fig. 6B shows mean spike counts fitted with the normalization model (model with *d*′) as a function of behavioral *d*′s in two hemifields for an example unit in a task difficulty session. Most units in the two task contexts were better fit with the attention model of normalization compared with alternate *d*′-independent normalization models [difficulty session: model with *d*′, 910/1194 (76%); model without *d*′, 626/1194 (52%); model without *d*′ and background display, 479/1194 (40%); reward session: model with *d*′, 1006/1331 (76%); model without *d*′, 614/1331 (46%); model without *d*′ and background display, 486/1331 (37%); variance explained >80%; Fig. 6, C and D]. At the population level, the quality of normalization model fits of spike responses in the two task contexts did not differ ($P = 0.23$, chi-square test). The normalization model with *d*′ captured multiple features of observed spike counts across different stimulus and attention conditions, including relative changes in spike counts across different attention conditions, and neuronal modulation indices for attentional selectivity and effort (fig. S16 and Fig. 6, E to H). This model also predicts a suppression of spiking for

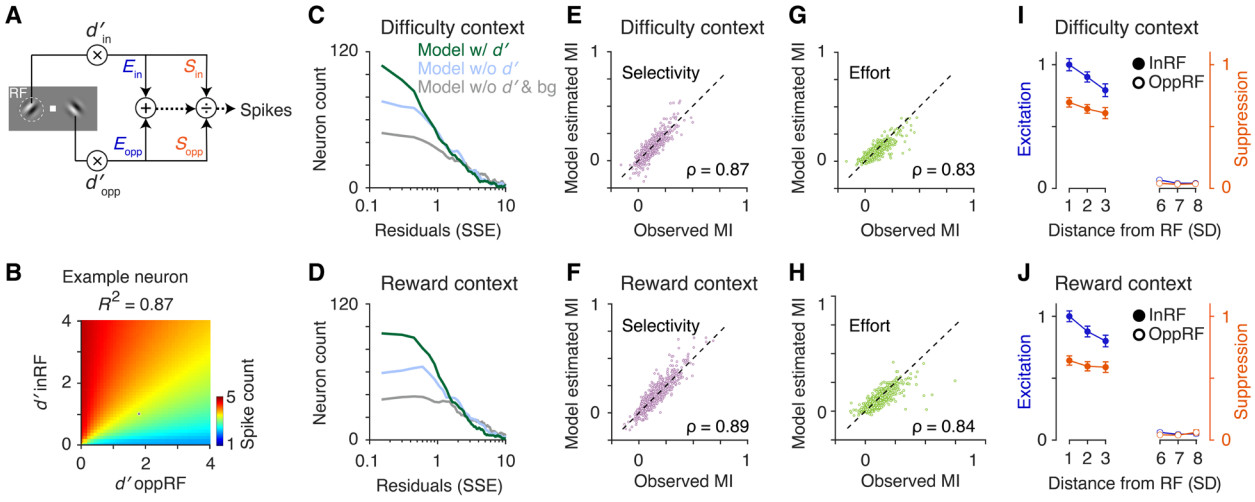

Fig. 6. Normalization model of attention can account for attentional effort effects on spike counts. (**A**) Normalization model of attention. Spike count, $r = (d'_{in} * E_{in,G} + d'_{opp} * E_{opp,G})/(d'_{in} * S_{in,G} + d'_{opp} * S_{opp,G} + \sigma)$, where, d'_i is the behavioral d' at location i (inRF or oppRF). $E_{i,G}$ and $S_{i,G}$ are excitation and suppression, respectively, at location i due to either the Gabor stimulus ($G = 1$) or the background ($G = 0$). σ is a constant. Spike counts were fitted with three different models (see Materials and Methods). Model w/ d': Contains behavioral d' values in two stimulus locations. In addition, there are excitation and suppression terms for background display. Model w/o d': Same as the previous model except without the d' terms ($d'_i = 1$). Model w/o d' & bg: Same as the previous model, except without d's and excitation/suppression parameters due to the background ($d'_i = 1$, $E_{i,G=0} = 0$, $S_{i,G=0} = 0$). (**B**) An example neuron fit with the normalization model with d'. Heatmap, fitted spike counts. (**C** and **D**) Fit qualities [sum of squared error (SSE)] for three different normalization models for all units in task difficulty and reward sessions. (**E** and **G**) Comparing spike count modulations [modulation index (MI)] with attentional selectivity (E) and effort (G) between observed and model fits for cross-validation test datasets in task difficulty sessions. MIs from observed and fitted spike counts are strongly correlated (Spearman correlation coefficient: selectivity, $\rho = 0.87$, $P < 10^{-100}$; effort, $\rho = 0.83$, $P < 10^{-100}$). (**F** and **H**) Same as in (E) and (G) for reward sessions (Spearman correlation coefficient: selectivity, $\rho = 0.89$, $P < 10^{-100}$; effort, $\rho = 0.84$, $P < 10^{-100}$). (**I** and **J**) Model fitted (model with d') population averaged excitation (E) and suppression (S) and binned according to the RF-stimulus distance (same or opposite of the RF hemifield).

high selective attention in the opposite RF location compared with low attentional effort (shift from right to left, Fig. 6B). Owing to differences in the overall effort levels, we expected to see a smaller change in spiking when attention shifts from high selective in oppRF location to low effort compared with the behavioral d' change of the same extent from high effort to high selective attention in inRF location. Spike rate PSTHs (Figs. 2B and 4, A and B) indicate such effects in few of the later time bins when top-down attentional influence in V4 might dominate stimulus drive.

Population correlations between the observed and normalization model–estimated spike count modulation indices for attentional selectivity and effort were strong across the task contexts (difficulty context: attentional selectivity, $\rho = 0.87$, $P < 10^{-10}$; attentional effort, $\rho = 0.83$, $P < 10^{-10}$; reward context: attentional selectivity, $\rho = 0.89$, $P < 10^{-10}$; attentional effort, $\rho = 0.84$, $P < 10^{-10}$; Spearman correlation coefficient; Fig. 6, E to H). Further, the model fitted excitatory and suppressive stimulus drives of recorded neurons decreased with the proximity of the neuron's RF and the stimulus irrespective of task context that motivated animals to attend (Fig. 6, I and J). Together, these results show that a simple normalization model captures the effects of attentional engagement at a distant site on the spike response in the RF location regardless of the stimulus or attentional context of that RF response.

DISCUSSION

We isolated the contributions of attentional selectivity and nonselective effort to the activity of individual V4 neurons while precisely and independently controlling monkeys' behavioral d' at the RF location and a distant location in the opposite hemifield. Changes in either attentional selectivity or effort independently are associated with overall increases in V4 spike rates and decreases in V4 spiking variability and pairwise spike count correlations. Further, spike rates were reduced when behavioral d' increased at a distant location in the opposite hemifield. A spatially tuned response normalization model explained all these changes in spike rate across attention conditions and task contexts. Last, single-trial encoding of attentional selectivity, effort, and their interaction in V4 neurons were found to be independent of the way the subject is motivated to regulate its attention.

A functional role for the spiking of V4 neurons is supported by a correlation with enhanced sensory processing of an attended stimulus at the RF location, as well as impaired behavioral detection in subjects with V4 lesions. V4 lesions in macaques and humans impair attentional performance by making it difficult for the subjects to exclude irrelevant distractors (*30*, *31*). However, our results reveal a mismatch between V4 spiking and behavioral performance: lower spike rates for the same behavioral performance when monkey's attention strategy shifts from spatially selective to nonselective high effort at the RF location (Figs. 2B and 4A). Although the reduced spike rates do not alter encoding of relevant sensory information (figs. S5 and S6), this raises questions about how neuronal signals in V4 are relayed to downstream brain areas and influence a subject's performance. Other studies also found a dissociation between behavioral performance and activity in cortical visual areas, including V4 (*5*, *28*, *32*). That work showed dissimilar dynamics of top-down attentional control signals, sensory modulation, and executive action. Specifically, when attention shifted, behavioral changes lagged changes in the spike rates of sensory neurons by seconds to minutes, breaking the link between spikes and performance (*28*). Our results show

that a mismatch between V4 activity and behavior can also go beyond lags: When animals spread their field of attention beyond the RF, V4 spike rates go down even though the animals maintained the same behavioral d'. While a normalization mechanism can explain the reduced spiking (Fig. 6), it does not address this disconnect with behavior. Changes in pupil diameter suggest that the animal increased its total effort at both sites to maintain performance in the high effort condition (Figs. 1F and 3E), making the reduced spike rates even less expected. Nevertheless, there is little reason to believe that behavioral performance should be uniquely determined by the strength or quality of sensory signals in any one brain region. A primate brain doing an attentionally demanding task depends on contributions from many structures throughout the neuraxis. Even if perceptual stages perform perfectly, the overall behavioral performance can be affected by distractions and lapses associated with activity in other brain stations. The somewhat reduced spike rates in V4 that occur with high attention to both hemifields might lower the performance but be counterbalanced by attention-related changes in other structures that enhance performance by reducing errors related to factors like distractions or motor error.

In the current experiments, monkeys were motivated to direct their spatial attention by expectation of either larger rewards or higher task demands associated with two locations. The invariance of neuronal modulations in V4 with these two distinct motivational factors suggests that these effects depend on a common top-down cognitive control and do not represent reinforcement signals to any appreciable extent. This notion is further supported by previous reports of weak evidence of encoding of reward information by single-trial spike counts of V4 neurons across different attention states and during state transitions (*28*) [but see (*5*)]. Motivation plays a crucial role in influencing attention control, giving priority to the most appropriate goal among multiple competing targets. Both humans and animals can be motivated to execute actions either for intrinsic pleasure or for satisfying some basic needs such as hunger and thirst. Several brain structures within the frontoparietal network are modulated by motivationally salient signals such as errors, rewards, and penalties (*33*). Although various forms of motivations (intrinsic and extrinsic) can have different origins, they have common nodal points in the striatum and prefrontal cortex that receive dopaminergic afferents that play a crucial role in reward learning (*34*). Consistent with this, previous evidence points to common cortico-limbic neural pathways that are activated by either changes in expectation of reward or changes in task difficulty (*35*).

Our results show a close relationship between attentional effort and attentional "intensity" (*28*). In both the manipulations, neuronal modulations with nonselective attention are spatially nonselective unlike selective attention as they were independent of the neuron's RF location and less likely due to split-foci selective attention (*28*). Previous work has associated effort exclusively with changes in task difficulty and viewed it as a specific type of "arousal" that is distinguishable from other generic forms of arousal elicited by factors such as stress, novel stimuli, and drugs, and often detrimental to the performance (*16*) [but see (*36*)]. The attentional effort in our study can be conceptualized as a form of arousal or readiness (*37*, *38*) that facilitates performance in top-down attention tasks. The manifestation of attentional effort on regulating visual sensory processing of V4 neurons in both of our task contexts supports identifying attentional effort as an intensive aspect of attention (*16*). It is possible that the component of bottom-up stimuli-driven arousal that affects performance might map well onto the neuronal modulation associated with attentional effort or intensity. Future experiments with precise and independent control of these cognitive components in simultaneous tasks might identify their precise relationships and how subjective experience of arousal relates to attentional effort. Other important questions to be addressed concern the cellular and molecular mechanisms that mediate attentional effort. These could involve activation of diverse neuromodulatory systems such as norepinephrine, acetylcholine and serotonin (*6*, *39*–*43*).

V4 spike responses for varying attentional selectivity and effort were well explained by an extension of a normalization model of visual responses (*44*) with spatially tuned attentional gain factors represented by behavioral d'. A similar normalization model with uniform attentional effects on excitation and surround suppression has been previously used in explaining neuronal modulations in V4 with spatially selective attention (*29*). Previous normalization models considered attentional effects on stimulus-induced excitation and surround suppression. To explain a modest but nonzero neuronal modulation during the fixation period in the absence of any visual stimulus, our normalization model included an attention effect on the background display similar to the visual stimulus. This is consistent with the evidence that attention acts as a constant gain factor (*3*), which affects baseline neuronal activity. We considered the excitatory and inhibitory intrinsic sensory drives in the absence of attention to a location as proportional to the Gabor stimulus contrast and orientation. In our normalization model of attention, the effective stimulus drive (excitatory and inhibitory drive) is the product of intrinsic sensory drive and a constant multiplicative factor of behavioral d' at that location. We have assumed that this multiplicative factor remains uniform during three task phases: over 200 ms presample fixation, during sample, and test stimulus. We think that this is a reasonable assumption because attention was controlled within blocks of trials in our task, as opposed to other designs where different attention conditions are interleaved trial by trial (*5*). A previous study using a similar task showed that block-averaged single-trial behavioral d' remains stable over time within an attention condition (*28*). Future studies with combinations of multiple stimulus configurations and attention conditions would be ideal to tease apart whether stimulus drives alter within a trial. Similar to reward expectations, increased task difficulty associated with spatial attention increases visual excitation (*13*) and response suppression (*20*) of V4 neurons. Further, response suppression with high task load is considered to serve as a mechanism for reducing peripheral interference and improving signal detection (*45*). These reports are inconsistent with the correlated decrease in model-estimated excitation and suppression with an increase in the proximity of neuron's RF and attended stimulus across task difficulty and reward contexts. Together, this normalization model of attention provides a canonical neuronal computation to explain how distributed spatial attention influences neuronal responses. Future experiments are required to examine how other forms of attention such as feature-based or bottom-up attention act on normalization mechanisms across different visual areas that are responsive to attentional modulation.

Together, our results provide previously unidentified experimental evidence revealing how attentional selectivity and nonselective effort interact and modulate sensory processing in visual cortex in reference to behavioral performance. Moreover, our study identified a computational mechanism of normalization through which spatially distributed attentional performances interact.

MATERIALS AND METHODS

Subjects and surgery

Two adult male rhesus monkeys (*Macaca mulatta*, 13 and 9 kg) were implanted with a titanium head post using aseptic surgical techniques before training began. After the completion of behavioral training (3 to 5 months), we implanted 10 × 10 array microelectrodes with 400 μm spacing (Blackrock Microsystems) into the dorsal visual area V4 of one hemisphere, between the lunate and superior temporal sulci. The same two monkeys were used in a previous study that included some of the same neuronal responses but described different findings (*28*). All experimental procedures were approved by the Institutional Animal Care and Use Committee protocols (no. 72355) of the University of Chicago and were in compliance with the U.S. National Institutes of Health guidelines.

Behavioral task

During training and neurophysiological recording, the monkey sat in a primate chair facing a calibrated CRT display (1024 × 768 pixels, 100 Hz refresh rate) at 57 cm viewing distance inside a darkened room. Binocular eye position and pupil area were recorded at 500 Hz using an infrared camera (EyeLink 1000, SR Research). Trials started once the animal fixated within 1.5° of a central white spot (0.1° square) presented on a mid-level gray background (Fig. 1A). The animal had to maintain fixation until its response at the end of the trial. After a fixation period of 400 to 800 ms, two achromatic Gabor sample stimuli appeared for 200 ms, one in each visual hemifield. After a variable delay of 200 to 300 ms, a Gabor test stimulus (test 1) appeared for 200 ms at one of the two target locations, randomly selected with equal probability. The test stimulus was identical to the preceding sample stimulus, except potentially its orientation. On half of the trials, the test 1 stimulus had a different orientation (nonmatch trial), and the monkey had to make a saccade to that target to receive an apple juice reward. On the remaining half of the trials, the test 1 stimulus had the same orientation as the corresponding sample stimulus (match trial), and the monkey had to maintain fixation until a second test stimulus with a different orientation (test 2, 200 ms) appeared in the same location after an additional delay of 200 to 300 ms. The monkey then had to saccade to that target to get a reward. Intertrial intervals varied from 2 to 3 s. Stimuli were presented always in the lower hemifields at eccentricity of 2° to 4°. Gabors were static and odd-symmetric with the same average luminance as the background. Spatial frequency, size, and base orientation of Gabor stimuli were optimized for one neuron recorded each day and remained unchanged throughout each session (left: azimuth, −2.5° to −4.5°; elevation, −0.5° to −4.0°; sigma, 0.35° to 0.70°; spatial frequency, 0.6 to 3.5 cycles per degree; right: azimuth, 1.8° to 5.5°; elevation, −0.5° to −4.0°; sigma, 0.25° to 0.58°; spatial frequency, 0.7 to 3.0 cycles per degree). On every trial, the orientation of the sample stimuli randomly took one of two values (independently): base orientation or orthogonal (base +90°). Stimulus parameters and orientation changes remained fixed within a session and varied across sessions. Orientation changes differed between blocks when task difficulty was manipulated, but every block contained probe trials that had the same orientation change throughout a session (24° to 40° for monkey S, and 20° to 40° for monkey P). Reward sizes for hits (correct response in nonmatch trial) and correct rejections (CRs in match trial) were adjusted by <10% as needed to encourage the animal to maintain a behavioral criterion close to zero. Behavioral task was controlled using custom-written software (https://github.com/MaunsellLab/Lablib-Public-05-July-2016.git).

Behavioral task contexts

Animals were motivated to allocate their spatial attention using two different task contexts: varying task demand or varying reward size. In alternate sessions, animal's spatial distribution of behavioral d' at two locations in opposite hemifields was controlled by either of the task contexts. In the task difficulty context, selective attention and nonselective attentional effort were controlled over interleaved blocks of trials (160 to 440 trials per block; monkey S, mean = 254, SD = 41; monkey P, mean = 285, SD = 76) by changing relative task difficulty at the two locations in opposite hemifields (Fig. 1C). Orientation change randomly took one of two values: probe orientation change (~30% of the trials) or contextual orientation change (~70% of the trials). Contextual orientation change was small for difficult task (high task demand) and large for easy task (low task demand) compared to the probe orientation change. A high behavioral d' (selective attention) at location 1 relative to location 2 was achieved by making the task difficulty high at the location 1 ($\Delta\Theta_{context}$, 17° to 20° for monkey S and 8° to 22° for monkey P) and easy at the location 2 ($\Delta\Theta_{context}$, 80° to 90° for monkey S and 80° to 90° for monkey P). A high nonselective behavioral d' (high nonselective attentional effort) was obtained by making the task difficulty high at both the locations ($\Delta\Theta_{context}$, 15° to 18° for monkey S and 6° to 22° for monkey P). A low nonselective behavioral d' (low nonselective attentional effort) was obtained by making the task difficulty easy at both the locations ($\Delta\Theta_{context}$, 80° to 90° for monkey S and 80° to 90° for monkey P) (Fig. 1B). Reward values for correct behavior responses were always the same across blocks on both sides and fixed.

In the differential reward context, selective attention and nonselective attentional effort were controlled over interleaved blocks of trials (120 to 220 trials per block) by changing reward size at the two locations (Fig. 3B). Reward manipulation sessions were interleaved with the task difficulty manipulation session over the course of the experiment (fig. S4). There was only a single orientation change that remained fixed throughout each experimental session (for monkey S: mean = 33.4°, SEM = 1.5°, N = 20; for monkey P: mean = 30.6°, SEM = 1.1°, N = 16). A high selective behavioral d' (selective attention) at location 1 relative to location 2 was achieved by delivering high rewards at location 1 compared to location 2. High nonselective behavioral d' (high nonselective attentional effort) was controlled by delivering high rewards for correct responses at both locations. Similarly, a low nonselective behavioral d' (low nonselective attentional effort) was controlled by giving low rewards for the correct responses at both locations. Previous studies have shown that when monkeys are instructed to shift from attending to one location to attending equally to two locations, behavioral performance is impaired (*5*, *46*). To achieve equivalent performance, we elevated monkeys' motivation in our task during high nonselective d' conditions compared with high selective d' condition either by increasing the reward size (reward context) or increasing task difficulty (difficulty context). Animals were encouraged to maintain a behavioral target/nontarget criterion close to zero by small adjustment of trial-by-trial reward ratio for hits and CRs (figs. S1 and S7 and tables S3 and S4). We expect that any differences in the criteria across conditions would not be reflected in the neuronal responses of V4 neurons, as was shown in previous studies (*14*).

Electrophysiological recording and data collection

Extracellular neuronal signals from the chronically implanted multielectrode array were amplified, band-pass filtered (250 to 7500 Hz), and sampled at 30 kHz using a data acquisition system (Cerebus, Blackrock Microsystems). We simultaneously recorded from multiple single units and multiunits over 42 differential task difficulty sessions (714 units and 20 sessions for monkey S; 480 units and 22 sessions for monkey P) and 36 differential reward sessions (850 units and 20 sessions for monkey S; 481 units and 16 sessions for monkey P). At the start of each experimental session, we mapped RFs and stimulus preferences of neurons while the animal fixated. These RFs were used to optimize the stimulus parameters. Spikes from each electrode were sorted offline (Offline Sorter, Plexon Inc.) by manually well-defining cluster boundaries using principal components analysis and waveform features. Well-isolated clusters were classified as single units from multiunits based on the isolation quality of unit clusters. The degree to which unit clusters were separated in two-dimensional (2D) spaces of waveform features (first three principal components: peak, valley, and energy) was measured by multivariate analysis of variance (MANOVA) F statistic using Plexon Offline Sorter (Plexon Inc.). A unit cluster of MANOVA $P < 0.05$ was considered as a single unit that indicates that the unit cluster has a statistically different location in 2D space and that the cluster is statistically well separated. The proportions of single units (single units/total units) between the task contexts did not differ for either monkey (fig. S4).

We performed a calculation to explore whether the units modulated in one context might come from a different population than those modulated in the second context. The probability of sampling n units out of N (average sample size in a session) in context 1 that do not overlap with units sampled from population 2 (context 2) can be represented by hypergeometric distribution

$$P(n \mid N, N_2, N_1) = \frac{\binom{N-N_2}{n} * \binom{N_2}{N_1-n}}{\binom{N}{N_1}}; n < N_1$$

where (:) are binomial coefficients, N is the average units recorded in a session, N_1 is the average number of units with positive modulation with attention [modulation index (MI) > 0] in context 1, and N_2 is the average number of units with positive modulation in context 2. Probabilities for different n's (new samples) were estimated (fixed N, session averaged observed value). The percent of new samples (100*n/N) for which the probability was maximum measures the highest proportion of nonoverlapping populations between the two contexts.

Data analysis

Behavioral sensitivity (d') and criterion

All completed trials in the reward context and all probe orientation trials in the task difficulty context were included in our analysis. Behavioral sensitivity (d') and criterion (c) at a spatial location were measured from hit rates within nonmatch trials and false alarm (FA) rates within match trials as

$$d' = \Phi^{-1}(\text{hit rate}) - \Phi^{-1}(\text{FA rate})$$

$$c = -\frac{1}{2}[\Phi^{-1}(\text{hit rate}) + \Phi^{-1}(\text{FA rate})]$$

where Φ^{-1} is inverse normal cumulative distribution function. We measured the average d' and c within a session across all trials across blocks separately for four different attention conditions.

Index of attentional selectivity and effort

Attentional selectivity was measured by the relative value of the behavioral d' inside the RF location with respect to opposite RF location. This measure mapped directly onto polar angle in d' space (Figs. 1D and 3C) and was normalized to a range from −1 (inside RF hemifield $d' = 0$) to 1 (opposite RF hemifield $d' = 0$)

$$\text{selectivity index} = \frac{4}{\pi}\tan^{-1}\left(\frac{d'_{\text{inRF}}}{d'_{\text{oppRF}}}\right) - 1$$

where d'_{InRF} and d'_{oppRF} are the sensitivities in the two hemifields, inside and outside the recorded neurons' RFs, respectively. Attentional effort represented the absolute value of total behavioral d' (distance from the origin in d' space)

$$\text{effort index} = \sqrt{(d'^2_{\text{inRF}} + d'^2_{\text{oppRF}})}$$

Pupil area

All pupil area measurements were measured binocularly at 500 Hz while monkeys maintained fixation in the absence of a luminosity change using infrared camera (EyeLink 1000, SR Research). Raw pupil areas were z scored for each session and each eye separately. Mean pupil area was measured by averaging the z-scored pupil area during the 400 ms after sample appearance.

Neuronal response modulation

Only neurons with an average spike rate 60 to 260 ms after sample stimulus onset that was significantly ($P < 0.05$) greater than the rate of 0 to 250 ms before sample onset were used in the analysis. To construct PSTHs for figures, spike trains were aligned to sample stimuli onset and averaged across trials and smoothed with a half Gaussian kernel (rightward tail, SD = 15 ms). Most analyses of neuronal response modulation used spiking during a 200 ms sample stimulus window (60 to 260 ms after sample onset). Behavioral performance depends critically on attentional engagement to the sample stimuli. The sample stimuli were kept brief to reduce opportunities for the animal to adjust its attentional allocation in response to those stimuli and to ensure stable attention across stimulus conditions. In addition, the sample stimulus period, unlike the test stimulus period, was relatively free from contaminating factors such as choice confidence, response preparation, and reward expectation. A spike rate modulation (Figs. 2, C and D, and 4, C and D) was quantified by neuronal d' as

$$d'_{\text{neuron}} = \frac{\langle r_{\text{high}} \rangle - \langle r_{\text{low}} \rangle}{\sqrt{\frac{1}{2}(\sigma^2_{\text{high}} + \sigma^2_{\text{low}})}}$$

where $<r_x>$ and σ_x are the average and SD of the spike counts within 60 to 260 ms from sample stimuli onset. Neuronal d's were calculated for each unit separately for attentional selectivity and effort. Neuronal MI (Fig. 6, E to H) was measures as

$$\text{MI} = \frac{\langle r_{\text{high}} \rangle - \langle r_{\text{low}} \rangle}{\langle r_{\text{high}} \rangle + \langle r_{\text{low}} \rangle}$$

Visual response latency

Spike rate latency (fig. S14) was measured by the time to half peak response (*47*). Mean spike rate PSTHs (aligned to sample stimulus onset) of each unit were smoothed with Gaussian filter (sigma, 8 ms) for different attention conditions. The SEM of baseline response was estimated from the presample (−200 to 0 ms) responses during fixation. The peak response between 50 and 200 ms from the sample onset was measured with the requirement that it exceeded mean baseline by 3.72*SEM. Units were not considered if the above criteria were not met. The response latency was computed as the time the PSTH reached half the difference between the peak response and mean baseline response. Latencies across attention conditions and contexts were compared using repeated-measures ANOVA with one between-subject (context: task difficulty and reward) and two within-subject factors (attentional selectivity: low versus high; and nonselective effort: low versus high).

Stimulus orientation decoding using demixed principal components analysis

Sensory information encoding was compared between the high selective attention and high effort conditions using decoding accuracy for sample stimulus orientations at the two stimuli locations (fig. S6) based on demixed principal component decompositions (*27*). Stimulus orientation information was isolated from other task variables using simultaneously recorded population spike trains. Details of the demixed principal component decompositions have been described by Kobak *et al.* (*27*). Briefly, population activity patterns from all simultaneously recorded neurons in every session were decomposed into a linear combination of specific components, each of which carries information of a single task variable. Mean-subtracted and trial-averaged spike trains of each neuron were decomposed into sum of marginalized averages, each corresponding to a task variable and a noise term. This marginalization process ensures that the individual components are uncorrelated. A loss function that penalizes the difference between the marginalized data and the reconstructed full data is minimized using the least-square method. The reconstructed data are the full data projected with the decoders onto a low-dimensional latent space and then reconstructed with the encoders. Trials were classified into two stimulus conditions: according to the Gabor orientation in the RF location (two orientations: $base_{inRF}$ and $base_{inRF}$ + 90°) and in the opposite RF location (two orientations: $base_{oppRF}$ and $base_{oppRF}$ + 90°) during the sample period. Thus, there were four different configurations of stimulus orientations. Single-trial spike rates were filtered with a half Gaussian kernel (σ = 50 ms) and subsampled at 100 Hz. Spike rates over 400 ms (40 time points) starting from −50 to 350 ms from the sample stimulus onset were used for the analysis. Decomposition into demixed components was performed on training datasets (leave-one-out, 1000 repetitions). The stimulus orientation was then decoded on the remaining cross-validated test trials using the top three components to estimate the decoding accuracy.

Proximity between neuron's RF and sample stimuli

Proximity between each neuron's RF and sample stimulus was estimated by the Mahalanobis distance (*28*) (standardized distance, Figs. 2D and 4D). For each neuron, the spatial RF was measured and fit using a bivariate Gaussian. We then calculated the Mahalanobis distance between probability densities of the RF and the Gaussian contrast profile of the Gabor stimulus in the RF location. Spatial overlap varied from 0% (minimal overlap) to 100% (maximum overlap). This measure of the RF-stimulus overlap (proximity) accounts for both the alignment of the stimulus with the neuron's RF and the correspondence between stimulus and RF size. It does not take stimulus orientation and other nonspatial factors into account. Neuronal modulation was measured using spike counts in both trial types (two sample orientations, base and base +90°). Across the population of recorded units, there was no systemic bias on the neuron's preferred orientation. Further, the sample orientation (base) varied in every session. Hence, we expect an unbiased distribution of overlaps between sample orientation and the neuron's preferred orientation across clusters with different spatial RF-sample stimulus proximity. The correlation coefficients between the neuronal modulations (neuronal d') and the RF-stimulus proximity was measured by Spearman correlation coefficients that are insensitive to the magnitudes of neuronal modulations.

Fano factor

Mean-matched Fano factor (Figs. 2E and 4E) was measured using spike counts over 100-ms sliding windows in 2-ms steps for each neuron according to procedures described previously (*28*, *48*). The variance and mean across trial were computed at every time bin. The greatest common distribution of means across neurons, attentional intensities, and time bins was measured. To match the mean distribution to the common mean distribution, a different subset of neurons was randomly chosen (50 times) at every time bin, and the average Fano factor was computed (ratio of the variance to the mean). Mean-matched Fano factors in different attention conditions were compared using repeated-measures ANOVA with three within-subject factors: attentional selectivity (low versus high), nonselective effort (low versus high), and the time from sample onset (60 to 260 ms).

Spike count correlations

Pearson correlation coefficients were computed for pairs of simultaneously recorded units on spike counts over 200 ms (60 to 260 ms from sample stimuli onset), defined as the covariance of spike counts normalized by the variances of individual neurons

$$\rho_{12} = \frac{\mathrm{Cov}(r_1, r_2)}{\sqrt{\mathrm{Var}(r_1) * \mathrm{Var}(r_2)}}$$

where r_1 and r_2 are spike counts of neuron 1 and neuron 2 across trials, respectively. Pairwise spike count correlations were binned according to the geometric mean of the evoked responses of the two neurons in 5 Hz intervals. Evoked response was computed by subtracting the trial-averaged baseline spike rate (−200 to 0 ms from sample onset) from the trial-averaged spike rate during the sample (60 to 260 ms from sample onset) (Figs. 2F and 4F). Spike count correlations were compared using repeated-measures ANOVA with three within-subject factors: attentional selectivity (low versus high), nonselective effort (low versus high), and sample stimulus orientation (base and base +90°).

Generalized linear encoding model

Generalized linear model (GLM) regression was used to estimate the encoding of different attention components and task variables in single-trial spike trains. Spike counts (r) over 50 ms bins with 10 ms shift in single trials were modeled to follow a negative binomial distribution. The negative binomial distribution is well suited for the purpose, as spike count variances of cortical neurons are most often equal to or greater than their means (*28*, *48*, *49*). The details of the model implementation were described in an earlier study (*28*). Briefly, expected value of spike count at each time bin according to the GLM was represented as

$$\mu_r = \exp(\beta_0 + \beta_{sel}\,Sel + \beta_{effort}\,Effort + \beta_{sel*effort}\,Sel*Effort + \beta_{ori}\,Ori + \beta_{sac}\,Sac)$$

where β_i is the coefficient for the predictor variable i. The signs of β coefficients indicate the direction of the spiking change between different levels of the corresponding predictor variable. Sel is the session averaged attentional selectivity, ratio of the behavioral d' at the RF location to the d' at the opposite hemifield location; Effort is the session averaged attentional effort, distance from the origin in d' space; Sel * Effort is the interaction between selectivity and effort; Ori is the orientation of sample stimulus inside the neuron's RF; and Sac is the saccade choice (1 for saccade toward the RF, −1 for saccade opposite to the RF, and 0 for saccade withheld). Encoding results did not change whether the selectivity was represented by d' ratios or selectivity index (fig. S13). For the reward context dataset, we also used another alternate GLM containing an additional predictor variable, reward history (fig. S12), where the expected spike count was represented as

$$\mu_r = \exp(\beta_0 + \beta_{sel}\,Sel + \beta_{effort}\,Effort + \beta_{sel*effort}\,Sel*Effort + \beta_{ori}\,Ori + \beta_{sac}\,Sac + \beta_{reward}\,Reward)$$

Here, reward represents the average of reward values on the three immediately preceding trials. GLM was implemented in MATLAB separately for each neuron. To compare different predictor coefficients, predictor variables were converted to z-scored values and fitted with GLMs to obtain standardized β coefficients. Goodness of fit for a given GLM was measured by residual deviance, pseudo R-squared value (Cragg and Uhler's method), and F statistics compared to a null model.

Predictive performance of the GLMs was measured by cross-validation. Observations in each neuron's dataset were split at random into 10 partitions. GLM fit was done on nine training partitions, and the remaining partition was used for cross-validation. This cross-validation error for each neuron was measured by

$$\text{error} = \frac{1}{nk}\sum_{t=1}^{k}\sum_{i=1}^{n}(y_{i,t} - \hat{y}_{i,t})^2$$

where n is the number of cross-validation trials; k is the number of time bins; and $y_{i,t}$ and $\hat{y}_{i,t}$ are recorded and GLM-estimated spike counts, respectively, at tth time bin in ith trial. The quality of cross-validation for neuron populations was measured by the Spearman correlation coefficients between the observed and GLM fit spike counts.

Relative importance of each predictor variable was measured by PI in Fig. 5 expressed as the absolute z statistic of each fitted predictor coefficient (both positive and negative)

$$PI_{j,t} = \left|\frac{\beta_{j,t}}{SE(\beta_{j,t})}\right|$$

where j = 1, 2, 3, …, m are the predictor variables and t is the time bin. A neuron with predictor coefficient different from zero ($P < 0.05$; t test) during the sample stimulus presentation (60 to 260 ms from sample on) was classified as sensitive to that predictor. Because of the exponential nonlinearity in the GLM, exponentiated fitted predictor coefficients were used to illustrate how attentional selectivity, effort, and their interaction (selectivity by effort) influence on spike rates as a function of attentional selectivity and effort (Fig. 5C, left) as well as a function of behavioral d's at the RF location and opposite hemifield location (Fig. 5C, right). These components are combined multiplicatively to drive instantaneous spike rates of individual units.

Normalization model

Trial-averaged spike counts over 200 ms across all attention conditions and stimulus configurations were fit with three different normalization models. Two were simple stimulus normalization models without any spatial tuning (model without d' and background display, and model without d'), and the third model was an extension of spatially tuned normalization model (model with d'). All the three models were fit with nine nonnegative parameters and used the same set of 27 of 36 training data points using nonlinear least-squares solver (MATLAB lsqnonlin.m, MathWorks). The quality of fit was measured by residuals using nine cross-validation test data points (fourfold cross-validation). The training dataset consisted of mean spike counts during three presamples (−200 to 0 ms), 12 sample stimuli (60 to 200 ms) across two Gabor orientations and four attention conditions, six test 1 stimuli (60 to 200 ms) on left hemifield, and six test 1 stimuli on right hemifield across all four attention conditions and two Gabor orientations. Cross-validation (fourfold) test dataset contained spike counts of 1 during presample, 4 during the sample stimuli, 2 during test 1 stimulus inside the RF location, and 2 during test 1 stimulus in the opposite RF location across all attention conditions and Gabor stimulus configurations.

Model without d′ and background display. Mean spike counts during the sample stimulus period according to the normalization model without d' and background display are described as

$$r_{in,opp} = \frac{E_{in} + E_{opp}}{S_{in} + S_{opp} + \sigma}$$

where fit parameters E_i and S_i represent excitation and suppression drives, respectively, due to a Gabor stimulus at ith location (inRF or oppRF), and σ represents a constant baseline suppression. For every E_i (or S_i), there are two parameters, $E_{i,base}$ and $E_{i,base+90}$ (or $S_{i,base}$ and $S_{i,base+90}$), associated with each of the two Gabor orientations (base and base +90°). The contrast term was 1 for the Gabor stimulus and 0 for no stimulus. Thus, the mean spike counts during the inRF and oppRF test 1 presentations, respectively, are

$$r_{in,0} = \frac{E_{in}}{S_{in} + \sigma} \text{ and } r_{0,opp} = \frac{E_{opp}}{S_{opp} + \sigma}$$

The mean of spike counts during the presample period was zero.

Model without d′, with background display (model without d′). According to the normalization model without d', mean spike counts during the sample, test 1, and presample periods are described, respectively, as

$$r_{in,opp} = \frac{E_{in} + E_{opp}}{S_{in} + S_{opp} + \sigma} \text{ (sample)}$$

$$r_{in,0} = \frac{E_{in} + E_{opp,0}}{S_{in} + S_{opp,0} + \sigma} \text{ (test 1 inRF)}$$

$$r_{0,opp} = \frac{E_{in,0} + E_{opp}}{S_{in,0} + S_{opp} + \sigma} \text{ (test 1 oppRF)}$$

$$r_{0,0} = \frac{E_{\text{in},0} + E_{\text{opp},0}}{S_{\text{in},0} + S_{\text{opp},0} + \sigma} \text{ (presample)}$$

where E_i (or S_i) could be either $E_{i,\text{base}}$ or $E_{i,\text{base+90}}$ (or $S_{i,\text{base}}$ and $S_{i,\text{base+90}}$) excitatory (or suppressive) stimulus drives at ith location (inRF or oppRF) due to the two Gabor orientations; $E_{i,0}$ (or $S_{i,0}$) is the excitatory (or suppressive) drive due to the background display in the absence of Gabor stimulus. Only one common parameter was used for excitatory (or suppressive) drives at the opposite RF location for both Gabor stimulus and background display, i.e., $E_{\text{opp,base}} = E_{\text{opp,base+90}} = E_{\text{opp},0}$ and $S_{\text{opp,base}} = S_{\text{opp,base+90}} = S_{\text{opp},0}$. A total of nine parameters were fit with the model.

Spatially tuned normalization model with d′ and background display (model with d′). According to the normalization model with d', mean spike counts during the sample, test 1, and presample periods are described, respectively, as

$$r_{\text{in,opp}} = \frac{d'_{\text{in}} E_{\text{in}} + d'_{\text{opp}} E_{\text{opp}}}{d'_{\text{in}} S_{\text{in}} + d'_{\text{opp}} S_{\text{opp}} + \sigma} \text{ (sample)}$$

$$r_{\text{in},0} = \frac{d'_{\text{in}} E_{\text{in}} + d'_{\text{opp}} E_{\text{opp},0}}{d'_{\text{in}} S_{\text{in}} + d'_{\text{opp}} S_{\text{opp},0} + \sigma} \text{ (test 1 inRF)}$$

$$r_{0,\text{opp}} = \frac{d'_{\text{in}} E_{\text{in},0} + d'_{\text{opp}} E_{\text{opp}}}{d'_{\text{in}} S_{\text{in},0} + d'_{\text{opp}} S_{\text{opp}} + \sigma} \text{ (test 1 oppRF)}$$

$$r_{0,0} = \frac{d'_{\text{in}} E_{\text{in},0} + d'_{\text{opp}} E_{\text{opp},0}}{d'_{\text{in}} S_{\text{in},0} + d'_{\text{opp}} S_{\text{opp},0} + \sigma} \text{ (presample)}$$

where d'_{in} and d'_{opp} are behavioral d' at the RF and opposite RF locations, respectively. The fit parameters are the same as in model without d'. In Fig. 6 (I and J), excitatory (or suppressive) drives across stimulus types were averaged at the RF location for each neuron, and then, the neurons were sorted according to the distance.

Statistical analysis

Unless otherwise specified, we used paired t test and multifactor ANOVA for comparing normally distributed datasets. Normality was checked using a Kruskal-Wallis test.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abc8812

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. M. R. Cohen, J. H. R. Maunsell, Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
2. R. J. Krauzlis, L. P. Lovejoy, A. Zénon, Superior colliculus and visual spatial attention. *Ann. Rev. Neurosci.* **36**, 165–182 (2013).
3. C. J. McAdams, J. H. R. Maunsell, Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
4. J. Moran, R. Desimone, Selective attention gates visual processing in the extrastriate cortex. *Front. Cogn. Neurosci.* **229**, 342–345 (1985).
5. J. K. Baruni, B. Lau, C. D. Salzman, Reward expectation differentially modulates attentional behavior and activity in visual area V4. *Nat. Neurosci.* **18**, 1656–1663 (2015).
6. S. Bouret, B. J. Richmond, Sensitivity of locus ceruleus neurons to reward value for goal-directed actions. *J. Neurosci.* **35**, 4005–4014 (2015).
7. T. Ikeda, O. Hikosaka, Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron* **39**, 693–700 (2003).
8. A. Ramakrishnan, Y. W. Byun, K. Rand, C. E. Pedersen, M. A. Lebedev, M. A. L. Nicolelis, Cortical neurons multiplex reward-related signals along with sensory and motor information. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E4841–E4850 (2017).
9. M. R. Roesch, C. R. Olson, Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* **304**, 307–310 (2004).
10. A. E. Rorie, J. Gao, J. L. McClelland, W. T. Newsome, Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLOS ONE* **5**, e9308 (2010).
11. L. Stănişor, C. van der Togt, C. M. Pennartz, P. R. Roelfsema, A unified selection signal for attention and reward in primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9136–9141 (2013).
12. J. H. R. Maunsell, Neuronal representations of cognitive state: Reward or attention? *Trends Cogn. Sci.* **8**, 261–265 (2004).
13. H. Spitzer, R. Desimone, J. Moran, Increased attention enhances both behavioral and neuronal performance. *Science* **240**, 338–340 (1988).
14. T. Z. Luo, J. H. R. Maunsell, Neuronal modulations in visual cortex are associated with only one of multiple components of attention. *Neuron* **86**, 1182–1188 (2015).
15. J. F. Mitchell, K. A. Sundberg, J. H. Reynolds, Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**, 879–888 (2009).
16. D. Kahneman, *Attention and Effort* (Prentice Hall, 1973), vol. 1063.
17. D. Kahneman, J. Beatty, Pupil diameter and load on memory. *Science* **154**, 1583–1585 (1966).
18. J. W. Bisley, M. E. Goldberg, Attention, intention, and priority in the parietal lobe. *Ann. Rev. Neurosci.* **33**, 1–21 (2010).
19. S. Treue, J. H. R. Maunsell, Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).
20. C. E. Boudreau, T. H. Williford, J. H. R. Maunsell, Effects of task difficulty and target likelihood in area V4 of macaque monkeys. *J. Neurophysiol.* **96**, 2377–2387 (2006).
21. J. Beatty, Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* **91**, 276–292 (1982).
22. B. Laeng, S. Sirois, G. Gredebäck, Pupillometry: A window to the preconscious? *Perspect. Psychol. Sci.* **7**, 18–27 (2012).
23. T. Piquado, D. Isaacowitz, A. Wingfield, Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* **47**, 560–569 (2010).
24. S. J. Luck, L. Chelazzi, S. A. Hillyard, R. Desimone, Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).
25. J. H. Reynolds, T. Pasternak, R. Desimone, Attention increases sensitivity of V4 neurons. *Neuron* **26**, 703–714 (2000).
26. T. Williford, J. H. R. Maunsell, Effects of spatial attention on contrast response functions in macaque area V4. *J. Neurophysiol.* **96**, 40–54 (2006).
27. D. Kobak, W. Brendel, C. Constantinidis, C. E. Feierstein, A. Kepecs, Z. F. Mainen, X. L. Qi, R. Romo, N. Uchida, C. K. Machens, Demixed principal component analysis of neural population data. *eLife* **5**, e10989 (2016).
28. S. Ghosh, J. H. R. Maunsell, Single trial neuronal activity dynamics of attentional intensity in monkey visual area V4. *Nat. Commun.* **12**, 1–15 (2021).
29. B.-E. Verhoef, J. H. R. Maunsell, Attention operates uniformly throughout the classical receptive field and the surround. *eLife* **5**, e17256 (2016).
30. P. DeWeerd, M. R. Peralta, R. Desimone, L. G. Ungerleider, Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. *Nat. Neurosci.* **2**, 753–758 (1999).
31. J. L. Gallant, R. E. Shoup, J. A. Mazer, A human extrastriate area functionally homologous to macaque V4. *Neuron* **27**, 227–235 (2000).
32. A. Zénon, R. J. Krauzlis, Attention deficits without cortical neuronal deficits. *Nature* **489**, 434–437 (2012).
33. K. R. Ridderinkhof, M. Ullsperger, E. A. Crone, S. Nieuwenhuis, The role of the medial frontal cortex in cognitive control. *Science* **306**, 443–447 (2004).
34. F. Kouneiher, S. Charron, E. Koechlin, Motivation and cognitive control in the human prefrontal cortex. *Nat. Neurosci.* **12**, 939–945 (2009).
35. E. Vassena, M. Silvetti, C. N. Boehler, E. Achten, W. Fias, T. Verguts, Overlapping neural systems represent cognitive effort and reward anticipation. *PLOS ONE* **9**, e91008 (2014).
36. M. Sarter, W. J. Gehring, R. Kozak, More attention must be paid: The neurobiology of attentional effort. *Brain Res. Rev.* **51**, 145–160 (2006).
37. M. I. Posner, S. E. Petersen, The attention system of the human brain. *Ann. Rev. Neurosci.* **13**, 25–42 (1990).
38. M.-M. Mesulam, Attentional networks, confusional states, and neglect syndromes. In M.-M. Mesulam (Ed.), *Principles of Behavioral and Cognitive Neurology* (Oxford University Press, 2000) pp. 174–256.
39. G. Aston-Jones, J. D. Cohen, An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Ann. Rev. Neurosci.* **28**, 403–450 (2005).
40. J. Krueger, A. A. Disney, Structure and function of dual-source cholinergic modulation in early vision. *J. Comp. Neurol.* **527**, 738–750 (2019).
41. L. Pinto, M. J. Goard, D. Estandian, M. Xu, A. C. Kwan, S. H. Lee, T. C. Harrison, G. Feng, Y. Dan, Fast modulation of visual perception by basal forebrain cholinergic neurons. *Nat. Neurosci.* **16**, 1857–1863 (2013).

42. L. Seillier, C. Lorenz, K. Kawaguchi, T. Ott, A. Nieder, P. Pourriahi, H. Nienborg, Serotonin decreases the gain of visual responses in awake macaque V1. *J. Neurosci.* **37**, 11390–11405 (2017).
43. A. Thiele, M. A. Bellgrove, Neuromodulation of attention. *Neuron* **97**, 769–785 (2018).
44. D. J. Heeger, Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992).
45. Y. Chen, S. Martinez-Conde, S. L. Macknik, Y. Bereshpolova, H. A. Swadlow, J. M. Alonso, Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nat. Neurosci.* **11**, 974–982 (2008).
46. J. P. Mayo, J. H. R. Maunsell, Graded neuronal modulations related to visual spatial attention. *J. Neurosci.* **36**, 5353–5361 (2016).
47. J. Lee, T. Williford, J. H. R. Maunsell, Spatial attention and the latency of neuronal responses in macaque area V4. *J. Neurosci.* **27**, 9632–9637 (2007).
48. M. M. Churchland, B. M. Yu, J. P. Cunningham, L. P. Sugrue, M. R. Cohen, G. S. Corrado, W. T. Newsome, A. M. Clark, P. Hosseini, B. B. Scott, D. C. Bradley, M. A. Smith, A. Kohn, J. A. Movshon, K. M. Armstrong, T. Moore, S. W. Chang, L. H. Snyder, S. G. Lisberger, N. J. Priebe, I. M. Finn, D. Ferster, S. I. Ryu, G. Santhanam, M. Sahani, K. V. Shenoy, Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nat. Neurosci.* **13**, 369–378 (2010).
49. M. N. Shadlen, W. T. Newsome, The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).

Acknowledgments: We thank J. J. Cone and C. Cherian for the helpful discussion and/or comments on the manuscript. **Funding:** This work was supported by NIH grant R01EY005911. The funder had no role in study design, data collection and interpretation, or the decision to submit the work for publication. **Author contributions:** S.G. and J.H.R.M. designed the experiments, performed the surgeries, and wrote the paper. S.G. performed the experiments and analyzed the data. **Competing interests:** The authors declare that they have no competing interests.. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 23 August 2021
Accepted 28 April 2022
Published 10 June 2022
10.1126/sciadv.abc8812