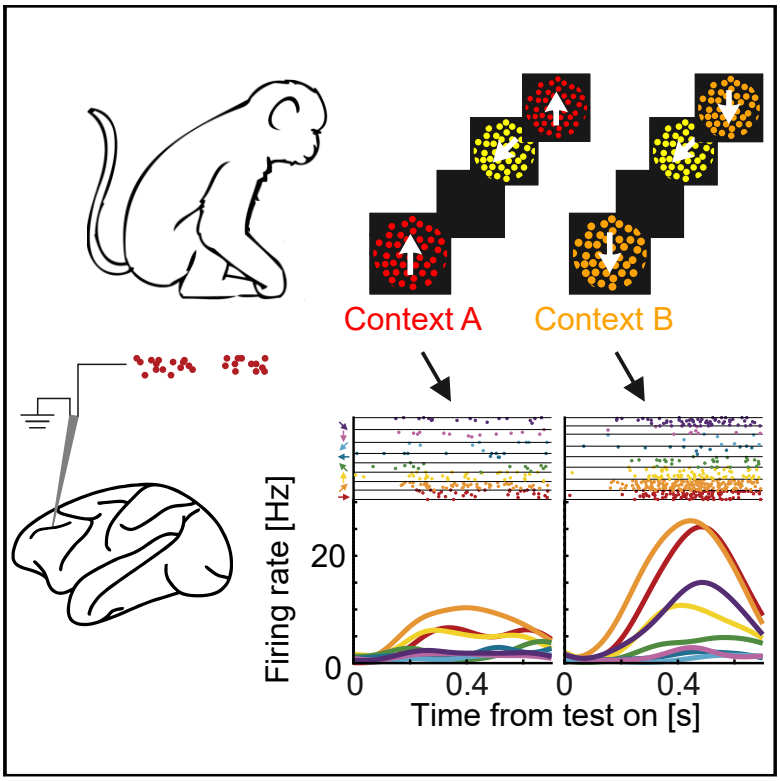
Cell Reports

Article

Feature-based attention processes in primate prefrontal cortex do not rely on feature similarity

Graphical abstract



Authors

Maximilian Stalter,
Stephanie Westendorff, Andreas Nieder

Correspondence

andreas.nieder@uni-tuebingen.de

In brief

Stalter et al. characterize the effects of feature-based attention on PFC neurons' tuning while monkeys switch attention based on two coupled features. Although PFC neurons experience strong gain modulation in response to attentional demands, neuronal gain modulation is not bound to the tuning preferences of neurons. This finding contradicts the feature-similarity gain model.

Highlights

- Prefrontal cortex neurons switch attention between conjunctive features
- Neurons experience strong gain modulation in response to attentional demands
- Attentional gain modulation is independent of neurons' feature tuning
- Feature similarity is not a general mechanism in feature-based attention

Stalter et al., 2021, Cell Reports *36*, 109470
August 3, 2021
https://doi.org/10.1016/j.celrep.2021.109470

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Feature-based attention processes in primate prefrontal cortex do not rely on feature similarity

Maximilian Stalter,[1,2] Stephanie Westendorff,[1,2] and Andreas Nieder[1,3,*]
[1]Animal Physiology Unit, Institute of Neurobiology, Eberhard-Karls-Universität Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany
[2]These authors contributed equally
[3]Lead contact
*Correspondence: andreas.nieder@uni-tuebingen.de
https://doi.org/10.1016/j.celrep.2021.109470

SUMMARY

Feature-based attention enables privileged processing of specific visual properties. During feature-based attention, neurons in visual cortices show "gain modulation" by enhancing neuronal responses to the features of attended stimuli due to top-down signals originating from prefrontal cortex (PFC). Attentional modulation in visual cortices requires "feature similarity:" neurons only increase their responses when the attended feature variable and the neurons' preferred feature coincide. However, whether gain modulation based on feature similarity is a general attentional mechanism is currently unknown. To address this issue, we record single-unit activity from PFC of macaques trained to switch attention between two conjunctive feature parameters. We find that PFC neurons experience gain modulation in response to attentional demands. However, this attentional gain modulation in PFC is independent of the feature-tuning preferences of neurons. These findings suggest that feature similarity is not a general mechanism in feature-based attention throughout the cortical processing hierarchy.

INTRODUCTION

Processing of sensory stimuli is strongly modulated by a subject's behavioral state. An important determinant of behavioral state is top-down selective attention, a powerful mechanism for privileged processing of behaviorally relevant stimuli at the expense of irrelevant information. Visual selective attention is typically classified into spatial and feature-based attention. Spatial attention, which has been investigated most intensively, prioritizes the processing of sensory input from a restricted part of the visual scene. In contrast, feature-based attention focuses on the selection of attended visual features (such as upward motion or the color red) irrespective of stimulus location (Liu, 2019; Maunsell and Treue, 2006).

Selective attention generally enhances (or gains) the responses of sensory neurons to stimuli/features in the focus of attention and diminishes responses to distractors, an effect termed "gain modulation" (Buschman and Kastner, 2015; Desimone and Duncan, 1995; Maunsell and Treue, 2006; Squire et al., 2013). However, the origins of these neuronal correlates of visual selective attention in sensory areas have been implicated to lie in parietal (Saalmann et al., 2007) and especially frontal cortical regions that exert top-down modulatory influence (Buschman and Miller, 2007; Everling et al., 2002; Rainer et al., 1998a). While the frontal eye field (FEF) has a key role in spatial attention (Gregoriou et al., 2014; Moore and Armstrong, 2003; Moore and Fallah, 2001), parts of the dorsolateral prefrontal cortex (PFC) constitute the causal basis for feature-based attention (Baldauf and Desimone, 2014; Bichot et al., 2015, 2019). In agreement with these differences in origin, physiological comparisons of feature-based and spatial attention suggest that these two forms of attention operate through different feedback networks (Bichot et al., 2015, 2019; Hayden and Gallant, 2005; Maunsell and Treue, 2006; McAdams and Maunsell, 2000). However, the nature of feature-based attentional top-down signals originating in PFC remains elusive.

Currently, it is tacitly assumed that feature-selective PFC neurons show basically the same attentional effects as have been described in visual cortical areas that receive top-down signals. In an influential electrophysiological experiment, the responses of middle temporal (MT) neurons were enhanced (i.e., gained) when the attended and the neurons' preferred motion directions coincided but suppressed when they differed (Treue and Martínez Trujillo, 1999). Similarly, response in V4 to a preferred stimulus in the receptive field (RF) was enhanced if this stimulus matched a searched-for stimulus, whereas response to a non-preferred stimulus showed no such enhancement. In both cases, spatial attention was directed somewhere else, such that the described effects can only be attributed to feature-based attention (Bichot, 2005; David et al., 2008). Single-unit findings like these were captured in the "feature-similarity gain model" according to which neuronal activity enhancement is a monotonic function of the similarity between the attended feature parameter and the neuron's preferred parameter of this feature. In other

words, neurons only increase their gain due to top-down signals when attention is directed to their preferred feature parameter (or location). In addition to a pure gain change, feature-based attention has been shown to alter the tuning preference of parietal neurons; tuning often shifts to more closely match the attended feature (David et al., 2008; Ibos and Freedman, 2014), an effect likely mediated by top-down feedback signals (Compte and Wang, 2006; Womelsdorf et al., 2008). Whether these attentional effects and the feature-similarity gain model established for early visual cortex extend to higher-order cortical areas such as the PFC as the source of top-down attentional signals has never been tested experimentally.

Here, we explore the coding properties of neurons of PFC, which operates at the apex of the cortical hierarchy and contributes key feedback signals during feature-based attention in behaving rhesus macaques. The monkeys were trained on a delayed conjunction matching task (DCMT) previously devised to investigate behavioral state effects in parietal lobe (Ibos and Freedman, 2014). While the monkeys were instructed to switch their attention between two combinations of a stimulus' motion direction and color, we explored the neurons' tuning properties in relation to the attended stimulus features and parameters.

RESULTS

Monkeys flexibly used contextual cues

We investigated the influence of feature-based attention on single neurons in the lateral PFC. To that aim, we trained two monkeys on a DCMT (Figure 1A) equivalent to the one used previously by Ibos and Freedman (2014). To identify target stimuli, the monkeys had to use two task-relevant visual features: the direction of motion and the color of the presented visual stimuli (Figure 1B). The relevant context was set by one of two sample stimuli that cued the monkeys on the relevant features, i.e., red dots moving upward (sample A) or yellow dots moving downward (sample B). After presentation of the sample stimulus and a brief delay, a succession of up to three test stimuli followed. The monkeys had to release a lever if a test stimulus matched the current sample in both features, i.e., color and motion direction. In both contexts, the monkeys saw the identical set of test stimuli. Because the sample stimulus identity varied on a trial-by-trial basis, the relevant motion direction and color varied too. This allowed us to assess how flexible attention to the two different stimulus features changed the responses of single neurons in the PFC depending on the behavioral context, i.e., when the monkeys searched for a certain motion direction and color. Trials in which a sample A or sample B stimulus was shown were defined as context A and context B trials, respectively. Since the sensory appearance of the different test stimuli remained constant between the contexts, the only aspect that changed was the monkeys' attention to the relevant features of the sample stimulus.

Both monkeys performed the task well above chance (monkey 1: 95%, n = 75 sessions with an average of 576 ± 12 trials [SEM]; monkey 2: 97%, n = 48 sessions with an average of 403 ± 10 trials [SEM]), indicating that they used both visual features for solving the task (Figure 1C). Overall, performance was close to perfect for both monkeys and stimulus contexts. Monkey 1 mildly preferred context B (96%) to context A trials (95%; $p < 0.001$; Wilcoxon signed-rank test; n = 75), and monkey 2 had the opposite preference (98% versus 97% on context A trials versus B trials, respectively; $p = 0.002$ Wilcoxon signed-rank test; n = 48). This mild sample preference was accompanied by a corresponding small difference in false alarm rates (FARs) for the two samples (ΔFAR (A-B) = 1%, $p < 0.001$ and ΔFAR = −0.5%, $p = 0.013$ for monkey 1 and 2, n = 75 and n = 48 respectively; Wilcoxon signed-rank test).

We measured the FAR to each of the possible test stimuli for both monkeys. The overall error pattern is as expected (Figures 1D and 1E). We compared the FAR at each stimulus to the average FAR across all stimuli. Asterisks mark the FAR to a stimulus significantly different from the average FAR across all stimuli (test for difference between two proportions; Zar, 2010; corrected for multiple comparisons, $p < 0.05$). Red asterisks mark FARs smaller than the average, and cyan asterisks mark FARs larger than the average (Figures 1D and 1E). The animals made nearly no false alarm to stimuli that were in their feature space very far away from either sample stimulus but made more errors to stimuli that were similar to the cued target stimulus and/or to the alternative sample stimulus. These results show that the animals learned the basic task structure. FARs were low for all stimulus combinations and monkeys. Importantly, this was also true for the test stimulus that matched the alternative and currently nonvalid sample stimulus, i.e., the yellow, downward stimulus in context A and the red, upward stimulus in context B (monkey 1: FAR = 0.16 in context A and 0.18 in context B; monkey 2: FAR = 0.02 in context A and 0.05 in context B; all $p < 0.01$, binominal test). This confirms that the monkeys were under stimulus control and were paying attention to the specific sample parameters; they correctly ignored test stimulus B during sample A trials and vice versa.

PFC neurons represent visual features in the test period

We recorded single-unit activity of 489 unbiasedly selected units (n = 208 and n = 281 for monkey 1 and 2, respectively) in the principal sulcus region of the lateral PFC while the monkeys performed the DCMT (Figures 1D and 1E). A large proportion of neurons (43% [208/489]) were selective to motion direction, whereas 24% (115/489) were selective for color. These neurons were counted as feature selective. The population of motion-direction-selective neurons and the population of color-selective neurons overlapped by 62 neurons (30% [62/208] and 54% [62/115], respectively).

Neuronal activity of both motion-selective and color-selective cell populations were modulated during the test period as a function of task context (i.e., whether the features of sample A or B were relevant). Figure 2 shows the responses of two example neurons, a motion-direction-selective neuron (#1) and a color-selective neuron (#2), during the test period for the two different task contexts. Context-dependent responses of the motion-direction selective neuron (#1) are displayed in Figures 2A and 2B. Neuron #1 displayed higher firing rates to different motion directions (i.e., gain modulation) in the context of sample B compared to sample A. A similar context-dependent effect was observed for the color-selective neuron (Figures 2C and 2D). Example neuron #2 displayed higher firing rates for colors

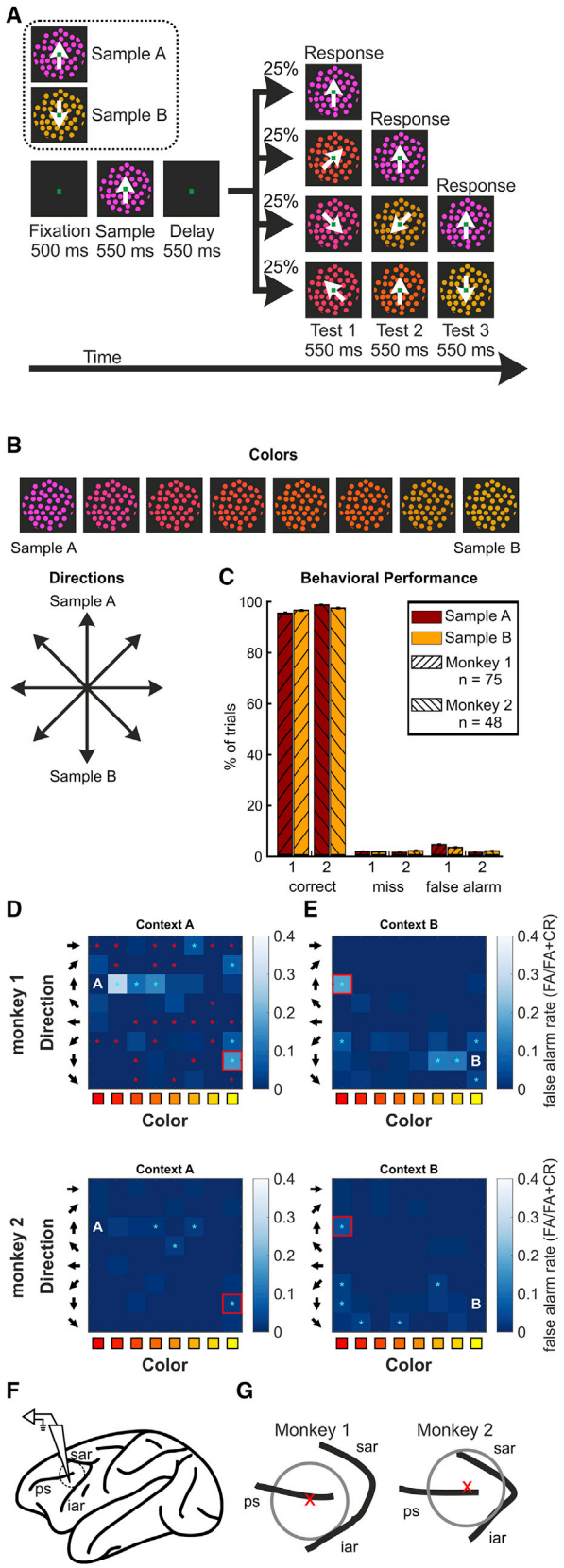


Figure 1. Task protocol, behavioral performance, and recording sites

(A) Delayed conjunction matching task (DCMT). A trial started when the monkey grabbed a bar and fixated on a fixation target in the center of the computer screen. After a fixation period, one of two sample stimuli (colored and coherently moving random-dot displays) were shown that instructed the monkey to which combination of features it had to attend throughout the trial. Sample A cued for upward motion of red dots (0° angular motion direction); sample B instructed downward motion of yellow dots (270°). After a brief memory delay, the monkey had to compare the features of the sample stimulus to the features of to three sequentially presented test stimuli. In equal proportions and pseudo-randomized presentation, either the first, second, third or none of the test stimuli showed both of the matching features. The monkey had to maintain hold of the bar if one or both features did not match and release the bar to indicate a match of visual features to receive a reward. If none of the test stimuli was a match, then the monkey was required to maintain hold of the bar for an additional 500 ms to be rewarded.

(B) Set of visual features. Top panel: test stimuli consisted of random-dot displays of one of eight specific color hues ranging from red (color of sample A) to yellow (color of sample B). Bottom panel: dots of the random-dot displays moved coherently in one of eight different directions, each 45° apart. For the test stimuli, each motion direction was paired with each color hue, resulting in a set of 64 conjunct stimuli.

(C) Behavioral performance. The proportions of correct, miss, and false alarm trials are shown for both monkeys (separately for sample A and B trials). Error bars represent SEM across sessions.

(D) Confusion matrix depicting the false alarm rates (FARs) for all 64 test stimulus combinations (8 directions and 8 colors) in context A. Top panel: monkey 1. Bottom panel: monkey 2. The letter A marks the test stimulus that matches the current sample stimulus A. The red square marks the test stimulus that matches the currently non-matching sample stimulus B. Asterisks mark FARs to a stimulus significantly different from the average FAR across all stimuli. Red asterisks mark FARs smaller than the average, and cyan asterisks mark FARs larger than the average.

(E) Same as in (D), but for context B.

(F) Recording location. The location of single-unit recording within PFC is shown for a schematic drawing of a rhesus macaque brain.

(G) Anatomical surface reconstruction of the recording site for monkey 1 (left) and monkey 2 (right). Recordings were performed dorsal and ventral to the posterior end of the principle sulcus. The red "X" indicates the center of the recording chamber. Abbreviations: PS, principle sulcus; SAR, superior sulcus arcuatus; IAR, inferior sulcus arcuatus.

in the context of sample A compared to sample B. In fact, the neuron was virtually inactive when colors were presented in the context of sample B.

We determined the preference of the directional motion tuning and color tuning during the DCMT for all feature-selective neurons. Across the populations of neurons, the preferred motion directions were evenly distributed. This was true for context A (Figure 3A; Rayleigh test for circular uniformity, $p = 0.47$, $n = 122/208$) and context B (Figure 3B; Rayleigh test, $p = 0.09$, $n = 146/208$). (Note that neurons that were feature selective in both contexts were included in the analysis of context A and B.) Similarly, the preferred color values were evenly distributed across color-selective neurons in context B (Figure 3D; chi-squared test, $p = 0.24$, $n = 64/115$) and context A (Figure 3C), albeit with one intermediate color underrepresented (chi-squared test, $p = 0.04$, $n = 60/115$). Also, there was no change in number of neurons selective for target motion/color in each context (percentage of neurons in context A versus context B, test for difference between two proportions; Zar, 2010; upward tuned, $p = 0.19$, downward tuned, $p = 0.55$; red tuned, $p = 0.24$; yellow tuned, $p = 0.03$). Note that

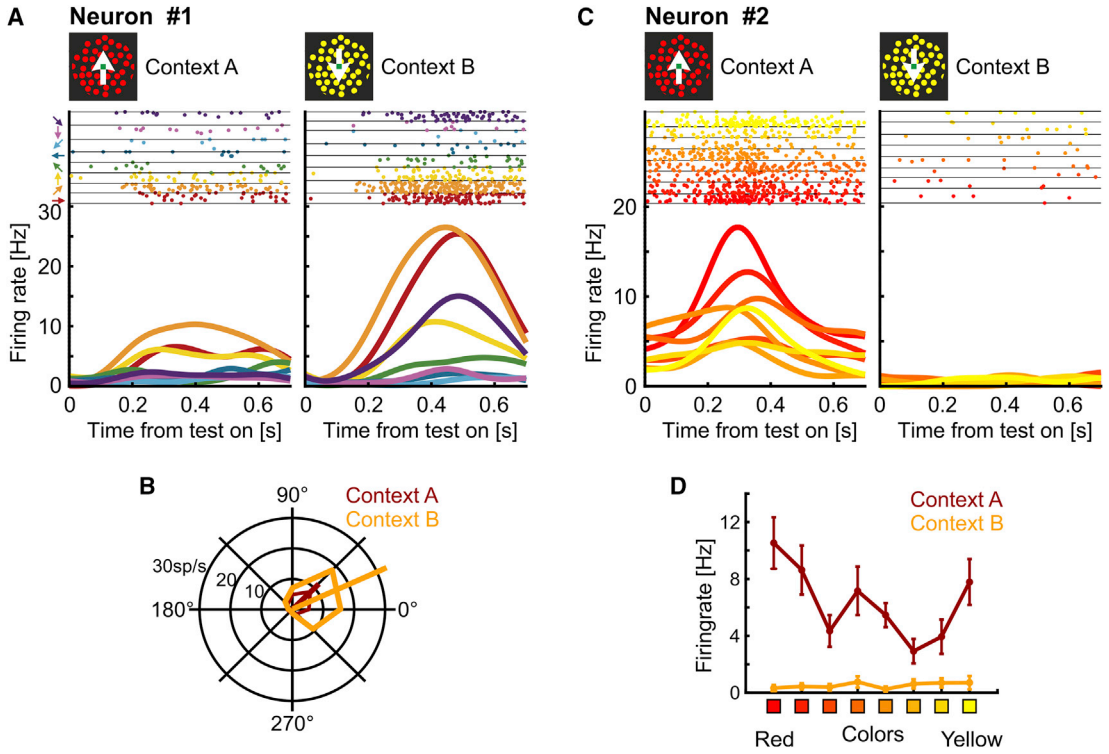


Figure 2. Feature-selective neurons experience context-dependent gain modulation during the test period

(A) Time-resolved responses of a motion-direction selective neuron (# 1). Left panel: responses during context A (attend to upward motion of red dots cued). Right panel: responses during context B (attend to downward motion of yellow dots). The top panels show dot-raster histograms depicting neuronal activity color-coded with respect to the eight feature levels (i.e., motion directions). Each line is a trial, and each dot represents an action potential. The bottom shows the corresponding, smoothed spike-density histograms for the same neuronal discharges. Note that the visual input during the test period is the same for both contexts.

(B) Polar plot showing the directional tuning of the direction-selective neuron #1 to the eight motion directions during context A and context B.

(C) Time-resolved responses of a color-direction-selective neuron (# 2). Same layout as in (A).

(D) Color tuning of color-selective neuron #2 to the eight color hues during context A and context B. Error bars represent SEM across sessions.

even though the difference for yellow-tuned neurons reaches significance, the numbers are opposite from what would be expected with more neurons tuned to yellow in context A (14 versus 6 in context B), in which the red, upward-moving sample stimulus was shown.

For a fraction of individual neurons (n = 418), we also compared the motion direction tuning while the monkeys alternatingly performed the DCMT while passively viewing moving random gray-dot patterns (passive task) that moved in eight different directions. Of the 73 neurons that were selectively tuned during the passive task (one-factorial ANOVA; $p < 0.05$), 45 neurons were also selectively tuned during the DCMT. Figures 3E–3H shows the responses of four exemplary motion-direction-selective neurons that were recorded in both tasks. The tuning profiles during the passive task matched the neurons' profiles during the attention-driven DCMT in both contexts. To quantify the similarity between the tuning in the passive task and the DCMT at the level of the neuron population, we calculated the angular differences between the preferred direction in the passive task and the preferred direction in the context the neuron was motion direction-selective in during the DCMT (Figure 3I). If the tuning profiles in both tasks were similar, then we would expect small angular differences that lie around 0° difference. Indeed, on average the angular differences, as captured by the population direction vector, were distributed around 0° (v-test, $p < 0.0001$, n = 71). This shows that motion-direction tuning of PFC neurons during the DCMT (with attention) was comparable to the neurons' spontaneous tuning in the absence of attentional selection.

Tuning sensitivity as a measure of gain modulation

Attentional gain modulation has been observed frequently in visual cortex (Lee and Maunsell, 2010; McAdams and Maunsell, 1999; Treue and Martínez Trujillo, 1999). We therefore explored gain modulation by quantifying the difference in tuning sensitivity of individual neurons between the two contexts.

The tuning sensitivities in context A and context B of all motion-direction-selective neurons are plotted in Figure 4A. Of these 208 neurons, 15% (32/208) showed significantly higher tuning sensitivity to one of the two contexts. During context A, 10 of 32 neurons exhibited higher tuning sensitivity, whereas in context B, 22 of 32 neurons showed significantly higher tuning sensitivity ($p < 0.05$, bootstrap analysis). An equivalent effect was seen for color-selective neurons (Figure 4C). Here, 17% of

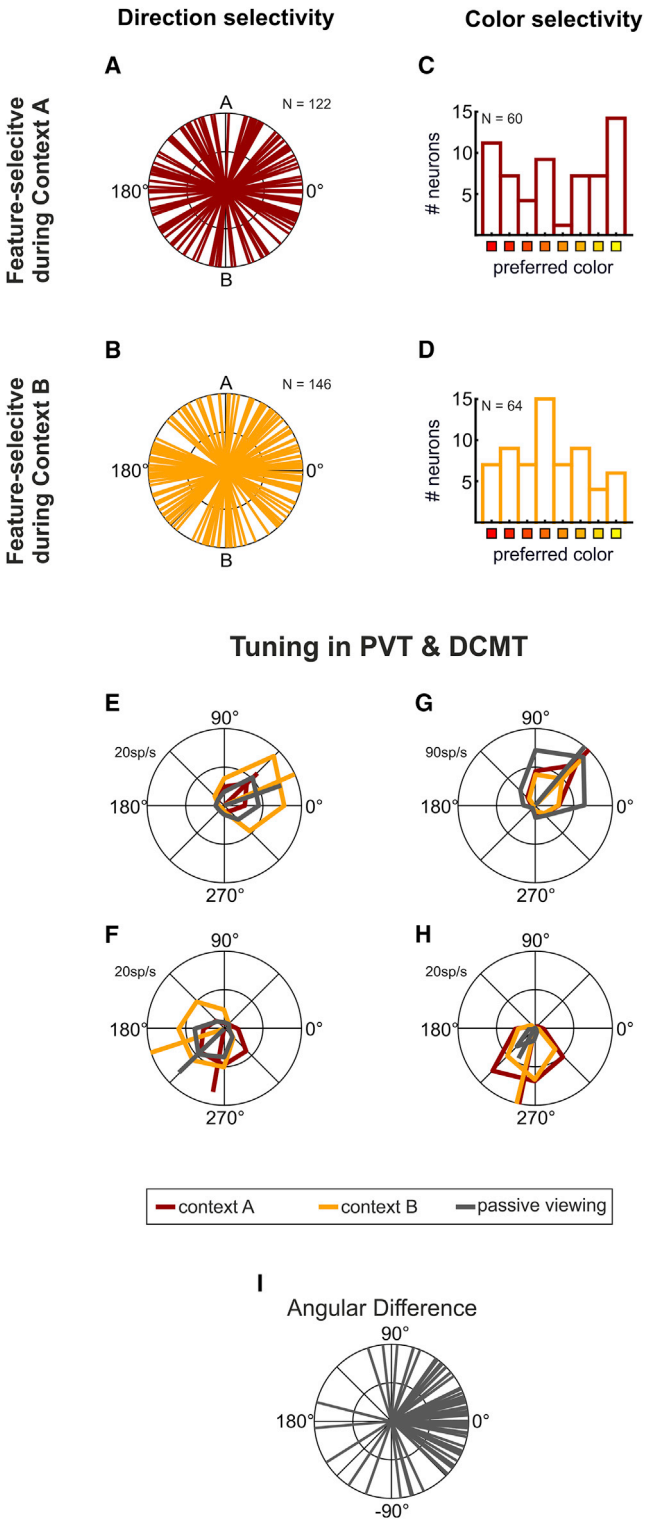


Figure 3. Neuronal tuning of feature-selective neurons to motion direction or color

(A) Distribution of preferred directions during context A for neurons that were selective for the direction of motion during context A or both contexts.

(B) Distribution of preferred directions during context B for neurons that were selective for the direction of motion during context B or both contexts.

neurons (19/115) showed significantly higher tuning sensitivity in one of the two contexts. During context A, 8 of 19 neurons had a higher tuning sensitivity, whereas 11 of 19 neurons exhibited higher tuning sensitivity in context B ($p < 0.05$, bootstrap analysis).

Not only single neurons but also the population as a whole modulated tuning sensitivity in response to the monkeys' attending to context A or B. For direction-selective neurons, the distribution of area under the receiver operating characteristic curve (AUROC) differences (context A minus context B) was significantly broader than expected by chance (Figure 4B), indicating an increase in selectivity at both sides of the distribution. The distribution of AUROC differences was different from a normal distribution ($p = 5.3 \times 10^{-28}$, single sample Kolmogorov-Smirnov goodness-of-fit test) and instead bimodal ($p = 0.033$, Hartigans dip test). Therefore, to test for significance, we used the interquartile range (IQR) as a measure for the width of the distribution and compared this to the IQRs of shuffled data (see STAR Methods). None of the IQRs of the shuffled distribution were larger than the IQR of the real data (randomization test, $p = 0$; real IQR, 1.75; shuffled distribution IQR minimum, 0.09; maximum, 0.148). The real IQR was 6.52 standard deviations away from the mean of the shuffled distribution, and thus, the distribution of AUROC differences was significantly wider than expected under the null hypothesis. Both effects (bimodality and the broadening of the distribution) support the conclusion that the context modulated the tuning sensitivity. The results were qualitatively similar if we used the tuning vector length instead of the ROC value as measure for tuning sensitivity (randomization test, $p = 0.02$; real IQR, 0.63; shuffled distribution IQR minimum, 0.403; maximum, 0.667). A similar effect was seen for the population of color-selective neurons, for which the distribution of AUROC differences (context A minus context B) was also bimodal ($p = 0.026$) and significantly broader than expected by chance (Figure 4D) (randomization test, $p = 0$; real IQR, 1.93; shuffled distribution IQR minimum, 0.088; maximum, 0.172). Here, the IQR of the real data was 5.34 standard deviations away from the mean of the shuffled distribution. This difference in tuning sensitivity indicates a gain modulation influenced by attentional demands for both direction- and color-selective neurons. When we repeated the analysis with a criterion of minimum seven trials per conditions, all the results confirm our original conclusions: the distribution of AUROC differences was still significantly broader than expected under the null hypothesis (direction: $p = 0$, $n = 173$; color: $p = 0$, $n = 88$, randomization test). To test if the effect is only driven by few selective neurons, we repeated the analysis without the

(C) Distribution of preferred colors during context A for neurons that were color selective during context A or both contexts.

(D) Distribution of preferred colors during context B for neurons that were color-selective during context B or both contexts.

(E–H) Directional tuning of four exemplary motion direction-selective neurons that were recorded during both the passive-viewing task (PVT) and the DCMT. The tuning profiles during the passive task (gray) match the tuning during the attention-driven DCMT (red and yellow).

(I) Angular differences for all direction-selective neurons that were recorded during both passive viewing and DCMT. The differences cluster around 0°, indicating similar preferred directions in both tasks.

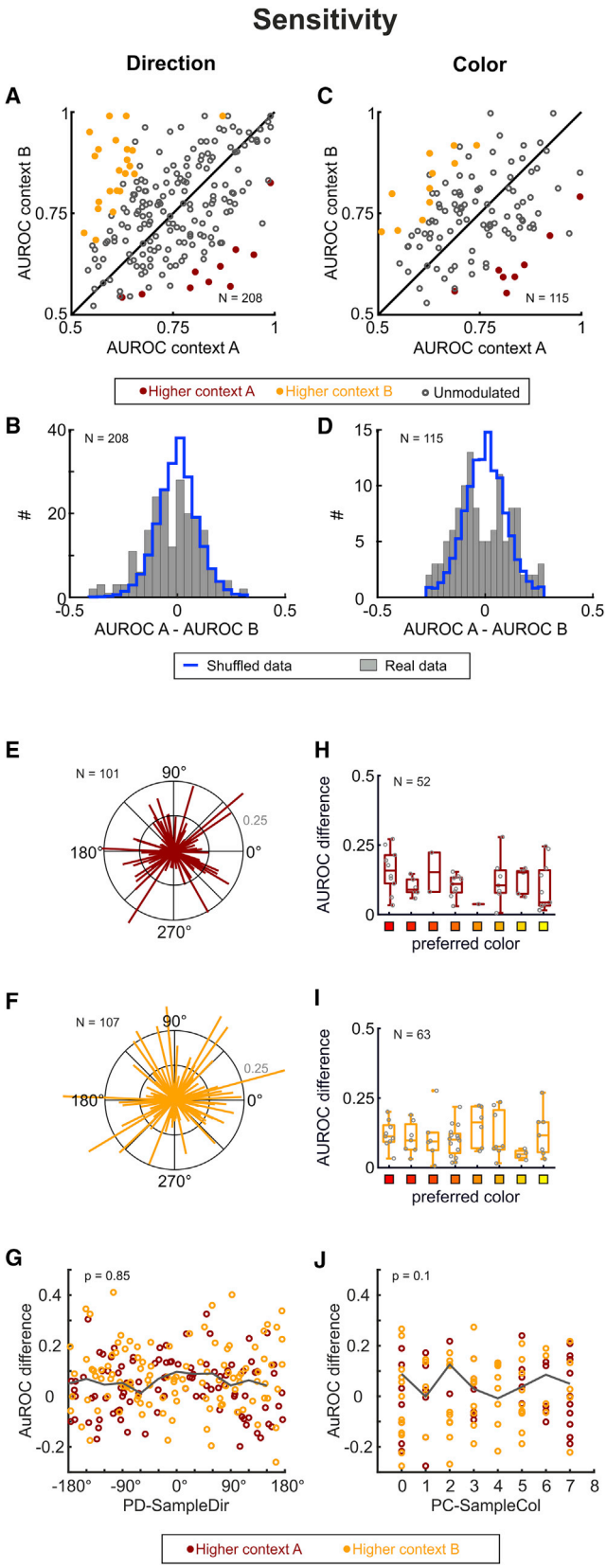


Figure 4. Attentional modulation of tuning sensitivity

(A) Comparison of the direction-selective neurons' tuning sensitivity (AUROC values) in the context of sample A versus the context of sample B. Each dot represents one neuron. Individual neurons with significantly higher AUROC values in one context over the other are color coded (red dots for higher sensitivity in context A; yellow dots for higher sensitivity in context B). Gray, open circles indicated neurons that showed no significant difference in amplitudes between the two contexts.

(B) Histogram showing the significantly broader distribution of the neurons' real AUROC differences compared to the distribution of differences derived from shuffled AUROC values (blue line).

(C) Comparison of the color-selective neurons' tuning sensitivity (AUROC values) in the context of sample A versus the context of sample B. Same layout as in (A).

(D) Histogram showing the significantly broader distribution of the color-selective neurons' real AUROC differences compared to the modeled distribution of differences derived from shuffled AUROC values (blue line). Same layout as in (B).

(E) Distribution of AUROC differences for all direction-selective neurons with higher sensitivity during context A. Each line represents a neuron's AUROC difference at its preferred direction.

(F) Distribution of AUROC differences for all direction-selective neurons with higher sensitivity during context B. Same layout as in (E).

(G) Comparison of AUROC differences for direction-selective neurons. The neurons are sorted by the distance of their preferred direction to the direction of the sample stimulus. For neurons with a higher firing rate in context A (red circles) the distance to sample A was taken, and for neurons with higher firing rate in context B (yellow circles) the distance to sample B was taken. Grey line represents average AUROC difference in 30°-bins.

(H) Distribution of AUROC differences for all color-selective neurons with higher sensitivity during context A. Boxplots show average AUROC differences for neurons tuned to each of the eight presented color hues. Whiskers extend to the most extreme data points not considered outliers.

(I) Distribution of AUROC differences for all color-selective neurons with higher sensitivity during context B. Same layout as in (H), Whiskers extend to the most extreme data points not considered outliers.

(J) Comparison of AUROC differences for color-selective neurons. Neurons are sorted by the distance of their preferred color to the color of the sample stimulus. Same layout as in (G).

neurons that show a significant effect on the single-cell level (without the red and yellow marked neurons in Figure 4A). The distributions of AUROC differences for direction and color were still significantly wider than the respective distributions under the null hypothesis (direction: $p = 0.006$; color: $p = 0.012$). Additionally, if we trained a support vector machine (SVM) to decode the feature (direction or color) of a test stimulus and the context, it was well able to do so (Figure S1). Given the fact that the visual input was the same, the information the SVM used to differentiate the context had to come from the gain modulation.

So far, changes of tuning sensitivity were explored irrespective of the neurons' motion-direction or color preferences. However, according to results found in visual cortices, one might expect that changes in tuning sensitivity, i.e., gain effects, are strongest for neurons tuned to the currently attended feature. In a next step, we therefore explored tuning sensitivity changes in relation to the neurons' preferred tuning. This was done separately for neurons with higher sensitivity during context A and context B.

Figure 4E depicts the relation between direction tuning and tuning sensitivity of individual neurons that showed higher tuning sensitivities during context A (i.e., all neurons contributing to data points below the unity line in Figure 4A). During context A, the monkey is attending to the upward motion direction. Based

on findings in visual cortex, one would expect that the majority of neurons with higher tuning sensitivity in context A show an upward tuning preference. The direction of the red lines in the polar plot indicate the preferred motion direction, while the length of each red line represents the magnitude of the difference in tuning sensitivity (AUROC difference) between both contexts. In contrast to predictions from visual cortex, neurons with higher sensitivity during context A did not show a preference for upward motion. Instead, they had preferred tuning directions distributed across the whole circle (v-test for 90°, $p = 0.07$, $n = 101$). Qualitatively similar results were obtained only for neurons showing significantly higher AUROC values in context A (neurons corresponding to red data points in Figure 4A)

The same analysis was applied to individual neurons that showed higher tuning sensitivities during context B (i.e., all neurons contributing to data points above the unity line in Figure 4A). For these neurons, one might expect that the preferred tuning directions should cluster around the downward direction, as the monkey was attending to downward motion during context B. However, as shown in Figure 4F tuning preferences were uniformly distributed around the circle (v-test for 270°, $p = 0.82$, $n = 107$). Qualitatively similar results were obtained only for neurons showing significantly higher AUROC values in context B (neurons corresponding to yellow data points in Figure 4A).

According to the feature-similarity gain model, the gain is expected to change as a function of tuning. Neurons with a gain effect are expected to show stronger gain modulation when the attended feature matches their preferred tuning. Thus, for neurons gain-modulated in context A, stronger gain modulation is expected if their preferred tuning direction is upward compared to downward. Vice versa, stronger gain modulation is expected for neurons gain-modulated in context B if their preferred tuning direction is downward.

To test this hypothesis, we compared the strength of gain modulation in dependence of the distance of the sample to the preferred direction (Figure 4G). If the effects followed the feature-similarity gain model, then we could expect the strongest effect at 0° distance, with effect sizes falling off with increasing distance. In contrast, we found no difference across the distances (Kruskal-Wallis test, $p = 0.85$).

We computed ε^2 as measure of effect size (Tomczak and Tomczak, 2014). ε^2 was 0.03, which is considered small. These results were confirmed when we repeated the analysis with a minimum trial criterion of seven trials per condition. Although the distribution of tuning vectors for cells with gain changes (equivalent to Figures 4E–4I) did show a weak significance for a preferred direction for cells modulated in context A (v-test for 90°, $p = 0.043$, $n = 82$), there was no significance for neurons modulated in context B (v-test for 270°, $p = 0.1$, $n = 49$). Also, there was no significant difference in the size of gain modulation with distance from the preferred direction (Kruskal-Wallis test, $p = 0.17$ $\varepsilon^2 = 0.089$). Therefore, surprisingly, our results do not support a relation between the neurons' tuning sensitivity as a function of attention to context A (upward) or B (downward) and the neurons' direction tuning preferences. We find no evidence that gain modulation during the monkeys' attention to upward and downward motion directions was dependent on the neurons' upward-versus-downward tuning preferences.

We also explored tuning sensitivity changes (i.e., gain modulation) in relation to the neurons' preferred tuning to colors. This resulted in very similar findings. Figure 4H shows the relation between color tuning and tuning sensitivity modulation of individual neurons that showed higher tuning sensitivities during context A (i.e., all neurons contributing to data points below the unity line in Figure 4C). Neurons with higher sensitivity during context A did not show a preference for red but had preferred tuning toward all colors (chi-squared test, $p = 0.05$, $n = 52$). Qualitatively similar results were obtained only for neurons showing significantly higher AUROC values in context A (neurons corresponding to red data points in Figure 4C). Similarly, neurons with higher sensitivity during context B (Figure 4I) did not show a preference for yellow but had preferred tuning toward all presented colors (chi-squared test, $p = 0.12$, $n = 63$). Similar indifferent results were obtained for the strength of gain modulation. The strength of gain modulation did not vary for different distances of the sample color to the preferred color of the neurons ($p = 0.1$, $\varepsilon^2 = 0.11$, Figure 4J). Again, these results were confirmed when repeating the analysis with a minimum of seven trials per condition. There was no preference for tuning to the color red for cells modulated in context A (chi-squared test, $p = 0.42$, $n = 39$), nor was there a preference for tuning to the color yellow for cells modulated in context B ($p = 0.1$, $n = 49$). In addition, the mean AUROC difference was not modulated by the distance from the preferred color (Kruskal-Wallis test, $p = 0.51$, $\varepsilon^2 = 0.03$).

Thus, based on tuning-sensitivity differences (derived from ROC analysis), we found no evidence that either the preferred motion direction or the preferred color of the neurons had any impact on attentional gain modulation.

Amplitudes of Gauss fits to direction tuning functions as a measure of gain modulation

In order to determine the magnitude of attentional gain modulation, Gauss functions are typically fitted to the tuning function of neurons in the visual system (McAdams and Maunsell, 1999; Treue and Martínez Trujillo, 1999). The amplitude of the Gaussian fits is then taken as quantitative parameter describing the response modulation of the tuning functions. We therefore explored amplitude differences of Gaussian functions fitted to the PFC neurons' tuning curves as a second parameter to explore attentional gain modulation. For this analysis, only tuning to motion direction was investigated, because color-tuning functions were severely truncated in color space.

The Gaussian fits described the tuning functions with decent quality (Figure 5A). As for the previous analysis, we included neurons that were direction selective in at least one context (see STAR Methods, direction and color-selective neurons). The black line in Figure 5A shows the average Gaussian fit to tuning functions in the respective selective context (mean r^2 goodness of fit = 0.63 ± 0.01 SEM; $n = 208$). The gray line shows the average Gaussian fit to the tuning function in the respective nonselective context (mean r^2 goodness of fit = 0.5 ± 0.02 SEM; $n = 208$). If neurons were direction selective in both contexts, the context with the higher average firing rate was defined as the selective context, and the context with the lower average firing rate was defined as the nonselective context

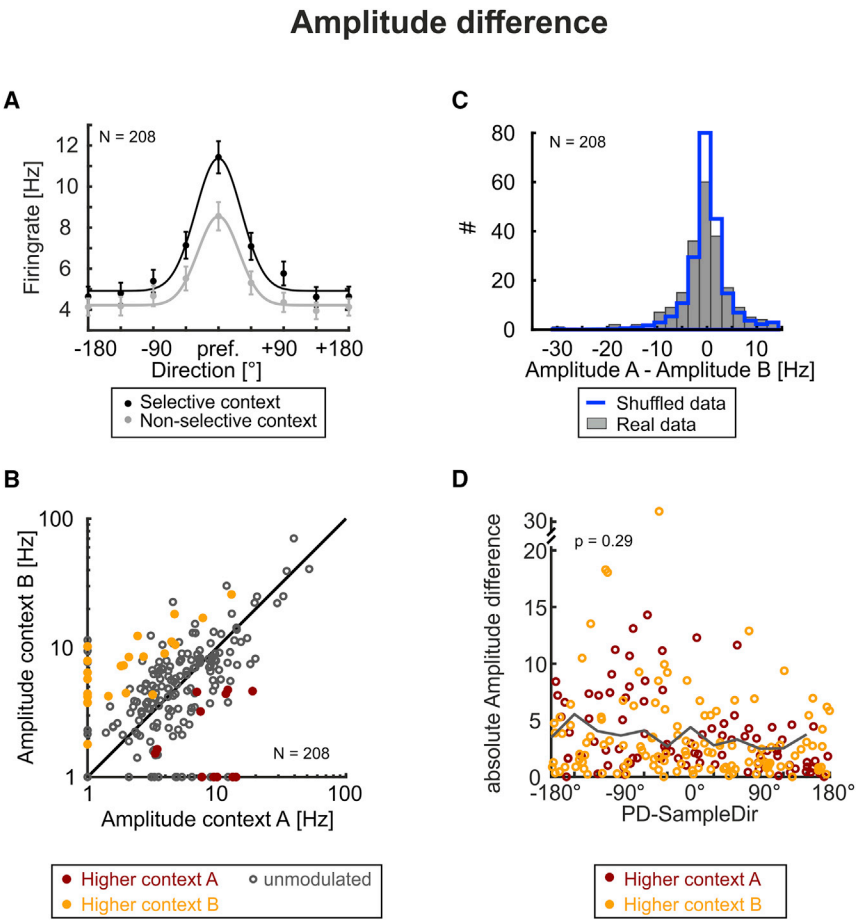


Figure 5. Attentional modulation of tuning amplitudes derived from Gauss fits

(A) Average Gauss fit to the tuning curves of all motion-direction-selective neurons. The fits to tuning functions in the selective context (black line) and the complementary nonselective context (gray line) are shown. Black and gray data points represent the average firing rates in the selective and complementary nonselective contexts; error bars represent SEM.

(B) Comparison of the direction-selective neuron (n = 208) amplitudes in the context of sample A against the context of sample B. Each dot represents one neuron. Individual neurons with significantly higher amplitude in one context over the other are color coded (red dots for higher sensitivity in context A; yellow dots for higher sensitivity in context B). Gray, open circles indicate neurons that showed no significant difference in amplitudes between the two contexts.

(C) Histogram showing the significantly broader distribution of the neurons' real firing rate amplitude differences compared to the modeled distribution of differences derived from shuffled amplitude rates (blue line).

(D) Comparison of amplitude differences for direction-selective neurons. The neurons are sorted by the distance of their preferred direction to the direction of the sample stimulus. For neurons with a higher firing rate in context A (red circles), the distance to sample A was taken, and for neurons with higher firing rate in context B (yellow circles), the distance to sample B was taken. Grey line represents absolute amplitude differences in 30°-bins.

Figure 5B shows a comparison of the neurons' amplitudes derived from the Gaussian fits during context A and B. Some of the neurons showed significantly higher amplitudes in one context over the other. These neurons are color coded in Figure 5B (red: higher amplitude in context A, n = 12; yellow: higher amplitude in context B, n = 23; bootstrap analysis, $p < 0.05$)

Not only single neurons but also the population as a whole showed significant modulation of tuning amplitudes in response to the monkeys' attending to context A or B. The distribution of the amplitude differences between context A and context B was significantly broader than expected under the null hypothesis based on shuffled amplitude differences (Figure 5C). None of the IQRs of a shuffled distribution was larger than the IQR of the real data (randomization test, p = 0; real IQR, 4.65; randomization distribution IQR minimum, 2.17; maximum, 3.94). The IQR of the real data was 5.35 standard deviations away from the mean of the randomization distribution. If the analysis was repeated with a more stringent criterion (r-square value > 0.6 in at least one context as goodness-of-fit measure for the Gaussian fits), then the distribution of amplitude differences between context A and context B was still significantly broader than expected under the null hypothesis (p = 0.00014, F-stat = 1.48, n = 161, two-sample F-test). Similarly, when we repeated the analysis without the neurons that showed a significant effect on the single-cell level, the distribution of amplitude differences was still significantly wider that the distribution under the null hypothesis (p = 0.046).

Next, we explored whether attentional amplitude (gain) changes were related to the neurons' preferred motion directions (as predicted by the feature-similarity hypothesis) and tested the strength of gain modulation across different distances of the neurons' preferred directions to the sample direction (Figure 5D). Instead of feature sensitivity (used before), we now used amplitude changes as a measure for gain modulation. The amplitude differences were not different across the different distances (p = 0.29, Kruskal-Wallis test, $\varepsilon^2 = 0.11$). If tested with the stricter criterion (r-square value > 0.6 in at least one context), again there was no significant difference in amplitude size for the different distances from the preferred direction (p = 0.38, Kruskal-Wallis test, $\varepsilon^2 = 0.2$). Thus, no evidence was found that gain changes in the two attentional contexts were dependent on the neurons' preferred tuning direction. None of the effects described above changed if we used amplitudes normalized between 0 and 1 in the Gaussian fits. The width of the distribution of amplitude differences was still significantly wider than the distribution under the null hypothesis (p = 0, randomization test; real IQR, 0.95; randomization distribution IQR minimum, 0.59; maximum, 0.90). and the amplitude differences did not vary with distance of the sample from the preferred direction (p = 0.73, Kruskal-Wallis-test).

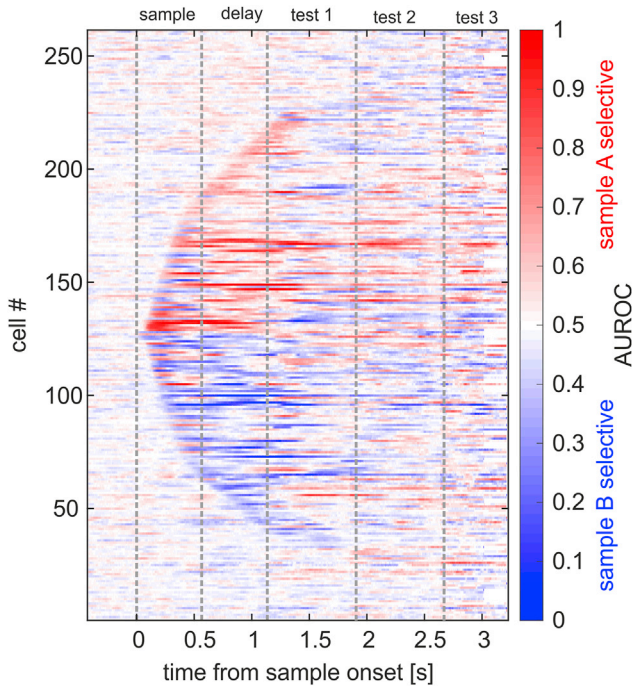


Figure 6. Sample selectivity of neurons throughout the trial

(A) Time course of sample selectivity. Each row depicts the sample selectivity (AUROC value, sample A versus sample B) for one neurons over time. Red signifies selectivity for sample A, and blue signifies selectivity for sample B. Neurons are sorted according to their selectivity in the delay period and latency.

Sample selectivity

Does sample selectivity during the sample and delay phase influence the observed gain effects during the test phase? To test if the gain modulation effects simply reflect sample selectivity that is already present beforehand in the trial (based on sustained activation), we derived the AUROC value as a measure of sample-related selectivity over the time course of the trial (Figure 6). Neurons are sorted according to their selectivity during the delay. It shows that the selectivity of many neurons is highly dynamic and changes between the different time periods. Furthermore, we correlated AUROC values during the sample period and delay period with the AUROC differences plotted in Figures 4B and 4D. We found no correlation for either gain modulations of directional tuning curves (sample: $r = 0.1$, $p = 0.15$; delay: $r = 0.03$, $p = 0.65$) or gain modulations of color tuning curves (sample: $r = -0.49$, $p = 0.6$; delay: $r = 0.13$, $p = 0.16$). Consistently, we found no significant correlation between sample selectivity in the sample or the delay and gain modulations measured as amplitude differences as plotted in Figure 5C (sample: $r = 0.085$, $p = 0.22$; delay: $r = 0.09$, $p = 0.18$). The gain modulation effects, therefore, cannot be explained by preceding sustained sample selectivity.

No tuning shifts of direction-selective PFC neurons

In visual cortex, the tuning preference of neurons sometimes shifts under the influence of attention (David et al., 2008; Ibos and Freedman, 2014). We therefore explored potential motion direction shifts for motion-direction-selective PFC neurons. In order to compare the neurons' preferred directions between context A and context B, we only used neurons that were motion-direction selective in both contexts ($n = 60$). Each individual neuron's directional tuning vector was calculated separately for context A and context B trials (see STAR Methods). Next, we calculated the angular differences between the reference direction of sample B (270°) and each neuron's tuning vector separately during context A and context B. If the context has no influence on the tuning direction, then, if plotted against each other (Figure 7A), the neurons are expected to lie on the unity line. In contrast, a shift toward the relevant context would result in more data points lying under the unity line, whereas a shift away from the attended context direction would result in data points above the unity line. Only 15% of the neurons (9/60; Figure 7A) shifted their tuning direction between contexts (bootstrap analysis, criterion $p < 0.05$). Of these, 6 neurons shifted their tuning profile toward the currently relevant sample (blue data points in Figure 7A), whereas 3 neurons shifted away (green data points in Figure 7A). The majority of neurons maintained their preferred direction as a function of the attended direction (open data points in Figure 7A).

This maintenance of tuning was confirmed at the population level. If tuning would systematically shift toward the attended direction, the tuning distances of the data points relative to the unity line in Figure 7A would increase. However, the distribution of these distance values remained centered around the unity line (distance value 0) (mean shift amplitude = +2.9° ± 3.3° [SEM], $p = 0.38$, $n = 60$, one-sample t test) (Figure 7B).

We also explored potential tuning shifts in color-selective PFC neurons. However, only a small fraction of neurons (8% [9/115]) were selective in both contexts. None of the neurons significantly shifted their color preference (bootstrap analysis, criterion $p < 0.05$). Due to the low number of neurons, we did not test for systematic shift on the population level. Taken together, these results suggest that attention had a negligible effect on shifting tuning preference of both motion-direction-selective and color-selective PFC neurons.

DISCUSSION

In this study, we found that PFC neurons experienced strong gain modulation in response to attentional demands. Surprisingly, and in contrast to assumptions made by the feature-similarity gain model established for attention effects in sensory cortices, neuronal gain modulation was not bound to the tuning preferences of the neurons.

PFC as central source of top-down feature-based attention

In order to investigate feature-based attention effects at the site of their origin, we recorded single-unit activity in the posterior principal sulcus region of the lateral PFC. This brain region is considered to be a central source of top-down attentional modulation (Everling et al., 2002; Rainer et al., 1998a). Our recording location included a subregion of ventrolateral PFC anterior and ventral to the arcuate sulcus termed the ventral pre-arcuate region (VPA). Several findings identify the VPA as a key area for feature-based

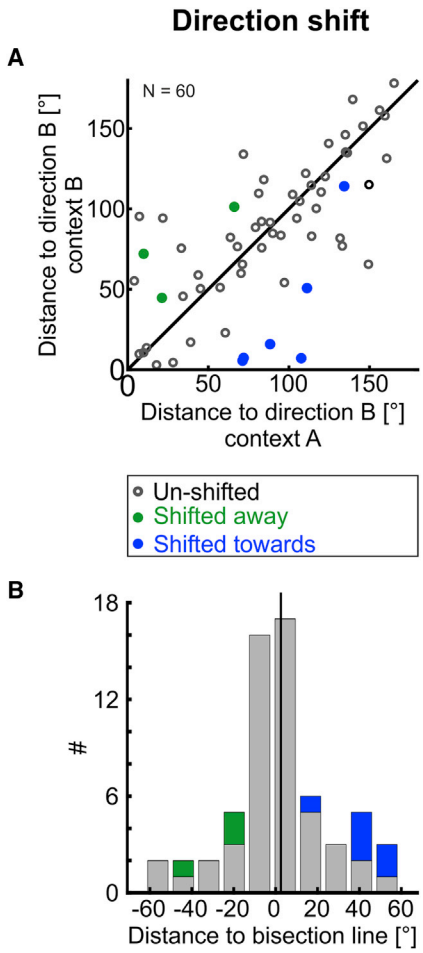


Figure 7. Exploration of potential direction tuning shifts in direction-selective PFC neurons

(A) Distribution of angular differences between the direction of sample B (270°) and each neuron's direction vectors during context A and context B. The angular distance in context A is shown on the x axis against the distance in context B on the y axis. Blue points indicate neurons whose preferred direction is significantly shifted toward the relevant direction, green points indicate neurons whose preferred direction is significantly shifted away from the relevant direction, and open circles indicate neurons that are not significantly shifted.

(B) The distribution of angular distances to the unity line in (A) is shown as a histogram. Green bars indicate shifted-away neurons, and blue bars indicate shifted-toward neurons. The average shift amplitude in the population is shown as a black vertical line.

attention in monkeys performing a feature-based search task (Bichot et al., 2015, 2019). First, neurons in VPA showed feature-based attentional modulation signals earlier than neurons in the neighboring FEF on which visuo-spatial attention depends on (Gregoriou et al., 2014; Moore and Armstrong, 2003; Moore and Fallah, 2001). Second, in contrast to other brain areas simultaneously recorded from (such as frontal eye field [FEF], inferior temporal cortex [IT], or ventral bank of the principal sulcus), only in VPA neurons were selective to the memorized object throughout an entire search period (Bichot et al., 2015). The PFC, in particular, is well known to exhibit object and feature selectivity in a variety of working memory tasks (Mendoza-Halliday et al., 2014; Rainer et al., 1998b; Sarma et al., 2016; Zaksas and Pasternak, 2006). The PFC, thus, seems to contain neuronal "target templates" kept in working memory for the searched-for object features of a target. Third, inactivation of VPA not only impaired the animals' ability to find the target (in the contralateral visual field) (Bichot et al., 2015),but also eliminated or greatly reduced feature-based attentional modulation in area V4 neurons in the same hemisphere (Bichot et al., 2019). This implies that this part of the PFC is a (direct or indirect) source of feature-based attention effects in V4 and most likely also other sensory areas. Results like these provide strong evidence for a prefrontal mechanism for feature-based attention via top-down feedback to upstream sensory brain areas. In the following, we discuss our findings concerning several key effects of feature-based attention in PFC.

Feature-based attention causes no systematic shifts in preferred tuning of PFC neurons

One effect we investigated were potential shifts in tuning preference. We found no evidence for feature-based attention to cause systematic tuning shifts of PFC neurons. In our study, attention had a negligible effect on shifting tuning preference of either motion-direction-selective or color-selective PFC neurons. This finding in PFC contrasts attentional effects reported for extrastriate cortex. Attention has been shown to alter the tuning of visual and parietal neurons; tuning often shifts to more closely match the attended feature or location. For example, feature-based attention shifted spatial frequency and orientation tuning of V4 neurons (David et al., 2008), and spatial attention shifted RF locations toward the attended location (Womelsdorf et al., 2006). Similarly, a larger proportion of direction-selective neurons in lateral intraparietal area (LIP) shifted their direction tuning profile toward the direction of the relevant sample, whereas fewer neurons shift their tuning away from the relevant direction (Ibos and Freedman, 2014).

Such tuning shifts in extrastriate cortex (just as gain modulation) are thought to be mediated by top-down feedback signals most likely originating from the PFC (Compte and Wang, 2006). However, irrespective of attentional demands, we find that PFC neurons maintained their preferred tuning comparable to situations without attentional demands in a passive-viewing task. Therefore, they seem to function as labeled lines that encode input features consistently, regardless of the state of attention.

Feature attention causes strong gain modulation in PFC neurons

We observed strong gain modulation of PFC neurons in response to the monkeys attending to different sensory features and parameters in the two behavioral contexts, A and B. As measures of attentional gain modulation, we compared not only the neurons' tuning sensitivity but also the differences in tuning amplitudes derived from Gauss fits modeled to the neuronal tuning functions. Both parameters showed that single neurons and the population of feature-selective neurons experienced significant gain modulation. Moreover, our population classifier analysis showed that feature-selective neurons were strongly and systematically influenced by the contexts the monkeys were attending to.

Since we did not map the RFs of PFC neurons, it is conceivable that some of the neurons had their RF outside of the visual stimulation zone in our fixating monkeys and therefore have not been stimulated. Thus, the overall frequency of tuned neurons might be higher than encountered. However, neurons in the frontal and parietal association cortices are known to be tuned to attended features irrespective of whether the displays and the RF of a neuron overlap or not (Freedman and Assad, 2009; Viswanathan and Nieder, 2020). If and to what extent RFs influenced the proportion of tuned neurons is therefore difficult to assess.

Previous work in extrastriate visual cortices found that feature-based attention causes changes in the activity of feature-selective neurons consistent with gain modulation. Attending to a particular direction increases the response gain of MT neurons tuned to the attended direction (Chen et al., 2012; Martinez-Trujillo and Treue, 2004; Treue and Martínez Trujillo, 1999). Similarly, attending to form (Bichot, 2005), color (Bichot, 2005), or orientation (McAdams and Maunsell, 1999) modulates V4 neurons. Such gain modulation is thought to increase the neuronal signal-to-noise ratio of relevant sensory features at the expense of irrelevant sensory features, which ultimately leads to increasing task performance for attended features (Cohen and Maunsell, 2009; Lee and Maunsell, 2009; McAdams and Maunsell, 1999; Reynolds and Heeger, 2009; Treue and Maunsell, 1996).

Feature-similarity gain modulation arises from feature-independent top-down modulation in PFC

Attentional gain modulation in extrastriate cortex is thought to be mediated by top-down feedback signals originating in the PFC (Ardid et al., 2007; Bichot et al., 2015, 2019). In visual cortex, the gain modulation affecting neurons shows a systematic layout; only those neurons experience gain modulation whose preferred feature parameter matches the currently attended feature parameters (Maunsell and Treue, 2006). For example, only neurons tuned to upward motion experience gain modulation when the subject attends to upward motion. Gain modulation was found to be progressively reduced if the distance between a neuron's preferred direction and the attended direction increased, with a suppression for large distances (Martinez-Trujillo and Treue, 2004; Treue and Martínez Trujillo, 1999). These effects in gain modulation as a function of the neurons' preferred tuning have been captured by the feature-similarity gain model (Bichot, 2005; Maunsell and Treue, 2006; Treue and Martínez Trujillo, 1999).

It is tacitly assumed that feature similarity also holds for top-down feedback signals (originating from PFC) that increase the gain of neurons whose tuning matches the target of attention (Ardid et al., 2007). However, we find this prediction not realized in primate PFC neurons. Thus, we found no evidence that the feature-similarity gain model accounts for PFC neurons that constitute the origin of top-down attentional modulation. We also found no correlation between sample selectivity in the delay period and the observed gain modulations in the test period. This suggests that the gain modulation is not simply sample selectivity spilling over from the delay period. However, it is conceivable that gain modulation in PFC reflects some kind of mixed selectivity (Mante et al., 2013; Rigotti et al., 2013), which could be part of the mechanism of how attentional modulation is realized in PFC and perhaps other higher-order areas.

This finding requires adjustment of computational models simulating attentional processing along the cortical attentional processing hierarchy, which so far assumed feature-similarity modulation up to the origin of the top-down attentional signal (Ardid et al., 2007). We suspect this observed type of feature dissimilarity provides PFC with more degrees of freedom when instructing more sensory brain areas that receive top-down projections during attentional modulation. For instance, PFC not only needs to enhance the signal-to-noise ratio of sensory neurons in accordance with the attended features but also at the same time suppresses other neurons that encode currently irrelevant features. Moreover, PFC not only directs attentional selection but also needs to flexibly deal with a multitude of cognitive processes in real-life situations, ranging from sensory processes (Jacob et al., 2013; Hage and Nieder, 2015; Stalter et al., 2020) to working memory (Rainer et al., 1998b; Jacob and Nieder, 2014) and conceptual decisions (Vallentin et al., 2012; Viswanathan and Nieder, 2015; Ramirez-Cardenas et al., 2016; Wallis et al., 2001). Here as well, the dissociation between preferred tuning of PFC neurons and other task demands may constitute a computational advantage.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Experimental setup
  - Surgical procedures
  - Task and stimuli
  - Extracellular recordings
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioral data analysis
  - Direction- and color-selective neurons
  - Single-cell and population responses
  - Direction tuning – polar plots
  - Uniformity of circular data
  - Comparison of direction tuning during DCMT and PVT
  - Direction- and color tuning shifts
  - Receiver operating characteristic analysis
  - Context-dependent feature tuning sensitivity
  - Gaussian tuning curve fit
  - Gauss fits to motion direction tuning profiles
  - Sample selectivity
  - Multi-Class Support Vector (SVM) classification.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109470.

ACKNOWLEDGMENTS

This research was supported by DFG FOR 1847 grants NI 618/5-1 and NI 618/5-2. We thank Katharina Brecht for helpful discussions.

AUTHOR CONTRIBUTIONS

M.S. and A.N. designed the experiments. M.S. recorded the data, and M.S. and S.W. analyzed the data. M.S., S.W., and A.N. wrote the paper. A.N. supervised the study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 29, 2020
Revised: May 31, 2021
Accepted: July 9, 2021
Published: August 3, 2021

REFERENCES

Ardid, S., Wang, X.-J., and Compte, A. (2007). An integrated microcircuit model of attentional processing in the neocortex. J. Neurosci. *27*, 8486–8495.

Baldauf, D., and Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. Science *344*, 424–427.

Bichot, N.P. (2005). Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4. Science *308*, 529–534.

Bichot, N.P., Heard, M.T., DeGennaro, E.M., and Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. Neuron *88*, 832–844.

Bichot, N.P., Xu, R., Ghadooshahy, A., Williams, M.L., and Desimone, R. (2019). The role of prefrontal cortex in the control of feature attention in area V4. Nat. Commun. *10*, 5727.

Buschman, T.J., and Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. Neuron *88*, 127–144.

Buschman, T.J., and Miller, E.K. (2007). Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. Science *315*, 1860–1862.

Chang, C., and Lin, C. (2011). LIBSVM. ACM Trans. Intell. Syst. Technol. *2*, 1–27.

Chen, X., Hoffmann, K.-P., Albright, T.D., and Thiele, A. (2012). Effect of feature-selective attention on neuronal responses in macaque area MT. J. Neurophysiol. *107*, 1530–1543.

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

Compte, A., and Wang, X.-J. (2006). Tuning curve shift by attention modulation in cortical neurons: a computational study of its mechanisms. Cereb. Cortex *16*, 761–778.

David, S.V., Hayden, B.Y., Mazer, J.A., and Gallant, J.L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron *59*, 509–521.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. *18*, 193–222.

Everling, S., Tinsley, C.J., Gaffan, D., and Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. Nat. Neurosci. *5*, 671–676.

Freedman, D.J., and Assad, J.A. (2009). Distinct encoding of spatial and nonspatial visual information in parietal cortex. J. Neurosci. *29*, 5671–5680.

Gail, A., Klaes, C., and Westendorff, S. (2009). Implementation of spatial transformation rules for goal-directed reaching via gain modulation in monkey parietal and premotor cortex. J. Neurosci. *29*, 9490–9499.

Gregoriou, G.G., Rossi, A.F., Ungerleider, L.G., and Desimone, R. (2014). Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4. Nat. Neurosci. *17*, 1003–1011.

Hage, S.R., and Nieder, A. (2015). Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. J. Neurosci. *35*, 7030–7040.

Hayden, B.Y., and Gallant, J.L. (2005). Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4. Neuron *47*, 637–643.

Ibos, G., and Freedman, D.J. (2014). Dynamic integration of task-relevant visual features in posterior parietal cortex. Neuron *83*, 1468–1480.

Jacob, S.N., and Nieder, A. (2014). Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. Neuron *83*, 226–237.

Jacob, S.N., Ott, T., and Nieder, A. (2013). Dopamine regulates two classes of primate prefrontal neurons that represent sensory signals. J. Neurosci. *33*, 13724–13734.

Lee, J., and Maunsell, J.H.R. (2009). A normalization model of attentional modulation of single unit responses. PLoS ONE *4*, e4651.

Lee, J., and Maunsell, J.H.R. (2010). The effect of attention on neuronal responses to high and low contrast stimuli. J. Neurophysiol. *104*, 960–971.

Liu, T. (2019). Feature-based attention: effects and control. Curr. Opin. Psychol. *29*, 187–192.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

Martinez-Trujillo, J.C., and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr. Biol. *14*, 744–751.

Maunsell, J.H.R., and Treue, S. (2006). Feature-based attention in visual cortex. Trends Neurosci. *29*, 317–322.

McAdams, C.J., and Maunsell, J.H.R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J. Neurosci. *19*, 431–441.

McAdams, C.J., and Maunsell, J.H.R. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. J. Neurophysiol. *83*, 1751–1755.

Mendoza-Halliday, D., Torres, S., and Martinez-Trujillo, J.C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. Nat. Neurosci. *17*, 1255–1262.

Moore, T., and Armstrong, K.M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. Nature *421*, 370–373.

Moore, T., and Fallah, M. (2001). Control of eye movements and spatial attention. Proc. Natl. Acad. Sci. USA *98*, 1273–1276.

Rainer, G., Asaad, W.F., and Miller, E.K. (1998a). Selective representation of relevant information by neurons in the primate prefrontal cortex. Nature *393*, 577–579.

Rainer, G., Asaad, W.F., and Miller, E.K. (1998b). Memory fields of neurons in the primate prefrontal cortex. Proc. Natl. Acad. Sci. USA *95*, 15008–15013.

Ramirez-Cardenas, A., Moskaleva, M., and Nieder, A. (2016). Neuronal Representation of Numerosity Zero in the Primate Parieto-Frontal Number Network. Curr. Biol. *26*, 1285–1294.

Reynolds, J.H., and Heeger, D.J. (2009). The normalization model of attention. Neuron *61*, 168–185.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585–590.

Saalmann, Y.B., Pigarev, I.N., and Vidyasagar, T.R. (2007). Neural Mechanisms of Visual Attention: How Top-Down Feedback Highlights Relevant Locations. Science *316*, 1612–1615.

Sarma, A., Masse, N.Y., Wang, X.-J., and Freedman, D.J. (2016). Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. Nat. Neurosci. *19*, 143–149.

Squire, R.F., Noudoost, B., Schafer, R.J., and Moore, T. (2013). Prefrontal contributions to visual selective attention. Annu. Rev. Neurosci. *36*, 451–466.

Stalter, M., Westendorff, S., and Nieder, A. (2020). Dopamine Gates Visual Signals in Monkey Prefrontal Cortex Neurons. Cell Rep. *30*, 164–172.e4.

Thiele, A., Delicato, L.S., Roberts, M.J., and Gieselmann, M. (2006). A novel electrode-pipette design for simultaneous recording of extracellular spikes and iontophoretic drug application in awake behaving monkeys. J. Neurosci. Methods *158*, 207–211.

Tomczak, M., and Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. Trends Sport Sci. *1*, 19–25.

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

Treue, S., and Maunsell, J.H.R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. Nature *382*, 539–541.

Vallentin, D., Bongard, S., and Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. J. Neurosci. *32*, 6621–6630.

Viswanathan, P., and Nieder, A. (2015). Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. Curr. Biol. *25*, 1259–1269.

Viswanathan, P., and Nieder, A. (2020). Spatial Neuronal Integration Supports a Global Representation of Visual Numerosity in Primate Association Cortices. J. Cogn. Neurosci. *32*, 1184–1197.

Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature *411*, 953–956.

Womelsdorf, T., Anton-Erxleben, K., Pieper, F., and Treue, S. (2006). Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nat. Neurosci. *9*, 1156–1160.

Womelsdorf, T., Anton-Erxleben, K., and Treue, S. (2008). Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. J. Neurosci. *28*, 8934–8944.

Zaksas, D., and Pasternak, T. (2006). Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. J. Neurosci. *26*, 11726–11742.

Zar, J.H. (2010). Biostatistical analysis, Fifth Edition (Prentice-Hall: Pearson).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: organisms/strains		
Macaca mulatta	German Primate Centre, Göttingen	https://www.dpz.eu/
Software and algorithms		
NIMH Cortex	National Institute of Mental Health	c598; https://www.nimh.nih.gov/research/research-conducted-at-nimh/research-areas/clinics-and-labs/ln/shn/software-projects.shtml
MAP Data Acquisition System	Plexon	https://plexon.com/
MATLAB R2018a		https://www.mathworks.com
MATLAB code	MathWorks	https://figshare.com/articles/software/Matlab_files_for_Feature-based_attention_processes_in_primate_prefrontal_cortex_do_not_rely_on_feature_similarity_/14899245
LIBSVM version 3.23	Chang and Lin, 2011	https://www.csie.ntu.edu.tw/~cjlin/libsvm/
Other		
Dental Cement	Heraeus	Paladur, ISO 20795, CE 0197
Microdrives	Modified NAN C-drive	https://naninstruments.com/
Electrodes	Animal Physiology Unit	Custom fabrication
Luminance meter	Konica Minolta	LS100

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Andreas Nieder (andreas.nieder@uni-tuebingen.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Analyzed data reported in this paper will be shared by the lead contact upon request.

All original code has been deposited at https://figshare.com/articles/software/Matlab_files_for_Feature-based_attention_processes_in_primate_prefrontal_cortex_do_not_rely_on_feature_similarity_/14899245 and is publicly available as of the date of publication. DOIs are listed in the key resources table.

Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We used two healthy 8-year-old, male rhesus monkeys (Macaca mulatta). The monkeys were obtained from the German Primate Center (DPZ) in Göttingen. The animals were housed indoors in social groups. During the training and recording period the monkeys were on a controlled feeding protocol. They received their daily amount of water as reward during the recording sessions. If necessary additional water was given after the sessions. The animal's body weight was measured daily during that time. Food was available *ad libitum* in the group cage. All experimental procedures were in accordance with the guidelines for animal experimentation determined by the responsible authority, the Regierungspräsidium Tübingen.

METHOD DETAILS

Experimental setup

During the experiment the monkeys sat in a darkened operant conditioning chamber in front of a computer screen. Visual stimuli were displayed on this monitor and the monkey had to respond to the stimuli by releasing a metal bar inside the primate chair. They received fluid reward via a mouthpiece attached to the chair. During each trial, the monkey had to maintain eye fixation within 3.5° visual angle of the central fixation target (ISCAN). Display of visual stimuli and control of the animals' behavior, as well as behavioral data acquisition was done by CORTEX software (NIMH, Bethesda, MD). Neuronal data was recorded with a PLEXON system (MAP box, Plexon Inc., Dallas, Texas).

Surgical procedures

We implanted two healthy, 8-year-old male rhesus macaques (*Macaca mulatta*) with titanium head posts and one recording chamber each. The chamber was centered on the right hemisphere over the principal sulcus of the lateral PFC, anterior to the frontal eye field in both monkeys. All surgical procedures were performed under aseptic conditions using general anesthesia. Chamber implantations were guided by landmarks obtained through magnetic resonance imaging prior to the surgeries.

Task and stimuli

The monkeys were trained to match visually presented random dot patterns based on two conjunct visual features, namely the direction of motion and the color of the dots, against one of two possible sample stimuli. The test stimuli varied in eight directions, evenly spaced across the full 360°, and in eight color hues, ranging from red to yellow. Every combination of direction and color was possible, resulting in a set of 64 unique stimuli. All eight directions and colors will be referred to as levels of their respective feature. The first sample stimulus (A) consisted of red dots moving upward, the second sample stimulus (B) consisted of yellow dots moving downward. The two sample stimuli were the same for each day and each session. The test stimuli were chosen from a set of all possible combinations of directions and colors for both contexts. Therefore, non-matching stimuli belonged to one of four categories of stimuli: (1) the test stimulus matched the sample stimulus only in color but not direction, (2) the test stimulus matched the sample stimulus only in direction but not color, (3) the test stimulus matched the sample stimulus that was not presented in the current trial (e.g., sample stimulus B during context A trials and vice versa), or (4) the test stimulus matched the sample stimulus neither in color nor in direction. Stimuli form category 3 were presented three times as often as stimuli from the other categories to prevent monkeys from simply ignoring the sample identity.

The monkeys initiated each trial by holding a metal bar and fixating on a central fixation target (fixation period). After 500 ms one of the two samples appeared for 550 ms (sample period) setting the context and cueing the monkey about the two relevant features for this trial. In the following delay period (550 ms) the screen was black, except for the fixation target, and the monkeys had to keep the presented sample stimulus in memory. In the subsequent test period, the stimulus matching the presented sample in direction and color was shown for 550 ms in 25% of the trials (match trial). In the other 75% of the trials, one (25%), two (25%) or three (25%) of non-match test stimuli were shown (550 ms each). One or two non-match test stimuli were followed by a matching test stimulus. Three non-match test stimuli were never followed by a matching stimulus (catch trials). The inter stimulus time between all test stimuli was 150 ms. The monkeys were rewarded with fluid if they released the bar in response to a test stimulus matching in both visual features, or for maintaining the bar for further 20 ms after the offset of the last non-matching stimulus in catch trials. All trials types were presented pseudo randomly.

Stimuli were circular patches of random dots 5° of visual angle (dva) in diameter. The overall dot density was 12 per dva^2 with a radius of 0.04 dva for each dot and they moved with 100% coherence at a speed of 4 dva/s. All stimuli were generated using MATLAB (The Mathworks). Colors were generated in the LAB color space (1976 CIE L*a*b) and all colors were controlled for iso-luminance under experimental conditions, using a LS-100 luminance meter (Konica Minolta).

Extracellular recordings

On each session, up to three custom-made, tungsten-in-glass electrodes (Thiele et al., 2006) with two flanking barrels were lowered transdurally into the brain using a modified electrical drive (NAN Drive). We randomly recorded single neurons and made no attempt to preselect neurons. Signal acquisition, filtering, amplification and digitalization were accomplished with the MAP system (Plexon). Waveform separation was performed offline (Offline Sorter; Plexon). Electrode impedances were measured after the recordings and ranged between 0.5 and 3.5 MΩ (measured at 500 Hz; Omega Tip Z; World Precision Instruments).

QUANTIFICATION AND STATISTICAL ANALYSIS

All data analysis was performed using the R2019a release of MATLAB software (The MathWorks). In order to eliminate hand movement related activity, we excluded the responses to matching stimuli and only analyzed responses to non-matching stimuli. For population analysis the neuronal data were merged for both monkeys, as they were similar. We used the same time window for all analysis as described in the section 'Direction- and color-selective neurons'.

The following statistical tests were used as appropriate for the data: a Kruskal-Wallis test was used to define feature selective neurons. A Rayleigh test was to test for circular uniformity in the distribution of tuning vectors. A v-test was used in the instances in which the distribution of tuning vector directions was tested for an orientation into a specific direction against uniformity. A one-sample t test was used to test if the distribution of tuning vector shift angles was significantly different from zero. We used randomization tests to test if a distribution of gain changes was broader than expected under the null-hypothesis. We created a respective distribution under the null-hypothesis by randomly assigning trials to one of the two possible contexts. To test if sensitivity differences were different between two different groups of neurons, we used a Mann-Whitney U-test.

Data are presented as mean ± standard error of the mean (SEM) unless indicated otherwise. $p < 0.05$ was considered to be statistically significant. For all analyses the exact statistical test including p values, dispersion and precision measures are given in the result section.

Behavioral data analysis

The false alarm rate in response to every possible test stimulus (64 combinations: 8 direction x 8 colors) was analyzed. False alarm rate (FAR) was determined as the number of erroneous bar releases (false alarms; FA) to non-matching test stimuli divided by the sum of false alarms plus correct bar maintenance (correct rejections; CR): FAR = FA/(FA+CR).

Direction- and color-selective neurons

We were interested in how the context, i.e., attention to a currently relevant feature, affected the feature tuning of PFC neurons. Importantly, the visual input during the analyzed test-period was identical for both contexts. Any activity differences can be explained only by the fact that the monkey was attending to the currently relevant feature. Neurons were included in the analysis if they were recorded with at least 4 correct trials per direction and color. Similar results were obtained if we used a minimum of 7 correct trials per condition (data not shown). In total we recorded 489 neurons that fulfilled this criterion. By visual inspection, many neurons differed in their time course of selectivity to the stimulus features direction and color. Additionally, selectivity depended on the currently relevant context. Therefore, for each context separately, a sliding window Kruskal-Wallis test was performed on the firing rates in the test period from 0.1 s to 0.7 s after test stimulus onset (bin width 150 ms, shifted in steps of 5 ms). In each context, onset and offset of selectivity were defined as the time from the first to the last significant window ($p < 0.05$), but only if at least 10 consecutive bins were significant. If onset and offset of selectivity could be determined only in one context, the same time window was taken for the other context. In cases where onset and offset of selectivity could be determined in both contexts, the window started at the earliest onset and ended at the latest offset from both contexts. All further analysis concerning the test period used the average firing rate in the thus defined time window, which was tailored to best capture each neurons stimulus-feature-selectivity.

A neuron was counted as direction or color selective, if the average firing rates in the above defined time window were significantly different for the analyzed feature in at least one context ($p < 0.05$) using a Kruskal-Wallis test. None of the reported results depended on the exact choice of trials or time window to analyze. Similar results were achieved using a different set of parameters.

Single-cell and population responses

For single-cell responses of example neurons, spike-density functions were generated, i.e., individual trials were parsed in 10 ms bins and spikes convoluted with a Gaussian function with a width of sigma = 75 ms. Activity was then averaged for every 10 ms bin over all trials of a given stimulus dimension.

Direction tuning – polar plots

The selectivity profile and selectivity strength of all direction selective single neurons was assessed by computing a directional vector separately on sample A and B trials using Equation 1:

$$\begin{cases} X = \sum_{i=1}^{8} FR(i) * \cos(direction(i)) \\ Y = \sum_{i=1}^{8} FR(i) * \sin(direction(i)) \end{cases} \quad \text{Equation 1}$$

Here, FR(i) is the neurons' firing rate to the i^{th} direction, yielding the Cartesian coordinates [0 X] and [0 Y] of the directional vector.

Uniformity of circular data

To test whether the preferred directions of our recorded motion direction-selective population were uniformly distributed, we used circular statistics (Rayleigh test) as conventional statistics are not suited for circular data.

Comparison of direction tuning during DCMT and PVT

For a subset of neurons (n = 418), we also recorded activity during a sensory-driven passive viewing task (PVT). PVT trials were randomly interleaved with trials of the delayed conjunction matching task (DCMT). The monkeys watched a series of visual random dot patterns moving in the center of the screen. Stimuli were circular patches of random dots 5° of visual angle (dva) in diameter. The

overall dot density was 12 per dva^2 with a radius of 0.04 dva for each dot and they moved with 100% coherence at a speed of 4 dva/s. The movement of the dots varied in eight directions, spaced evenly across the full 360°. The color of all dots was a light gray. A trial was initiated by holding a metal bar and maintaining gaze on a central fixation target (fixation period). After 300 ms, a series of eight stimuli of pseudo-randomized motion directions was presented in sequence. At the end of each trial, a fluid reward was delivered if the monkey maintained holding the bar and maintained fixation throughout the stimulus presentation sequence. Each stimulus was presented for 300 ms and contained dot patterns moving in one of eight directions. Within one trial every possible direction was presented only once and the time between each stimulus was 100 ms.

Neurons were included in the analysis of the PVT if they were recorded with at least 10 trials per direction. We performed a Kruskal-Wallis test on the firing rates during stimulus presentation (0.18 s to 0.48 s after stimulus onset). A neuron was counted as direction-selective during the PVT if the p value of the Kruskall-Wallis test was significant ($p < 0.05$). In total we found 73 neurons to be selective for the direction-of motion during the PVT. From this pool of neurons, 45 were also motion direction-selective during the DCMT. These neurons were used for the comparison of motion direction tuning during the DCMT and PVT. None of the reported results depended on the exact choice of time window for the PVT analysis. Similar results were obtained using different parameters.

Tuning vectors for PVT data were calculated as described for the DCMT data. By visual inspection, the tuning profiles of the sensory PVT and the attentional DCMT task were very similar. To test, that the neuronal tuning profile is indeed similar in both tasks, we calculated the angular differences of the tuning vectors during the PVT and the DCMT. Under the null hypothesis the differences between the tuning vector in the PVT and the DCMT should be uniformly distributed around the circle. The alternative hypothesis is that the angle differences cluster around 0°. We tested this, using a v-test for circular data at an alpha of 0.05.

Direction- and color tuning shifts

For the population of neurons that were motion direction-selective during both contexts during the DCMT (n = 60), we normalized the firing rates to the neurons maximum in the test period. For each neuron we shuffled firing rates between sample A and B, within their respective direction and with replacement (analysis was analog to Ibos and Freedman (2014)). This was repeated 1000 times and for each repetition we calculated the preferred directions for both contexts individually according to Equation 1, yielding 1000 values of PrefDIR-A-shuffle and PrefDIR-B-shuffle. The preferred directions of shuffled data were then compared to direction of sample B (270°), i.e., DIR-BtoPrefDIR-A-shuffle and DIR-BtoPrefDIR-B-shuffle. We calculated all the possible combinatory differences between the 1000 DIR-BtoPrefDIR-A-shuffle and DIRBtoPrefDIR-B-shuffle values, yielding 1000000 differences. In this distribution, negative values mean that PrefDIR-A-shuffle is closer to the direction of sample B then PrefDIR-B-shuffle. We counted a neurons preferred direction as significantly shifted toward the relevant direction of the current context, if at least 97.5% ($p \leq 0.05$) of these differences were positive. If at least 97.5% ($p \leq 0.05$) of the differences were negative, we counted the preferred direction as significantly shifted away from the relevant direction. To see if the context dependent shifts exist also on the level of the population of these direction selective neurons, we calculated the distance to the bisection line for each neuron individually. A one-sample t test was used to test if the distribution of distances to the bisection line was significantly shifted from zero.

To analyze tuning shifts for the color data on the single cell level, we performed a bootstrap analysis similar to the direction data, however, our linear color space prevented us from using vector-based analyses of color-tuning shifts. For each neuron we shuffled firing rates between sample A and B, within their respective color and with replacement. This was repeated 1000 times and for each repetition we calculated the index of preferred colors for both contexts individually, yielding 1000 values of PrefCOL-A-shuffle and PrefCOL-B-shuffle (e.g., for a preferred color red the index was 1 and increased by one with each intermediate color. The index for a preferred color yellow was 8).

The preferred colors of shuffled data were then compared to the color index of sample B (index 8), i.e., COL-BtoPrefCOL-A-shuffle and COL-BtoPrefCOL-B-shuffle. We calculated all the possible combinatory differences between the 1000 COL-BtoPrefCOL-A-shuffle and COL-BtoPrefCOL-B-shuffle values, yielding 1000000 differences. In this distribution, negative values reveal that PrefCOL-A-shuffle is closer to the color of sample B then PrefCOL-B-shuffle. We counted a neurons preferred color as significantly shifted toward the relevant color of the current context, if at least 97.5% ($p \leq 0.05$) of these differences were positive. If at least 97.5% ($p \leq 0.05$) of the differences were negative, we counted to preferred color as significantly shifted away from the relevant color.

Receiver operating characteristic analysis

Tuning sensitivity for both stimulus features (direction and color) was quantified using receiver operating characteristic (ROC) analysis, derived from signal detection theory. For each neuron, each context and each feature separately, the true positive rate (discharge rate in response to the preferred feature level; direction or color, respectively) and false positive rate (discharge rate in response to the least-preferred feature level) were calculated in order to generate the ROC curve. Then the area under the ROC curve (AUROC), which is a nonparametric measure of the discriminability of two distributions, was calculated. It denotes the probability with which an ideal observer can tell apart a meaningful signal from a noisy background. Values of 0.5 indicate no separation, and values of 1 signal perfect discriminability. The AUROC is a good indicator of signal quality as it takes into account both the difference between distributions means as well as their width.

Context-dependent feature tuning sensitivity

To explore the influence of feature-based attention on the tuning sensitivity of the population of direction and color selective neurons, the difference of the AUROCs (see above) for both contexts (AUROC-A-trials – AUROC-B-trials) was calculated for each neuron individually. Values of zero indicate no difference in the tuning sensitivity between the two contexts. A value close to −0.5 means that the neuron has a high sensitivity during context B and low sensitivity during context A. A value of 0.5 indicates the reverse. We found many neurons that differ in sensitivity between contexts. These effects are, as expected, observed with equal frequencies with preferences for context A and preferences for context B. Therefore, we have effects of AUROC differences on both tails of the distribution. Consequently, if testing for a population effect, it is not possible to test for a shift of the mean of the distribution away from zero. However, if there are effects to both sides, the hypothesis is that the real distribution is wider than a distribution under the null-hypothesis of no difference between context A and B. We created this Null-distribution by randomly relabelling trials as A- or B-trials, while keeping direction- or color labels intact. We calculated 1000 permutations of AUROC-A-trials and AUROC-B-trials. The differences of the shuffled AUROC values were taken to estimate the null distribution of AUROC differences (AUROC-shuffle-A – AUROC-shuffle-B with $N_{Shuffle} = N_{Neurons}$ x 1000). If feature-based attention has no influence on the population sensitivity in the two contexts, the distribution of real AUROC differences and the null distribution should be equally wide. In contrast, if, as expected, feature-based attention exerts effects in both contexts, the distribution of the real data should be wider than the null distribution (Gail et al., 2009). We used the interquartile range to calculate the width of the distribution. We did this for the distribution of real data and each of the 1000 distributions of shuffled data. If the interquartile range of the real data is outside the 95th percentile (i.e., smaller than the 2.5th percentile or larger than the 97th percentile) of the distribution of interquartile ranges for the shuffled data, the effect was considered significant. The p value is given as the percentage of shuffled values that are larger/smaller than the real interquartile range. Additionally, we used the shuffle distribution to detect single neurons that were significantly sensitivity modulated by the context of sample A or B. We tested for each neuron if its real AUROC difference between A and B trials was outside of 95% of its null-distribution of shuffled AUROC differences ($p < 0.05$, two-sided).

During feature-based attention according to the feature-similarity-gain model (Treue and Martínez Trujillo, 1999), neurons that are tuned to the currently attended direction are gain modulated in visual area MT. The size of the gain modulation decreases with increasing distance of the preferred feature to the attended feature. Therefore, the hypothesis is that in PFC, the potential source for the top-down attention signal, a similar mechanism is in place that produces a gain specifically for neurons with tuning matching the currently attended feature. Therefore, we plotted the size of the gain effects in dependence of the distance of the preferred color/direction to the sample color/direction. For direction, we then separated these distances into bins of 30° ranging from −180° to 180°. For direction within each bin, and for color at each distance, we computed the mean size of the respective gain effect and tested with a Kruskal-Wallis test at an alpha level of 0.05 for differences between the different distances.

Gaussian tuning curve fit

The activity of each motion direction-selective neuron in the DCMT was fitted with a Gaussian function to analyze the individual tuning curve, in each context separately.

$$f(x) = a * e^{-\left(\frac{x-b}{c}\right)^2} + d \qquad \text{Equation 2}$$

We fitted three parameters: amplitude (a), width (c) and offset (d); see Equation 3. MATLABs 'nonlinearleastsquares' algorithm was used to find optimal values for each neuron. Before fitting, the x-values were shifted in such a way that the position of the maximum firing rate (b) was fixed in the center of the curve to deal with the circularity of the data. We used the amplitude values of these fits for the analysis of context dependent gain modulation. We used all such generated fits for further analysis. However, our results do not change qualitatively if we apply a minimum r^2-value of 0.6 for either one of the contexts or both contexts as quality criterion (data not shown).

To generate population tuning curves in Figure 5D, all responses were aligned to the preferred direction and averaged over all neurons for the context the neurons were motion direction-selective in or the respective other context, separately. The average response was then used to fit the population tuning curve using the same fitting procedure as described before. For each direction we calculated the error across neurons as standard error of the mean (SEM).

Gauss fits to motion direction tuning profiles

To test whether feature-based attention modulated the amplitude in the tuning profiles of direction-selective neurons on the population level, we first fitted each neurons response with a Gaussian function (see paragraph "Gaussian tuning curve fit"; above), for each context separately. We then used the amplitude value from our Gaussian fit in each context, which corresponds to the difference between the peak and the tail of the fitted function, to calculate the difference in the amplitudes between the two contexts for each neuron individually (Amplitude-A-trials – Amplitude-B-trials). By randomly relabelling trials as A- or B-trials, while keeping direction labels intact, we calculated 1000 permutations of Amplitude-A-trials and Amplitude-B-trials. The differences of the shuffled amplitude values were taken to estimate the null distribution of amplitude differences (Amplitude-shuffle-A – Amplitude-shuffle-B with $N_{Shuffle} = N_{Neurons}$ x 1000).

As described for the sensitivity modulation, we expected feature-based attention to exert amplitude-effects in both contexts. Therefore, we tested with an F-test, if the distribution of real amplitude differences was wider than the null distribution ($p < 0.05$, two- sided). For detecting single neurons that were significantly context modulated in feature coding strength, we tested for each neuron if its real amplitude difference between A and B trials was outside the 95% confidence interval of the distribution of shuffled amplitude differences.

To analyze if the amplitude differences in motion-direction selective neurons differed between the group of 'matching-feature'-encoding and non-matching feature'- encoding neurons, we performed the same analysis as described for the sensitivity analysis but used the amplitude differences (see last section in the paragraph ''Context-dependent stimulus feature tuning sensitivity'' above).

Sample selectivity

To analyze sample selectivity we computed a sliding window ROC analysis between the two sample conditions (context A and B). For each neuron we computed the AUROC value in a 150 ms window at 5 ms time steps starting 0.5 s before sample onset and continuing until 3.3 s after sample onset. To test how sample selectivity before test onset correlated with gain effects during the test phase we correlated (Pearson correlation) sample selectivity measured as AUROC value in the sample period (0.1 s to 0.65 s after sample onset) and in the delay period (0.65 s to 1.1 s after sample onset) to the gain effects analyzed in the test period, measured as AUROC differences (Figure 4) and as amplitude differences (Figure 5).

Multi-Class Support Vector (SVM) classification.

To assess the amount of feature information contained in the populations of direction- and color-selective neurons and to evaluate how this information depends on the current context, we trained a multi-class SVM classifier (Chang and Lin, 2011) (LIBSVM version 3.23). We used a linear SVM-kernel with default parameter settings and applied 'one-versus-one' classification to distinguish our sixteen classes (eight feature levels in two contexts). A separate classifier was built for the population of direction- and color-selective neurons using 7 trials per feature level and neuron (n = 180 motion direction-selective and n = 92 color-selective neurons). For neurons with more than 7 trials we randomly sampled without replacement 7 trials. We normalized all firing rates by z-scoring and used leave-one-out cross-validation. In the confusion matrix the main diagonal contains correct labeling of the classifier. This process was repeated 100 times and the average decoding performance was taken to be the mean of all runs.

To assess the influence of context, we shuffled the context labels of the firing rates while keeping the feature-level labels intact. Using these shuffled data, we trained and tested a new classifier for each population of direction- and color-selective neurons. For training and testing of the multi-class SVM we used the exact same procedure as described in the previous paragraph. If the neuronal firing rates were unaffected by the context, the main diagonal would be less pronounced and two secondary diagonals would appear. These secondary diagonals would then indicate the misclassification of a given direction with that same direction of the other context (refer to Figure S1).