Behavioral/Systems/Cognitive

Neural Correlates of a Decision Variable Before Learning to Perform a Match/Non-Match Task

Xue-Lian Qi,[1] **Travis Meyer,**[1,2] **Terrence R. Stanford,**[1] **and Christos Constantinidis**[1]

[1]Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, and [2]Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

The lateral prefrontal cortex plays an important role in working memory and decision-making, although little is known about how neural correlates of these functions are shaped by learning. To understand the effect of learning on the neuronal representation of decision-making, we recorded single neurons from the lateral prefrontal cortex of monkeys before and after they were trained to judge whether two stimuli appeared at matching spatial locations. After training, and in agreement with previous studies, a population of neurons exhibited activity that was modulated depending on whether the second stimulus constituted a match or not, which had predictive ability for the monkey's choice. However, even before training, prefrontal neurons displayed modulation depending on the match or non-match status of a stimulus, with approximately equal percentages of neurons preferring a match or a non-match. The difference in firing rate and discriminability for match and non-match stimuli before training was of comparable magnitude as that after training. Changes observed after training involved an increase in the percentage of neurons exhibiting this effect, a greater proportion of neurons preferring non-match stimuli, and a greater percentage of neurons representing information about the first stimulus during the presentation of the second stimulus. Our results suggest that the neuronal activity representing some match/non-match judgments is present in the lateral prefrontal cortex even when subjects are not required to perform a comparison and before any training.

Introduction

The lateral prefrontal cortex plays an important role in executive function, allowing the control of behavior in accordance with previous information and future goals (Miller and Cohen, 2001). Accordingly, neural correlates of processes such as decision-making are readily observed in the discharges of prefrontal neurons during execution of tasks that require comparisons of stimuli or choices between alternatives (Kim and Shadlen, 1999; Matsumoto et al., 2003; Barraclough et al., 2004; Zaksas and Pasternak, 2006). For instance, when subjects are required to compare two vibratory stimuli and report if the second is higher in frequency, prefrontal neurons respond differently to the second stimulus depending on how it relates to the first (Romo et al., 1999; Romo and Salinas, 2003). Such activity is likely to form the basis of the comparison judgment, although the extent to which the manifestation of decision variables in neuronal activity is shaped by task demands is not clear. Typically, experimental subjects receive thousands of training trials on stereotypical tasks before recording neural data. At one extreme, activity that differ-

entiates between stimuli being compared may emerge only in the context of performing a well-learned task in which the correct discrimination is associated with a rewarded motor response. Execution of a set of rules that defines the task modulates responses to stimuli (Asaad et al., 1998; White and Wise, 1999; Wallis et al., 2001), possibly conflating the neural representation of the decision with that representing the task itself. At the other extreme, neural activity thought to represent a task-related decision variable may reflect automatic processes that occur even in the absence of an explicit judgment or choice. Few studies have examined the effects of training on prefrontal cortical activity (Sandrew et al., 1977; Kubota and Komatsu, 1985; Asaad et al., 1998; Rainer and Miller, 2000), thus it is not known which of these extremes is a more apt description of prefrontal encoding.

To address this issue, we recorded neuronal activity from the lateral prefrontal cortex of monkeys after they were trained to perform a task that required them to judge whether the locations of two sequentially presented stimuli were the same or different. We compared these responses to those obtained from the same animals, before learning the task at all, when the identical stimuli were presented for passive viewing. We have recently analyzed data obtained via this approach to characterize the responsiveness and selectivity to the stimuli themselves (Meyer et al., 2011; Qi et al., 2011). In the present study, we determined whether information relevant to location judgments is also represented in prefrontal activity before training. We reasoned that if activity differentiating the match or non-match status of a stimulus reflects the active decision process (or emerges solely in the context of task performance), then no such activity should be present before training. Conversely, the presence of such activity before

training would indicate that the prefrontal cortex automatically represents comparative information about stimuli, even when judgments are not explicitly required.

Materials and Methods

Three male, rhesus monkeys (*Macaca mulatta*) weighing 5–12 kg were used in this study. All animal experiments were performed in compliance with the guidelines set forth by the National Institutes of Health, as reviewed and approved by the Wake Forest University Institutional Animal Care and Use Committee. Neuronal recordings were obtained from the dorsolateral and ventrolateral prefrontal cortex of the monkeys before and after training in a working memory task. Detailed experimental procedures have been described previously (Meyer et al., 2011; Qi et al., 2011) and are only presented in brief here. Data analysis was performed using the MATLAB computational environment (MathWorks).

Behavioral task. Neuronal data were compared at two stages. In the first stage (before working memory training), the monkeys were only required to maintain fixation while visual stimuli were presented on a computer screen (Fig. 1A). The first stimulus appeared for 500 ms and was followed by a "delay period" that lasted for 1.5 s. After the delay period, a second white square appeared at either the same location or a diametric location for 500 ms. This was followed by a second 1.5 s delay period. Stimulus presentations were randomized in terms of the spatial location of the cue and whether a match or non-match stimulus appeared in each trial. The stimuli were 2° white squares that appeared randomly in one of nine locations arranged on a 3 × 3 grid; the distance between adjacent stimulus locations was 10° (Fig. 1B). For the analysis presented here, the foveal location was ignored because it never appeared as a non-match stimulus (and it is not shown in Fig. 1B). Therefore, all comparisons involved eight spatial locations. The numbers of match and non-match trials were balanced in this task, as were the presentations of each spatial location; a correct trial for each trial type and stimulus combination needed to be completed successfully before the next block of trials was presented. Typically, 10 trials for each S1–S2 combination were collected (10 match and 10 non-match trials after presentation of the first stimulus at each spatial location).

After recordings were obtained in this stage, the animals were trained to perform a spatial working memory task. The stimuli and timing of presentation were identical before and after working memory training for two of the three animals, allowing us to compare neuronal properties between stages. A third animal was trained in a variation of the working memory task always involving presentation of matching stimuli, and data from this animal were not used here. Average behavioral performance in the working memory task after training was 89%. The percentage of trials aborted as a result of fixation breaks (including blinks) was 15 and 23% before and after training, respectively.

Monkeys were additionally trained in other tasks, including a task requiring feature working memory, and a task relying on a combination of features and spatial locations. For the analysis presented here, we relied exclusively on the stimuli of the spatial set because they were parametric and easily distinguishable from each other even before training. All stimuli were presented using in-house software (Meyer and Constantinidis, 2005), making use of the Psychophysics Toolbox (Brainard, 1997).

Neurophysiology. Neuronal recordings were obtained from areas 8a, 9, and 46 of the dorsolateral prefrontal cortex and areas 12 and 45 of the ventrolateral prefrontal cortex, determined by means of anatomical MR imaging (Meyer et al., 2011). A 20 mm recording cylinder was implanted over the prefrontal cortex, and neuronal recordings were obtained through arrays of two to eight epoxylite- or glass-coated tungsten micro-
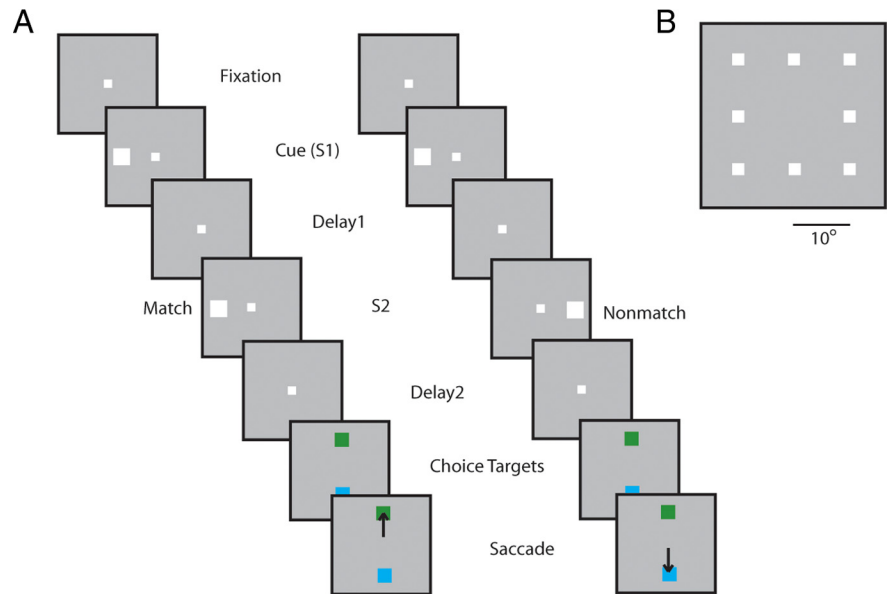

Figure 1. Schematic illustration of the task. *A*, Successive frames illustrate the sequence of stimulus presentations. Two stimuli were presented in sequence (S1 and S2), at either the same or diametric locations, followed by delay periods. After training, the second delay period was followed by two choice targets, and the monkeys were required to saccade to a green target if the two stimuli appeared at matching locations and to a blue target if they did not. In the pretraining stage, the animals were rewarded for maintaining fixation after the end of the second delay period. *B*, Stimulus set used for the analysis presented here.

electrodes spaced 0.2–1.5 mm apart. These were advanced into the cortex daily with a microdrive system (EPS drive; Alpha Omega Engineering). The electrical signal was bandpass filtered between 500 Hz and 8 kHz, and spike waveforms that exceeded a user-defined threshold were sampled at 25 μs resolution with a modular data acquisition system, digitized, and sorted offline (APM system; FHC).

Analysis of neural data. Action-potential waveforms were recorded from two to eight microelectrodes in the lateral prefrontal cortex and sorted into separate units with an automated cluster analysis method based on the KlustaKwik algorithm (Harris et al., 2000). Firing rate was determined in windows identical to the stimulus presentations (aligned to stimulus onset) and the entire delay periods, and then compared between conditions using *t* tests and ANOVA. Units with responses during the stimulus presentations or the delay periods that followed them were identified, evidenced by significantly elevated or decreased firing rate compared with the 1 s interval of fixation (paired *t* test, $p < 0.05$). Neurons with preference to a match or non-match stimulus were identified as those exhibiting a significant difference in mean firing rate between the match and non-match conditions for S2 stimuli presented at the eight spatial locations (two-way ANOVA, $p < 0.05$). In each case, the comparison involved responses to the same stimulus preceded by either a cue stimulus at the same location (therefore constituting a match) or a cue stimulus at a different location (therefore being a non-match). Firing rate of individual neurons was fitted to Gaussian curves of the form $r(x) = B + A \times \exp[-(x - \mu)^2/2\sigma^2]$, where *r* represents mean firing rate for either match or non-match responses at each location *x*, *B* the baseline, *A* the amplitude, μ the peak, and σ the SD of the Gaussian. We distinguished between neurons that had an overall higher response for the match and those that had a response for the non-match. We also identified neurons with a significant interaction between the two factors of the ANOVA, match/non-match status, and spatial location. Population peristimulus time histograms (PSTHs) were constructed by averaging responses from neurons with match or non-match preference together.

Analysis of error trials was performed for a subset of neurons for which both correct and incorrect trials were available. Only error trials involving incorrect selection of one of the two choice targets were processed for this analysis; trials that were aborted as a result of breaks in fixation were ignored. Additionally, we required at least three error trials in each of the match and non-match conditions, collected in presentations of the stim-

ulus at locations that elicited above average responses (four best locations of each neuron). To avoid a bias from neurons recorded in sessions in which more error trials were performed, we first averaged responses from all trials available for each neuron and then averaged across neurons.

A receiver operating characteristic (ROC) analysis was performed by comparing the distribution of firing rates elicited by the identical stimulus when it appeared as a match and when it appeared as a non-match; the ROC analysis represents the ability of an ideal observer to distinguish between these two stimuli based on the distribution of firing rates (Tolhurst et al., 1983). ROC values from the population of neurons with a match or non-match stimulus preference were averaged together. The ROC analysis was also performed in a time-resolved manner, computed in 100 ms bins, stepped by 50 ms windows.

A multivariate regression analysis was used to determine dependence of firing rate on the location of the first stimulus (s_1) and the second stimulus (s_2). For this analysis, we used the eight peripheral stimulus locations and identified the location that elicited the peak firing rate (as in the study by Meyer et al., 2011). We represented s_1 and s_2 as the distance from this location, in units of stimulus separation (corresponding to the unsigned, radial distance of each stimulus from the location that elicited the peak response). The process essentially collapses the two sides of the tuning curve around the peak. We then performed a regression analysis of firing rate r during an interval as a function of the s_1 and s_2 variables:

$$r = a_0 + a_1 s_1 + a_2 s_2 + \varepsilon.$$

The regression analysis was performed using the firing rate during the entire cue (S1) period (aligned on stimulus onset), the match or non-match (S2) period, as well as the delay periods that followed them.

Results
Database
We analyzed responses from three monkeys before and two monkeys after training in a task that required them to observe two stimuli presented in sequence and to make a judgment about whether or not they appeared at the same location (Fig. 1). Before training, the identical stimuli were presented passively, when the monkeys were just required to maintain fixation. The dataset analyzed here consisted of 607 lateral prefrontal neurons that were active during any task epoch before training (representing 47% of the total number of neurons sampled) and 661 neurons active during the task (representing 63% of the total number of neurons recorded after training). Of those, 366, 48, and 193 were recorded from each of the three monkeys, respectively, before training, and 484 and 177 were recorded from the two monkeys, after training. Our analysis sought first to confirm that neural correlates of decision variables were present in neuronal activity of trained animals and then went back in time to investigate how the same variables were represented in neuronal activity in naive animals, before learning the task.

Match/non-match responses
After training, and similar to previous studies that have identified neural correlates of judgments and choices in prefrontal neuronal activity, we observed neurons that responded differently to the same stimulus depending on whether it appeared as match or non-match. An example of a neuron that responded at overall higher levels for non-match stimuli is shown in Figure 2*A–C*; a neuron the responded better to match stimuli is shown in Figure 2*D–F*. We evaluated the influence of spatial location and match/non-match status on neuronal responses with a two-way ANOVA, testing differences between the two match/non-match conditions and the eight spatial locations of the S2 stimulus. Modulation of the neuronal responses by the match/non-match factor took two forms. A total of 139 of 661 (21%) of the neurons

that were active during any task epoch exhibited a significant main effect of the match/non-match factor in either the stimulus presentation period or the delay period that followed it (two-way ANOVA, $p < 0.05$). A total of 154 of 661 neurons (23%) exhibited a significant interaction between spatial location and the match/non-match effect. This included 41 of 661 neurons (6%) that exhibited both a main effect and an interaction, resulting in a total of 252 of 661 (38%) neurons exhibiting either type of modulation. The interaction involved either modified gain of neuronal responses (Fig. 2*F*) or an altered spatial preference for stimuli when they appeared as match versus non-match (Fig. 2*G–I*). Among the neurons that exhibited a significant main effect of match/non-match during the stimulus presentation, there was an overall bias toward non-match responses in this task; 32% of the neurons responded with a higher response to a stimulus when it appeared as a match, and 68% of the neurons responded with a higher response to a stimulus when it appeared as a non-match.

Having identified neural correlates representing the judgment of whether the two stimuli appeared at the same location or not, we tested whether neurons in the same animals discriminated between match and non-match stimuli before training in the task that required them to make such a decision (and therefore had no behavioral significance). We found that, even before any training, a population of lateral prefrontal neurons were modulated by whether a stimulus was preceded by the same stimulus (and therefore would constitute a match) or not (and therefore would constitute a non-match). Using a two-way ANOVA with the spatial location of the second stimulus and its status of match or non-match as factors, we found that 69 of 607 (11%) of neurons exhibited a significant main effect of match/non-match and 79 of 607 (13%) an interaction between match/non-match and spatial location (two-way ANOVA, $p < 0.05$). Fourteen neurons (2%) exhibited both a main effect and an interaction, resulting in 134 of 607 neurons (22%) with either type of modulation. We observed neurons with a significant match/non-match main effect in all three monkeys before training (representing 14, 7, and 10% of activated neurons, respectively). An example neuron with a match/non-match effect recorded before training is shown in Figure 3. Because the percentage of neurons that exhibited effects was relatively small and involved comparisons in both the cue and delay period (in one of which some neurons might not be active), we wanted to ensure that this finding was not the result of spurious variation of neuronal responses. We therefore repeated the analysis for neurons that were active during the stimulus presentation ($n = 315$). To avoid a bias toward enhanced match or non-match responses, we selected these neurons by identifying significant increases in firing rate during the stimulus presentation in the cue (S1) period compared with the fixation period. We then used this sample of neurons to compare their responses to match and non-match stimuli in the S2 period. A total of 41 (13%) of the neurons exhibited a significant main effect of match/non-match (two-way ANOVA, $p < 0.05$); this was significantly higher than the expected false-positive error of the ANOVA test (χ^2 test, $p < 10^{-5}$). The equivalent percentage of neurons with a match/non-match effect active during the stimulus presentation after training was 94 of 369 (25%). A higher than expected percentage of neurons was also observed if we relied on a more conservative significance level for the ANOVA test: 3% of the neurons recorded before training exceeded the $\alpha = 0.005$ level. The result confirmed that a small but statistically significant population of neurons represented decision variables before training in the task.
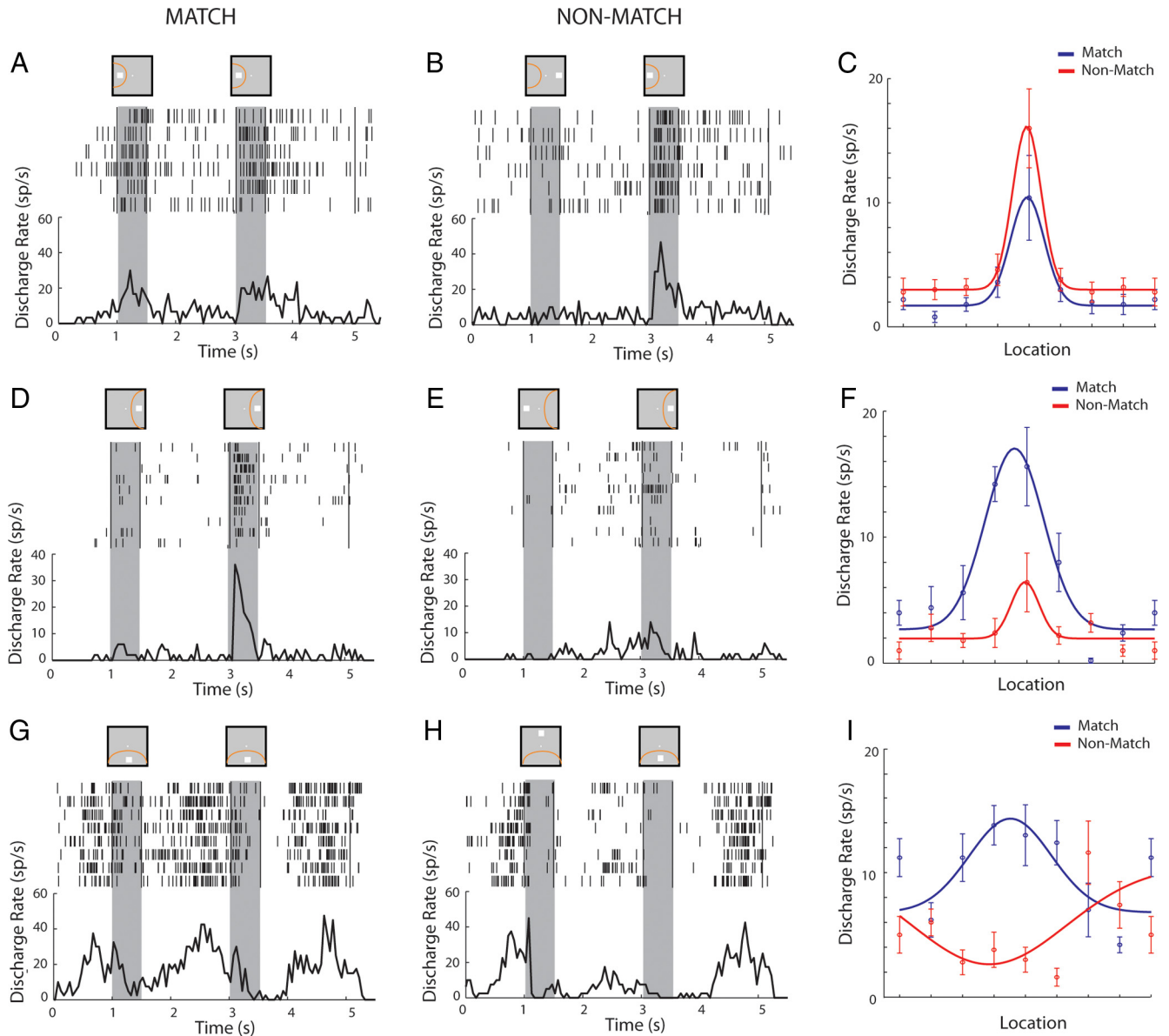
Figure 2. Example neurons with preference for a match or non-match stimulus. **A**, Rasters and PSTH representing responses of one lateral prefrontal neuron to a S2 stimulus (match) in the receptive field (yellow arc in the inset above the PSTH), preceded by a cue at the same location. Gray bars indicate times of stimulus presentations. **B**, Responses of the same neuron to the same S2 stimulus as a non-match. **C**, Firing rates of the same neuron for match and non-match responses appearing at each location fitted to a Gaussian curve. **D–F**, Responses of a second neuron with preference for non-match stimuli. **G–I**, Responses of a third neuron with different preference for spatial location depending on match/non-match status.
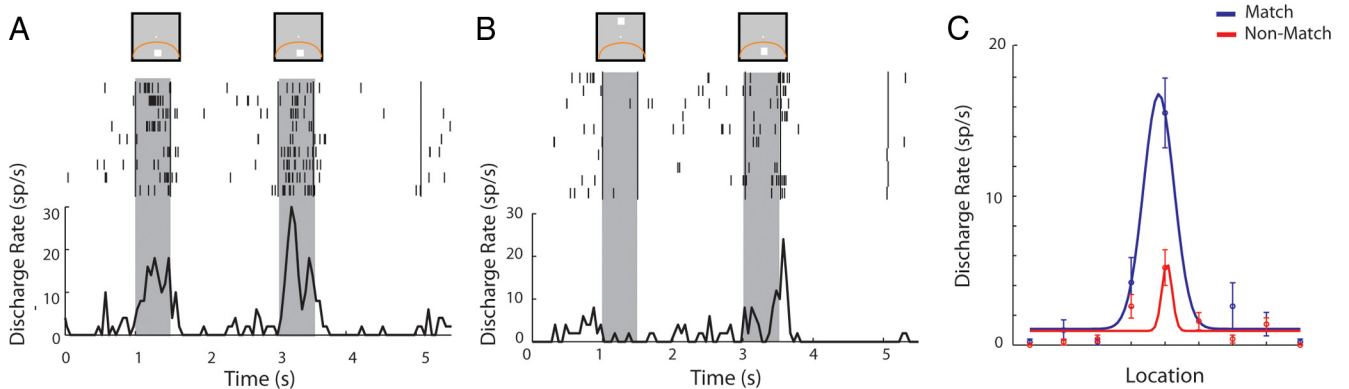

Figure 3. Example neuron with preference for a match stimulus, **A**, over a non-match stimululs, **B**, recorded before training. **C**, Gaussian fit. Conventions are the same as in Figure 2.
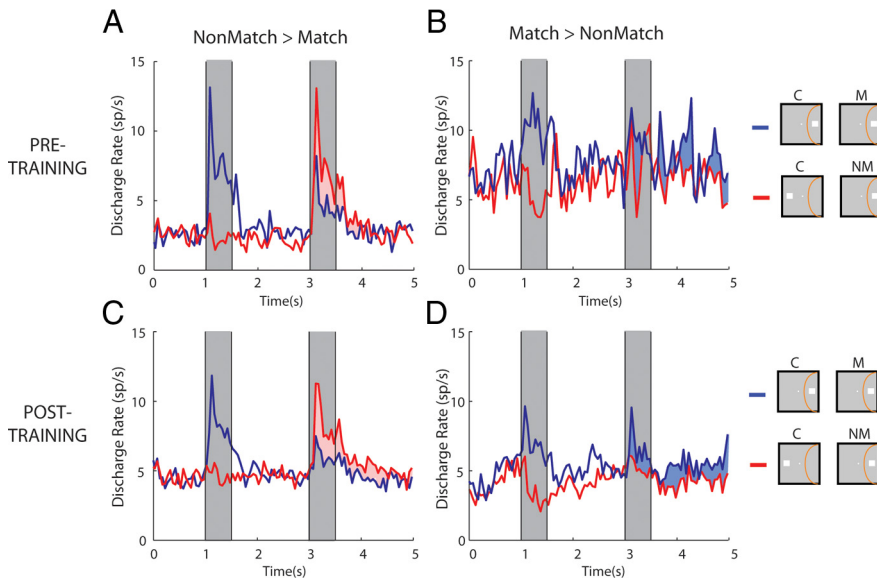
Figure 4. Magnitude of match/non-match preference before and after training. *A*, Population PSTH recorded before training, of neurons with significant preference for non-match over match stimuli. Blue line represents the match (non-preferred condition), and red the non-match (preferred), always presented at the best location in the receptive field. Gray bars represent times of stimulus presentation. *B*, Population PSTH of neurons with significant preference of non-match over match stimuli. Now the blue line representing the match is the preferred condition and red the non-preferred. *C, D*, Equivalent populations of neurons recorded after training in the task. Insets to the right of the figure are schematic; the location of the receptive field differed between neurons, and data were averaged from the best location of each neuron. C, Cue; M, match; NM, non-match.
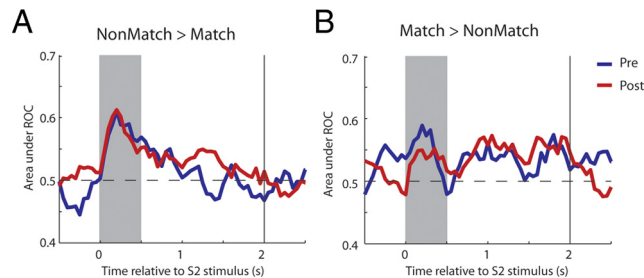

Figure 5. ROC analysis. *A*, Average ROC values comparing the distribution of responses to match and non-match stimuli presented for neurons with preference of non-match over match responses, recorded before (blue line) and after (red line) training. *B*, Average ROC values for neurons with preference of match over non-match stimuli.

Previous studies in several brain areas and using a variety of methodologies have indicated that stimuli presented repeatedly produce decreased responses, a phenomenon referred to as repetition suppression (Grill-Spector et al., 2006). We therefore wanted to test whether the match/non-match effect we observed was another demonstration of this phenomenon. Contrary to this prediction, equal percentages (50%) of neurons that had a significant preference for match or non-match stimulus preferred the match over the non-match and vice versa, although overall firing rate was higher for the non-match stimuli.

In terms of anatomical specialization, a higher percentage of neurons exhibiting a significant match/non-match main effect was observed in the dorsolateral than the ventrolateral prefrontal cortex (14 vs 7%). We previously documented a series of functional differences in terms of stimulus selectivity and responsiveness between the dorsolateral and ventrolateral prefrontal cortex, both before and after training (Meyer et al., 2011). However, an equivalent difference between areas was observed after training (25 vs 12%), and results from dorsal and ventral prefrontal cortex

were pooled together here, for the pretraining versus posttraining comparisons.

Magnitude of decision variables before and after training

It is possible that, before training, neurons signaled the difference between match and non-match only marginally, a difference in neuronal activity that would provide an inadequate basis for a judgment. To test this possibility, we sought to quantify how the discharge rate of match/non-match-sensitive neurons compared with that of neurons representing the equivalent difference after training. Among neurons that exhibited a significant match/non-match effect, the average discharge rate difference was comparable before and after training (Fig. 4). We split neurons into those that exhibited an overall preference of the non-match (Fig. 4*A, C*) and those preferring the match (Fig. 4*B, D*) stimulus and computed population discharge rates before and after training. The average absolute difference of match and non-match responses (Fig. 4, shaded areas) was 28% before and 32% after training, which was not significantly different (t test, $p > 0.2$). Preference for the match/non-match was essentially present from the first spike of the stimulus response and continued into the delay period after the stimulus presentation, both before and after training (Fig. 4*A, C*).

The comparable difference in firing rate we observed before and after training also resulted in approximately equal levels of discriminability, judged with an ROC analysis. In this context, the area under the ROC curve represents the ability of an ideal observer to determine whether a stimulus is a match or a non-match based on neuronal discharges. For this analysis, data were pooled from neurons with either a match or non-match preference, and the sign of the ROC curve around the chance level (0.5) was reserved for the former population before averaging. A time-resolved ROC analysis for neurons with a significant match/non-match effect illustrated that comparable values of discriminability were achieved before and after training and with very similar time courses (Fig. 5*A, B*). Mean ROC values across this population of neurons were not significantly different when computed for the entire stimulus presentation period (t test, $p > 0.7$) or delay period (t test, $p > 0.1$).

Although these differences in firing rate and discriminability indicate that an ideal observer can determine whether a stimulus constitutes a match or a non-match, they do not necessarily mean that, after training, the monkey bases its decision on this activity. For this reason, we compared firing rate differences between match and non-match responses in error trials after training. Analysis was limited to only those errors in which the monkey chose one of the two choice targets (trials that were aborted as a result of breaks in fixation were ignored). We examined the firing rate of neurons with significant match/non-match effects for which correct and error trials were available, from both the match and non-match conditions ($n = 20$). As in the analysis for Figures 4 and 5, we reversed the sign of the comparison for neurons that had match and non-match preferences. When we compared this firing rate difference in correct and error trials (Fig. 6*A, B*), we
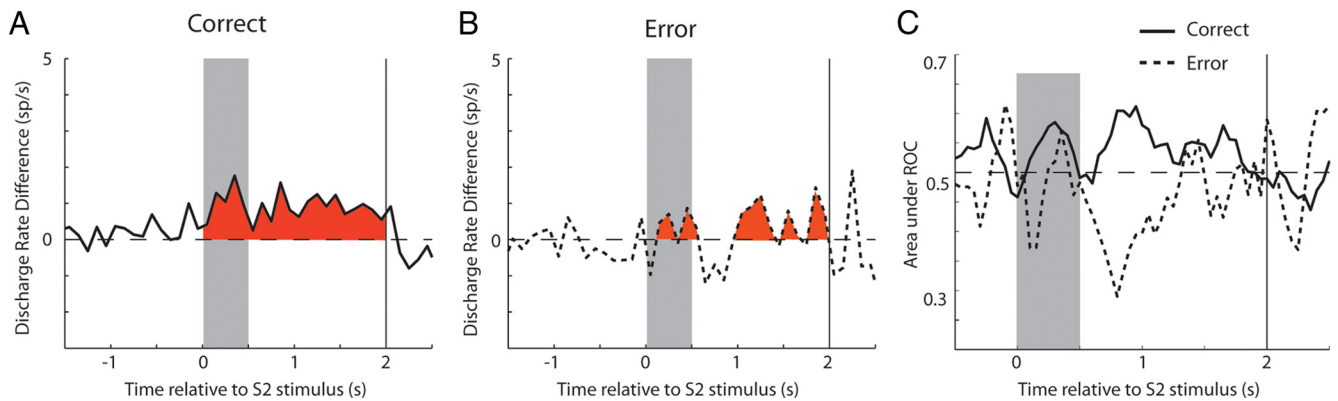
Figure 6. Correct and error trials. Average absolute difference in firing rate between match and non-match responses, for neurons with a significant difference between the two stimuli (shaded area in Fig. 3). **A**, Average from correct trials. **B**, Average from error trials. The same population of neurons with sufficient number of error trials ($n = 20$) is compared in both panels. **C**, Average ROC values are compared from correct and error trials.

found a significant decrease in error trials (t test, $p < 0.05$), in which the monkey erroneously chose the blue choice target for two matching stimuli, or the green choice target for non-match stimuli. Similarly, a comparison of ROC values comparing the responses to match and non-match stimuli obtained in correct and error trials indicated that the value of the ROC around baseline (chance levels) had the opposite sign for most of the stimulus presentation and delay period (Fig. 6C). The difference in average ROC values computed over the entire task epoch reached statistical significance for the delay period after the S2 stimulus (paired t test, $p < 0.05$). The result indicates that the activity distinguishing between match and non-match responses had predictive power of the monkey's decision after training.

Representation of cue information in the match/non-match period

Previous studies have demonstrated that prefrontal neurons represent information about the properties of the cue (S1) during and after the second stimulus presentation, even when just a comparison of the two is required by the task rather than a report of the first stimulus properties (Romo et al., 1999; Romo and Salinas, 2003). We therefore investigated how this concurrent representation of the two stimuli is achieved in the lateral prefrontal cortex before and after training. For this question, we performed a multiple linear regression analysis, analogous to that used in previous studies of neural correlates of perceptual decisions (Hernández et al., 2002). Firing rate in each task period was represented as a linear sum of variables representing the location of the first (cue) and second stimulus (match or non-match):

$$r = a_0 + a_1 s_1 + a_2 s_2 + \varepsilon.$$

We represented location in our analysis as distance from the best cue (S1) location (the location variable was rotated for each neuron before regression based on the cue location that produced the best response; see Materials and Methods). We then identified neurons with coefficients significantly different from zero for the first (S1) and second (S2) stimulus.

A neuron with a significant coefficient for S1 would indicate that it is modulated by the location of the cue. This is the pattern observed for a population of neurons during the interval of cue presentation when 53% of neurons recorded before training and 36% neurons recorded after training exhibited significant s_1 coefficients (Fig. 7A,C, blue dots). Similarly, during the time of the second stimulus presentation, 41 and 24% of pretraining and

posttraining neurons represented the location of the second stimulus, evidenced by significant s_2 coefficients (Fig. 7B,D, purple dots). The critical comparisons involved whether lateral prefrontal neurons continued to represent information about the cue (S1) location during the S2 interval and whether the incidence of these neurons increased after training. Indeed, during the second stimulus interval after training, we observed 14% of neurons whose firing rate continued to be significantly modulated by the location of the cue (S1) stimulus compared with 6% of neurons before training in either the sample (Fig. 7D) or delay period (data not shown). These included 7% of the neurons after training and 3% before training with significant coefficients for both S1 and S2 (Fig. 7C,D, yellow points). We should note that this analysis is much more stringent than the simple match/non-match comparison performed above because it reports significant coefficients for only those neurons with a linear modulation of firing rate as a function of the location variable. The increase in the percentage of neurons modulated by s_1 coefficients after training compared with before training was statistically significant (χ^2 test, $p < 0.001$). Points falling to the left of the vertical axis in Figure 7 correspond to neurons with better non-match than match responses at their best location: a negative s_2 coefficient indicates lower response to an S2 stimulus appearing at the best location when it is preceded by the cue (S1) at the same location and therefore constitutes a match. Points falling to the right of the vertical axis similarly correspond to neurons with match preference at their best location. A special case involves points falling in the $s_2 = -s_1$ line (Fig. 7B,D, dotted line). The output of these neurons is significant in a task that required subjects to judge whether the first stimulus frequency was higher than the second one (Romo et al., 1999; Romo and Salinas, 2003). In the context of our task, these are neurons for which non-match preference is absolute, in that the sum influence of the first and second stimulus is zero, and therefore the neuron responds with baseline firing rate for any match stimulus. Neurons approximating this property (values of coefficients within 50% of each other; falling within the dotted lines in Fig. 7B,D) made up 31% of the neurons with significant s_1 coefficients after training compared with 33% before training.

Discussion

Our results demonstrate that a population of lateral prefrontal neurons signals whether or not two stimuli match even before monkeys have been trained to make such a judgment. In the
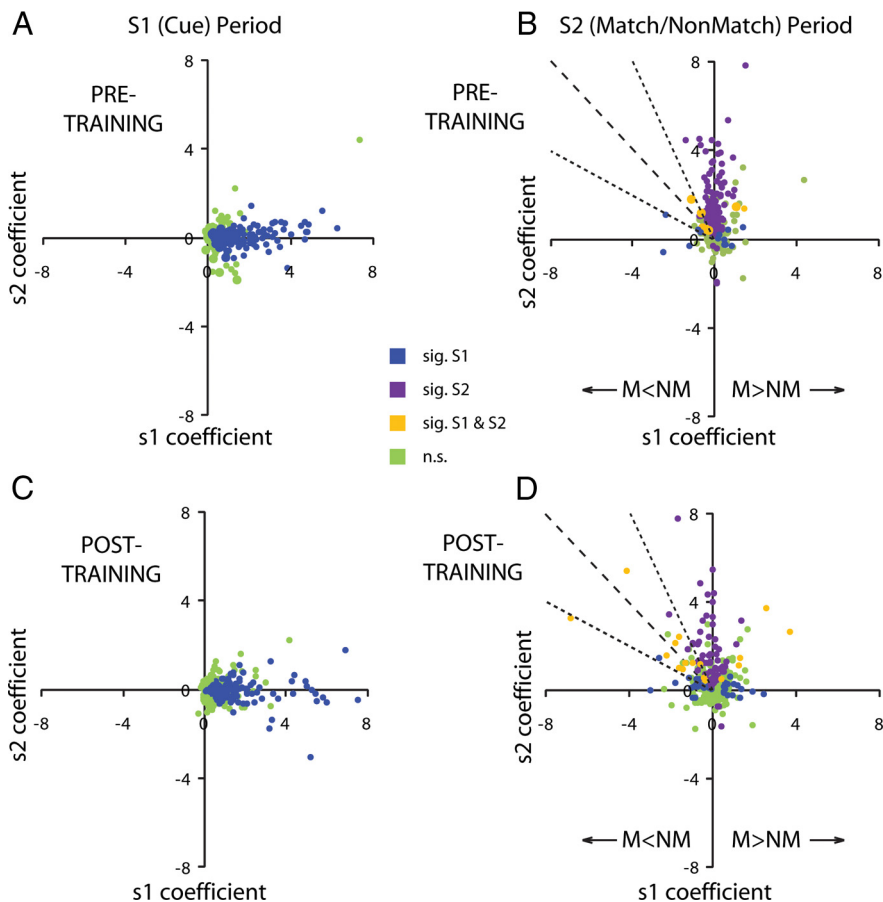
Figure 7. Regression analysis. **A**, Responses recorded during the cue period, before training. Each dot represents the s_1 and s_2 regression coefficients of a single neuron. Blue dots represent neurons with significant s_1 coefficients (i.e., significantly modulated by the location of the cue); green dots represent neurons with no significant s_1 coefficients. **B**, Responses recorded during the second stimulus presentation period. In addition to blue and green dots, purple dots represent neurons with significant s_2 coefficients (i.e., significantly modulated by the location of the second stimulus). Orange dots represent neurons with both significant s_1 and s_2 coefficients. **C**, **D**, Equivalent responses of neurons recorded after training. Up to three outliers appeared beyond the axis range in each panel. M, Match; NM, non-match.

context of task performance after training, similar activity was predictive of the impending match/non-match judgment. Our findings suggest that prefrontal cortex automatically encodes some stimulus relationships even when there is no explicit requirement for comparison, providing relational information that may be drawn on to guide goal-directed decision-making.

Match/non-match responses

Tasks requiring a judgment about stimuli presented in sequence have been used extensively in studies of the prefrontal cortex (Miller et al., 1996; Romo et al., 1999; Freedman et al., 2003; Pasternak and Zaksas, 2003; Romo and Salinas, 2003; Kusunoki et al., 2009) and areas that project to it (Miller et al., 1991; Steinmetz et al., 1994; Freedman and Assad, 2006; Woloszyn and Sheinberg, 2009). For example, inferior temporal and posterior parietal neurons respond differently to match and non-match stimuli in the context of delayed-match to sample tasks (Miller et al., 1991, 1993; Steinmetz et al., 1994; Steinmetz and Constantinidis, 1995; Woloszyn and Sheinberg, 2009). The overall preference for a match or non-match stimulus across the population of neurons can be reversed depending on the specific task that monkeys are trained to perform (Miller and Desimone, 1994; Rawley and Constantinidis, 2010), suggesting that this differential activity is modulated by the particular judgment imposed by the task. The

magnitude of response difference to match and non-match stimuli differs systematically in correct and error trials, suggesting that this activity guides the subject's decision (Zaksas and Pasternak, 2006). Our present results in trained animals replicated many of these findings. We found that a population of lateral prefrontal neurons responded differently to a stimulus depending on whether it appeared as a match or a non-match (which can also be viewed as suppression of match stimuli). Firing rate difference between match and non-match stimuli discriminated between the two conditions on correct trials but diminished or reversed on error trials. Additionally, we demonstrate for the first time that neurons with a significant preference for a match or a non-match stimulus are present before any specific training requiring monkeys to report a judgment about the sequential stimuli. The magnitude of the firing rate difference and area under the ROC curve for match and non-match responses among neurons exhibiting the effect were comparable before and after training. The finding reveals that lateral prefrontal neurons automatically represent some stimulus comparisons, independent of any requirement to use the information to achieve a behavioral objective.

We have shown previously that, in naive monkeys, a small but significant population of prefrontal neurons manifests persistent activity for stimuli viewed passively in the absence of any mnemonic requirement (Meyer et al., 2007). Analogous to our present findings, this persistent activity exhibited many of the same characteristics (e.g., spatial and feature selectivity) as activity recorded from animals actively engaged in working memory tasks. Changes in persistent activity observed after training were quantitative rather than qualitative, involving a higher percentage of neurons exhibiting delay period activity and with a higher mean firing rate (Meyer et al., 2011; Qi et al., 2011).

Interpreting differences in activity across sequential stimulus presentation requires consideration of the phenomenon of repetition suppression. This refers to the reduction in activity typically observed for the repeated presentation of a stimulus. Repetition suppression has been observed in multiple brain areas and using various methodologies, including single-neuron recordings, fMRI, and EEG/MEG (Grill-Spector et al., 2006). It remains a matter of debate whether repetition suppression is a result of local mechanisms acting at the level of the neurons excited by the stimulus, such as fatigue or sharpening of neuronal responses (Grill-Spector and Malach, 2001; Kaliukhovich and Vogels, 2011), or an effect of top-down influences (Friston, 2005; Summerfield et al., 2008; Engel and Wang, 2011). The effect we observed in naive animals may appear to be a case of repetition suppression because there was an overall higher response to non-match than to match stimuli (Fig. 4A,B). However, we observed equal percentages of neurons with match and non-match preference and an equally large population of neurons with significant

interaction between stimulus location and match/non-match status (including neurons that exhibited a different best location when a stimulus appeared as a match or a non-match), findings that cannot be explained by suppression of the repeated stimulus. If anything, weaker responses to match stimuli were observed among a greater percentage of prefrontal neurons after training, when the monkeys were performing the task. Prefrontal cortex is generally assumed to be the ultimate source of top-down signals in repetition suppression (Grill-Spector et al., 2006). Our current results suggest that prefrontal neurons exhibit a preference to non-match stimuli to a greater extent after learning to perform a task. However, if this differential activity constitutes a top-down signal to other areas, then this is present even before training.

In terms of anatomical specialization, we have reported previously that dorsolateral prefrontal neurons are more responsive and more selective for spatial stimuli than are ventrolateral prefrontal neurons; this was the case both before and after training (Meyer et al., 2011). We now report that a greater percentage of dorsolateral prefrontal neurons differentiated between match and non-match spatial stimuli. The dorsolateral bias was present before training and persisted as the effects of training were comparable for the two areas.

Parametric representation of stimulus location

We adopted an approach used in a series of studies conducted by Romo and colleagues to study the representation of a parametric variable (frequency of vibration of a tactile stimulus) when animals were required to judge the relative frequency of two stimuli. Our task is most similar to the "postponed decision" version of their paradigm, which imposes a delay period after the presentation of the second tactile stimulus (Lemus et al., 2007; Martínez-García et al., 2011). Romo and colleagues identified prefrontal neurons that were modulated by the frequency of the first stimulus during the second stimulus presentation (Romo et al., 1999; Romo and Salinas, 2003). We used a similar regression analysis method to identify neurons having activity that was modulated by the location of the first stimulus and found that the percentage of neurons exhibiting this quality was relatively small after training (14%), representing a quantitative rather than qualitative change in the nature of representation observed before training. In the vibratory task, a large population of neurons also represented the difference between the frequency of the two stimuli. Our task did not require a signed comparison for non-match stimuli and an equivalent variable (the difference between the s_1 and s_2 coefficients) was represented in a small population of neurons both before and after training in our task (which made up approximately equal proportions of the neurons modulated by the properties of the cue). It is possible to train subjects to perform judgments of all sorts, and some of these may rely on activity that only emerges after training; our results indicate that at least some comparisons are naturally represented in the neuronal activity of untrained animals.

Effects of training

Although our discussion emphasized the similarity in neuronal activity before and after training, differences were also observed. Previously, we reported that after training in a spatial and a feature working memory paradigm, task performance evokes activity in a greater percentage of neurons and a greater overall firing rate when compared with passive viewing of identical stimuli. Somewhat analogous findings were reported by Law and Gold (2008), who demonstrated that, over the course of training on a motion discrimination task, the activity or neurons in parietal

area LIP (lateral intraparietal area) increased. Paradoxically, because of a disproportionate increase in baseline activity compared with the peak responses, average selectivity for stimuli declined after training in our task (Meyer et al., 2011; Qi et al., 2011). Our present analysis indicated that, after training, a greater percentage of neurons show significant differences in activity between match and non-match stimuli, an overall greater percentage of these neurons prefer the non-match stimuli, and a higher percentage of neurons are modulated in a parametric manner by the first stimulus location during the presentation of the second stimulus. Changes after training were even more pronounced for feature stimuli, which were not as discriminable as the spatial stimuli at the outset of training (data not shown). We should note, however, that some of the changes we observed may be attributable to the effect of the task execution, which imposes greater demands for attention during the task, in addition to requiring the encoding of task demands, also known to modulate neuronal responses to stimuli (Asaad et al., 1998; White and Wise, 1999; Wallis et al., 2001). If it were possible to equalize attention and task representation factors across pretraining and posttraining phases, we might have found an even greater similarity in neural correlates of spatial judgments before and after training.

References

Asaad WF, Rainer G, Miller EK (1998) Neural activity in the primate prefrontal cortex during associative learning. Neuron 21:1399–1407.

Barraclough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci 7:404–410.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Engel TA, Wang XJ (2011) Same or different? A neural circuit mechanism of similarity-based pattern match decision making. J Neurosci 31:6982–6996.

Freedman DJ, Assad JA (2006) Experience-dependent representation of visual categories in parietal cortex. Nature 443:85–88.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci 23:5235–5246.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836.

Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst) 107:293–321.

Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14–23.

Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsáki G (2000) Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J Neurophysiol 84:401–414.

Hernández A, Zainos A, Romo R (2002) Temporal evolution of a decision-making process in medial premotor cortex. Neuron 33:959–972.

Kaliukhovich DA, Vogels R (2011) Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. Cereb Cortex 21:1547–1558.

Kim JN, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci 2:176–185.

Kubota K, Komatsu H (1985) Neuron activities of monkey prefrontal cortex during the learning of visual discrimination tasks with GO/NO-GO performances. Neurosci Res 3:106–129.

Kusunoki M, Sigala N, Gaffan D, Duncan J (2009) Detection of fixed and variable targets in the monkey prefrontal cortex. Cereb Cortex 19:2522–2534.

Law CT, Gold JI (2008) Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. Nat Neurosci 11:505–513.

Lemus L, Hernández A, Luna R, Zainos A, Nácher V, Romo R (2007) Neural correlates of a postponed decision report. Proc Natl Acad Sci USA 104:17174–17179.

Martínez-García M, Rolls ET, Deco G, Romo R (2011) Neural and computational mechanisms of postponed decisions. Proc Natl Acad Sci USA 108:11626–11631.

Matsumoto K, Suzuki W, Tanaka K (2003) Neuronal correlates of goal-based motor selection in the prefrontal cortex. Science 301:229–232.

Meyer T, Constantinidis C (2005) A software solution for the control of visual behavioral experimentation. J Neurosci Methods 142:27–34.

Meyer T, Qi XL, Constantinidis C (2007) Persistent discharges in the prefrontal cortex of monkeys naive to working memory tasks. Cereb Cortex 17 [Suppl 1]:i70–i76.

Meyer T, Qi XL, Stanford TR, Constantinidis C (2011) Stimulus selectivity in dorsal and ventral prefrontal cortex after training in working memory tasks. J Neurosci 31:6266–6276.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

Miller EK, Desimone R (1994) Parallel neuronal mechanisms for short-term memory. Science 263:520–522.

Miller EK, Li L, Desimone R (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. Science 254:1377–1379.

Miller EK, Li L, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J Neurosci 13:1460–1478.

Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci 16:5154–5167.

Pasternak T, Zaksas D (2003) Stimulus specificity and temporal dynamics of working memory for visual motion. J Neurophysiol 90:2757–2762.

Qi XL, Meyer T, Stanford TR, Constantinidis C (2011) Changes in prefrontal neuronal activity after learning to perform a spatial working memory task. Cereb Cortex 21:2722–2732.

Rainer G, Miller EK (2000) Effects of visual experience on the representation of objects in the prefrontal cortex. Neuron 27:179–189.

Rawley JB, Constantinidis C (2010) Effects of task and coordinate frame of attention in area 7a of the primate posterior parietal cortex. J Vis 10:1–16.

Romo R, Salinas E (2003) Flutter discrimination: neural codes, perception, memory and decision making. Nat Rev Neurosci 4:203–218.

Romo R, Brody CD, Hernández A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. Nature 399:470–473.

Sandrew BB, Stamm JS, Rosen SC (1977) Steady potential shifts and facilitated learning of delayed response in monkeys. Exp Neurol 55:43–55.

Steinmetz MA, Constantinidis C (1995) Neurophysiological evidence for a role of posterior parietal cortex in redirecting visual attention. Cereb Cortex 5:448–456.

Steinmetz MA, Connor CE, Constantinidis C, McLaughlin JR (1994) Covert attention suppresses neuronal responses in area 7a of the posterior parietal cortex. J Neurophysiol 72:1020–1023.

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11:1004–1006.

Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res 23:775–785.

Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. Nature 411:953–956.

White IM, Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. Exp Brain Res 126:315–335.

Woloszyn L, Sheinberg DL (2009) Neural dynamics in inferior temporal cortex during a visual working memory task. J Neurosci 29:5494–5507.

Zaksas D, 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.