

# Prefrontal Cell Activities Related to Monkeys' Success and Failure in Adapting to Rule Changes in a Wisconsin Card Sorting Test Analog

Farshad A. Mansouri, Kenji Matsumoto, and Keiji Tanaka

Cognitive Brain Mapping Laboratory, RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan

The cognitive flexibility to select appropriate rules in a changing environment is essential for survival and is assumed to depend on the integrity of prefrontal cortex (PFC). To explore the contribution of the dorsolateral PFC to flexible rule-based behavior, we recorded the activity of cells in this region of monkeys performing a Wisconsin Card Sorting Test (WCST) analog. The monkey had to match a sample to one of three test items by either color or shape. Liquid reward and a discrete visual signal (error signal) were given as feedback to correct and incorrect target selections, respectively. The relevant rule and its frequent changes were not cued, and the monkeys could find it only by interpreting the feedback. In one-third of cells, cellular activity was modulated by the relevant rule, both throughout the trial and between trials. The magnitude of the modulation correlated with the number of errors that the monkeys committed after each rule change in the course of reestablishing high performance. Activity of other cells differed between correct and error trials independently from the rule-related modulation. This difference appeared during actual responses and before the monkeys faced the problems. Many PFC cells responded to the error-signal presentation, and, in some of them, the magnitude of response depended on the relevant rule. These results suggest that the dorsolateral PFC contributes to WCST performance by maintaining the relevant rule across trials, assessing behavioral outcomes, and monitoring the processes that could lead to success and failure in individual trials.

**Key words:** WCST; short-term memory of rule; error-related activity; preparatory attention; history of success and failure; feedback representation

## Introduction

The Wisconsin Card Sorting Test (WCST) has been extensively used to assess the cognitive flexibility of humans (Berg, 1948; Grant et al., 1949; Heaton, 1993). In the WCST, the sensory dimension by which the subject should match test items with a sample changes frequently. The subject receives no cues about the relevant sensory dimension and its changes and therefore must find this based on the assessment of feedback and maintain it in working memory across trials. Patients with prefrontal cortex (PFC) damage show impaired performance on the WCST and other rule-switching tasks; they tend to apply the previously relevant rule even after it becomes inappropriate (Milner, 1963; Drewe, 1974; Anderson et al., 1991; Stuss et al., 2000; Goldstein et al., 2004). However, cognitive flexibility in rule-switching tasks such as the WCST relies on the interplay of multiple cognitive processes, and impairment in any one of them might lead to a performance deficit.

Neurological and lesion studies suggest an involvement of PFC in rule-switching tasks (Milner, 1963; Drewe, 1974; Anderson et al., 1991; Dias et al., 1996; Stuss et al., 2000; Goldstein et al., 2004). Imaging studies (Berman et al., 1995; Volz et al., 1997;

Mentzel et al., 1998; Konishi et al., 1999; Monchi et al., 2001; Wang et al., 2001; Nakahara et al., 2002) have shown activation of different areas of PFC in rule-switching tasks. Recordings from PFC neurons have shown that their activities might represent working memory of specific items (Niki and Watanabe, 1976; Goldman-Rakic, 1990; Pontecorvo et al., 1996; Rao et al., 1997; Levy and Goldman-Rakic, 1999; Funahashi, 2001) and the relevant rule (Sakagami and Niki, 1994; White and Wise, 1999; Asaad et al., 2000; Wallis et al., 2001). These previous results support the importance of PFC for the control of rule-based behavior. However, it remains unclear how PFC contributes to flexible rule-based behavior. Our knowledge of the neuronal processes underlying the assessment of behavioral outcome and short-term maintenance of appropriate rule and their relation to overall behavior is limited. We also do not know what kinds of neuronal events lead to errors. To explore neural bases of behavioral flexibility in adapting to frequent rule changes, we recorded cell activity from the dorsolateral PFC of monkeys performing a close analog of the WCST.

Because no cue was provided for the relevant rule and its changes in the WCST analog, it was a challenging task for monkeys. Their performance in adapting to rule changes varied among blocks of trials (a block starts with a rule change and ends with the next rule change), which in turn provided an opportunity to study the correlation of neuronal activities with the monkeys' performance. We found two types of modulations in cellu-

Received July 23, 2005; revised Jan. 13, 2006; accepted Jan. 13, 2006.

Correspondence should be addressed to Farshad A. Mansouri, Laboratory for Cognitive Brain Mapping, RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan. E-mail: farshad@postman.riken.go.jp.

DOI:10.1523/JNEUROSCI.5238-05.2006

Copyright © 2006 Society for Neuroscience 0270-6474/06/262745-12\$15.00/0

lar activity that correlated with the monkeys' performance in different timescales. The first type was a block-by-block modulation by the relevant rule, which correlated with the efficiency of monkeys to reestablish a high performance level after a rule change. The second type was a trial-by-trial modulation that correlated with the monkeys' success or failure in individual trials.

## Materials and Methods

**Behavioral task.** Two monkeys (*Macaca fuscata*; one male and one female weighing 6 and 7 kg, respectively) were trained to perform an analog of WCST. Monkeys had access to food pellets *ad libitum* and received the required water (diluted apple juice) while performing the task. They had *ad libitum* access to water for 1.5 d in each week. Additional fresh fruits and water was supplied at the end of recording session. If they scored >300 correct trials in a daily training session, a piece of chocolate was given. The main events in the task are illustrated in supplemental Figure 1a (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). The monkey sat in front of a touch-sensitive liquid crystal display monitor placed 28 cm from their eyes. CORTEX software (<http://www.cortex.salk.edu>) was used to control the task and data acquisition. A start cue (a gray circle of 11.2° of visual angle diameter with a grid pattern inside) appeared when the intertrial interval (ITI) (2400 ms) was over. The monkeys started the trial by pushing a bar (located at the bottom of the monitor) within 10 s after the onset of the start cue. The bar pressing changed the start cue to a fixation point (a white circle of 0.8° in diameter). If the monkeys kept pushing the bar and maintained gaze on the fixation point for 700 ms, a sample stimulus replaced the fixation point. An early release of the bar or fixation break terminated the trial, and a visual error signal (a purple annulus of 22.4° outer diameter and 11.2° inner diameter) was presented. If the monkeys maintained the eye fixation and bar press, three test items were added to the area surrounding the sample 630 ms after the onset of the sample presentation. Each of the sample and test items was 5–7° in size. The center-to-center distance between the test items and sample was 16° (left and right) or 14° (bottom). The monkeys had to touch the test item that matched the sample in color or in shape within 2500 ms from the onset of the test item presentation. A drop of juice was administered 800 ms after a correct target selection, and the sample and test items were extinguished 200 ms after the reward delivery. The visual error signal replaced the sample and test items 700 ms after an incorrect target selection and immediately after a fixation break, and it stayed on the monitor for 500 ms in both cases. Note that the sample remained after the test items appeared in the task. There was thus no requirement to keep the sample in memory. We did not study the effect of feedback omission on neuronal activity, because we found that the omission of feedback had a disruptive effect on the monkeys' performance and even led them to stop performing the task. The ITI, which was defined by the period from the disappearance of the sample and test items or the disappearance of the error signal to the onset of the start cue, was 2400 ms.

When the monkey reached 85% correct performance (17 correct responses in 20 trials), the relevant rule (matching by color or matching by shape) was changed without notice to the monkey. A minimum of 40 trials was provided in each block even when the monkey reached the criterion earlier. The monkeys had to keep their gazes fixed on the fixation point during the fixation and sample presentation periods. We kept the fixation point at the center of the sample. The size of fixation window was 4° ( $\pm 2^\circ$  from the fixation point center), which was smaller than any sample (5–7°). In some occasions when the monkeys committed very few or very high number of fixation errors, the size of the fixation window was adjusted on-line (within  $\pm 2^\circ$ ) to obtain  $\sim 10\%$  fixation breaks because we intended to analyze neuronal activities in fixation-break trials. The size of fixation window was immediately returned to 4° when the number of fixation breaks returned to the usual amount. Fixation-break trials were not included in the calculation of shift criterion. Eye position was monitored by an infrared system (<http://staff.aist.go.jp/k.matsuda/eye/>).

Ten sample stimuli were used in the recording sessions. Nine of them were created by combining three colors (red, green, and blue) with three shapes (square, circle, and cross). The 10th stimulus was a yellow hexa-

gon. A sample was randomly selected from the 10 stimuli without replacement (all 10 stimuli were used before starting another cycle). Three of the stimuli appeared as test items with a fixed arrangement (a red square on the left, a green circle on the right, and a blue cross at the bottom) in both color and shape blocks. When a yellow hexagon was presented as the sample, the blue cross was replaced with a yellow cross and the green circle with a green hexagon in the arrangement of test items. Thus, seven samples were shown under the incongruent condition, in which one test item matched the sample in color and another test item matched the sample in shape, whereas the other three samples were presented under the congruent condition, in which one test item matched the sample in both color and shape.

For the behavioral test with 36 samples, stimuli were created by combining six colors and six shapes. Seven and 29 samples were presented in the congruent and incongruent conditions, respectively. This test was conducted for both monkeys in 20 sessions.

Although the fundamental structure and requirements of a clinical WCST were maintained in our WCST analog for monkeys, the WCST analog differed from the clinical WCST in several ways. (1) The third dimension used in the clinical WCST, numerosity, was omitted, and only color and shape were implemented in the WCST analog. Previous studies have shown the competence of monkeys in numerical judgment after long-term training (Nieder, 2005). It took >18 months to train the monkeys on the WCST with two dimensions (head and eyes fixed), and we assumed that implementing the third dimension would make the WCST task inaccessible to the monkeys. The monkeys could shift to the other rule when they realized the rule change by assessing the feedback. However, they still needed to keep the currently relevant rule in short-term memory, because there was no cue to indicate it. (2) We required head and eye fixation in WCST analog to facilitate single-cell recording and interpretation of the neuronal activity. (3) In contrast to the clinical WCST, there was a time limit (2500 ms) for monkeys to make a decision and select the target in the WCST analog. This was required to keep the monkeys continuously attentive within the trial and also let us to compare the response latencies in different conditions.

**Recordings.** After 18 months of training, in which the monkeys completely learned the task and eye fixation, each monkey was prepared by an aseptic surgery for recordings. The skull over the recording region was removed, and a recording cylinder was implanted. The recording region was determined referring to the positions of the principal sulcus and the arcuate sulcus in the magnetic resonance images taken before the surgery. Recordings were made from the lateral surface both dorsal and ventral to the principal sulcus. The ventrodorsal extent of recordings was limited to within 5 mm from the center of the sulcus along the cortical surface. The most posterior positions of recordings were  $\sim 3$  mm anterior to the posterior end of the principal sulcus. Action potentials of neurons were recorded extracellularly with tungsten microelectrodes (Fredrick Haer Company, Bowdoinham, ME) inserted through the dura. The microelectrode was advanced through a guide tube by an oil-drive manipulator (MO-81; Narishige, Tokyo, Japan). Signals from a single cell were isolated on-line using a template-matching system (MSD; Alpha Omega Engineering, Jerusalem, Israel). The cells that changed their activity during the time course of a trial were preferentially recorded. All animal training, surgery, and experimental procedures were done in accordance with National Institutes of Health guidelines and were approved by the RIKEN Animal Experiment Committee.

**Data analyses.** Neuronal activity was analyzed in seven epochs: the ITI (600 ms immediately before start-cue appearance), "start cue" (600 ms before the appearance of the fixation point), "fixation" (600 ms starting from 100 ms after fixation point onset), "sample" (600 ms starting from 30 ms after sample onset), "decision" (600 ms after onset of test items), "before reward" (600 ms immediately before reward delivery), and "after reward" (600 ms after reward delivery). We did not observe any difference in the characteristics of PFC cells between the dorsal and ventral banks of the principal sulcus or between the posterior and anterior parts of the recorded area. Therefore, results are reported for all of the recorded cells.

Monkey M258 pushed the lever with its right hand, used its left hand to touch the left-side test item, and used its right hand to touch the test items

on the bottom and right side of the screen. Monkey M259 pushed the lever with its right hand and used only the left hand for touching the test items. Monkey M259 was generally much faster in his responses. We calculated the response latency as the time between the appearance of the test items and the subsequent first touch of the screen.

**Activity differences between color and shape blocks.** To assess the rule dependency of activity, a repeated-measure two-way ANOVA [epoch (within-subjects factor)  $\times$  rule (between-subjects factor)] was conducted for correct trials of all of the recorded cells. The sphericity assumption was tested with Mauchly's test, and Greenhouse-Geisser adjustment was implemented. A cell was considered as having rule dependency of activity if there was a significant ( $p < 0.025$ ) main effect of rule or its interaction with epoch.

The rule dependency of activity was examined between activities in two groups of blocks. The block design is inherent structure in the WCST and our analog. Investigating the neuronal activity across blocks of trials requires certain efforts to avoid the possible effects of slow changes in the excitability of the recorded cells (Asaad et al., 2000). We took several steps to make certain that our results were not the artifact of block design. We included those cells for data analysis for which the activity was recorded while monkeys completed at least six blocks. Therefore, the cellular activity was recorded while there were three repetitions of each rule. In the comparison of activity between color and shape blocks, the "local average" was subtracted from the activity in individual trials, epoch by epoch. The local average was calculated by adding the means from the previous, current, and next blocks, with weights of 0.25, 0.5, and 0.25, respectively. In the first and last blocks, the local average was calculated by taking the simple average of means from the first or last two blocks. The calculations of local average and subtraction were conducted independently for each of the seven epochs. Next, the activities in color and shape blocks were compared by the two-way ANOVA (rule  $\times$  epoch). Finally, the detection of a pseudo-rule dependency attributable to any residual block-to-block variation was examined by applying a one-way ANOVA (block as factor) to the activities in color and shape blocks, separately. Fourteen cells that showed significant variation among color blocks or among shape blocks were excluded. These approaches were taken only for the selection of the cells with rule dependency of activity, and other analyses were conducted with no local average subtraction.

Because the rule dependency was seen in different epochs in different cells, we determined the epochs in which the significant rule dependency was detected in each cell (by  $t$  test) and also the epoch in which the largest/consistent difference between color and shape was detected in each cell (smallest  $p$  value for the difference between color and shape obtained by a  $t$  test). The one-way ANOVA (block as factor) and calculation of the magnitude of rule dependency was done for the activity in these epochs.

**Correlation of the rule-related activity with monkeys' performance.** To examine the correlation between the magnitude of the rule-related activity and the monkeys' performance in adapting to rule changes (see Fig. 4), we determined the best-, intermediate-, and worst-performance blocks, separately for color and shape rules, according to the number of error trials in each block. For cells recorded in six blocks, the best-, intermediate-, and worst-performance blocks were defined as blocks with the largest, intermediate, and smallest number of error trials, respectively. For cells recorded in more than six blocks, the block with a number of error trials closest to the averaged number of error trials in the best- and worst-performance blocks was selected as the intermediate-performance block.

We used the number of error trials rather than the percentage of correct trials in each block as the measure of performance (monkeys' behavioral adaptability to rule changes), as done previously in WCST studies (Heaton, 1993; Stuss et al., 2000). In fact, the number of error trials dissociated from the percentage of correct trials in many cases.

Those epochs in which a significant rule dependency was found contributed to the three-way ANOVA analysis (rule  $\times$  performance  $\times$  cell) and the calculation of correlation between the magnitude of the rule-related activity and the number of errors (see Fig. 4*d,e*). For the calculation of correlation between the mean activity and number of errors in each block (see Fig. 4*f,g*), the mean activity of each block was calculated in

the epoch in which the smallest  $p$  value for rule dependency was obtained by the  $t$  test. The mean activity was normalized by the overall average activity of each cell. In the analysis of the correlation in population of cells (see Fig. 4*d–g*), the two levels of the rule (color and shape) were replaced with preferred and nonpreferred rules. The preferred rule (the rule in which the cell showed higher activity) was determined based on the mean activity averaged over the last 20 trials in the best-, intermediate-, and worst-performance blocks.

**Comparison of rule-related activity between correct and error trials.** We could not simply compare the rule dependency of activity between correct and error trials in the 71 cells with significant rule-related activity because these cells had been selected based on the presence of significant rule dependence in correct trials. A fake difference in the magnitude of rule dependence (larger in correct trials) could have appeared as a result of a selection-result circulation. Therefore, we used a split-data method to avoid any possible bias in the results attributable to the selection of cells. First, activities were taken from half of the correct trials, which were randomly selected from all of the correct trials. These activities were used to select cells according to the significant rule dependency [by a two-way ANOVA (rule  $\times$  epoch)]. For the selected cells, the activities in the same half of correct trials were used to determine the epochs with significant rule dependency ( $t$  test) and the preferred rule in which the cell showed higher activity. A three-way ANOVA (rule  $\times$  response type  $\times$  cell) was then applied to the activity of selected cells in the remaining half of correct trials and all error trials. The first error trial in each block was excluded from these analyses.

**Activity difference between correct and error trials.** To assess the activity difference between correct and error trials, a repeated-measure three-way ANOVA [epoch (within-subjects factor)  $\times$  response type (between-subjects factor)  $\times$  rule (between-subjects factor)] was conducted for all of the recorded cells. The sphericity assumption was tested with Mauchly's test, and Greenhouse-Geisser adjustment was implemented. We tested the homogeneity of variance between correct and error trials for each one of the 214 recorded cells by Hartley's  $F_{\max}$  test. Of 214 recorded cells, only 12 cells showed significant difference. This number of cells is not different from the number that we could get by chance (binomial test,  $p > 0.1$ ). When activity differences between correct and error trials were analyzed by the three-way ANOVA, the activities in individual correct and error trials in each block were subtracted by mean activity in correct trials of the block to remove the block-to-block variation in activity. This procedure was conducted separately for each epoch.

**Neuronal responses to feedback.** In fixation-break and error trials, responses to the presentation of visual error signal were defined by changes in firing rate between the 300 ms period immediately before error-signal onset and the period from 100 to 500 ms after error-signal onset. In correct trials, responses to the delivery of liquid reward were defined by changes in firing rate between the 300 ms period immediately before reward delivery onset and the 600 ms period immediately after reward delivery onset. The responses were compared between color and shape blocks by  $t$  test. The possible contribution of differences in the eye position to the differences in neuronal responses to the error-signal presentation between color and shape blocks was examined by applying  $t$  tests to the averaged  $x$  and  $y$  coordinates of the eye position during the error-signal presentation in individual trials of color and shape blocks.

**Modulation of neuronal activity by success/failure in the preceding trial.** When we compared the neuronal activity between trials after correct and those after error trials, both correct and error trials were used, but those trials that were terminated because of eye fixation breaks or early release of the bar were excluded from the analysis. In the comparison of after-correct and after-error trials between the early and last parts of the block, the cells were divided to two groups: (1) higher activity in after-correct and (2) higher activity in after-error. We calculated an index for the first and last 20 trials for each cell:  $(FR_{\text{Acorrect}} - FR_{\text{Aerror}})/(FR_{\text{Acorrect}} + FR_{\text{Aerror}})$ , where  $FR_{\text{Acorrect}}$  and  $FR_{\text{Aerror}}$  represent the firing rates (FR) in after-correct and after-error trials, respectively. The paired  $t$  test was separately applied to these cell groups to compare the index between the early and late parts of the block.



## Results

### Behavioral performance

We trained two monkeys to perform a WCST analog with two dimensions (Mansouri and Tanaka, 2002). The monkeys had to match a sample to one of three test items by either color or shape and touch the correct item to receive a reward (supplemental Fig. 1a, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). The relevant rule for matching (matching by shape or matching by color) was consistent within a block of trials, and it changed without any notice to the monkey when a criterion of 85% correct performance was achieved. The relevant rule was not cued. Even the reward and error signal by themselves did not indicate it directly. The monkeys could identify the relevant rule only by applying a rule and then interpreting the reward or error feedback in the context of the applied rule.

Both monkeys successfully adjusted their behavior to >10 changes in relevant rule per day. In a behavioral test in which new samples were introduced, both monkeys immediately generalized the matching rules to new samples (Mansouri and Tanaka, 2002). This demonstrated that the monkeys performed the matching by relying on the rules rather than on individual associations between sample and target items or location. In another behavioral test with 36 samples, in which one monkey made  $11.5 \pm 0.8$  (mean  $\pm$  SEM) shifts per day and the other one made  $11.3 \pm 1.1$  shifts per day, both monkeys attained 90% correct performance before any sample was repeated within the block in the majority (75%) of blocks (Mansouri and Tanaka, 2002). This showed that the monkeys inferred the relevant rule by receiving feedback to their responses for some of the samples and applied it to the rest within the block.

In a behavioral study with other monkeys performing a similar WCST analog (our unpublished observation), after the monkey attained a high performance with one rule, we let the monkey perform only one trial of another simple and unrelated task and then return to the WCST analog and continue to follow the same rule. The distraction for a brief period dramatically impaired their performance, indicating the transient nature of the memory of relevant rule and its vulnerability to distracting events. Together, the behavioral data indicated that the monkeys had a rule-based strategy, and the short-term memory of the relevant rule supported their performance.

While the activity of PFC cells was recorded, the monkeys made  $9.0 \pm 0.16$  (mean  $\pm$  SEM) rule shifts per day. Besides eye fixation errors, there were four other types of possible errors in each trial: (1) early release of the bar during the fixation and sample periods, (2) failure to touch the target in the response window (2500 ms), (3) touching the test item that did not match the sample in either color or shape dimensions, and (4) touching the test item that matched the sample according to the irrelevant rule.

A vast majority (93%) of the monkeys' errors were of the last type, i.e., those attributable to selecting the test item that matched the sample based on the irrelevant rule. Such erroneous selections were more frequent after a rule change, and their frequency decreased toward the end of blocks. However, unlike normal human subjects, the monkeys continued to commit occasional erroneous selections even at the last part of blocks (Mansouri and Tanaka, 2002). Error trials in the rest of this paper refer to trials in which the monkeys selected the target that matched the sample by the irrelevant rule.

In contrast to the relatively consistent performance of monkeys in adapting to reversals in previous studies with stimulus–response association tasks (Pasupathy and Miller, 2005), the

monkeys' performance in learning the new rule after rule changes considerably varied from block to block in the present study. During the recording of PFC cell activity (requiring head and eye fixation), the number of errors committed before completing a block ranged from 3 to 50. Even in highly trained monkeys performing a similar WCST analog in a freely moving condition, the number of errors varied from 1 to 20 (our unpublished observation). This might be attributable to the complexity of the WCST, which requires a collective contribution of multiple cognitive processes. It is possible that, although the monkeys had acquired the ability to shift by a few errors, they did not work at their maximum performance in all of the blocks.

### Neuronal activity

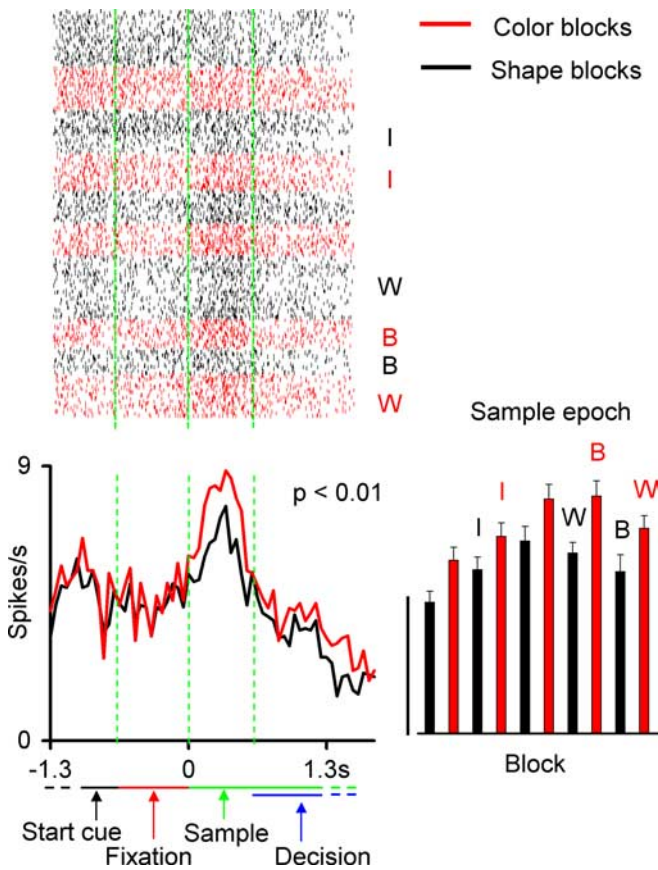
The activity of 214 (134 from one monkey and 80 from the other) neurons was recorded from the dorsolateral PFC (supplemental Fig. 1b, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) in at least six blocks of trials with five rule shifts. Most of the cells (207 of 214) significantly changed their firing rates during the time course of a trial [significant main effect of epoch in a two-way ANOVA (rule  $\times$  epoch),  $p < 0.05$ ; for definition of the seven epochs within a trial, see Materials and Methods, Data analysis].

### Activity differences between color and shape blocks

For one-third of the recorded cells (71 of 214 cells, 37% in one monkey and 28% in the other), the neuronal activity in correct trials was significantly different between color and shape blocks [significant main effect of rule or significant interaction between rule and epoch in the two-way ANOVA (rule  $\times$  epoch)]. Some cells showed higher firing rates in color blocks (35 cells), and others did so in shape blocks (36 cells). The possibility that slow changes in excitability led to the detection of a pseudo-rule dependence was carefully eliminated in this analysis (see Materials and Methods, Data analyses). Figures 1 and 2 show the activity of three PFC cells. The activity difference between color and shape blocks (rule dependence of activity) was the largest in the sample, fixation, and ITI epochs (defined by the smallest  $p$  value in *post hoc* Newman–Keuls test) for the cells shown in Figures 1 and 2, *a* and *b*, respectively. The bar graphs in the right part of individual figures, which represent block-to-block changes of mean firing rate in these epochs, show that the rule dependence of activity was consistent in consecutive blocks.

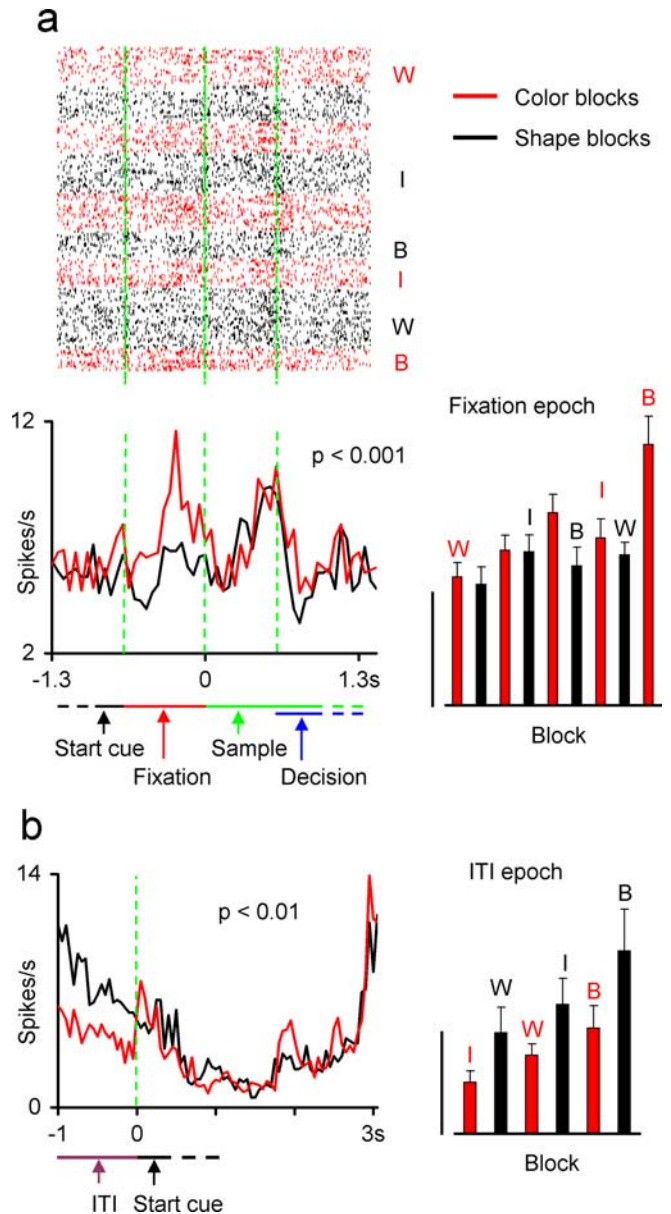
To examine the distribution of rule dependence of activity across different epochs of the trial,  $t$  test was applied to the activity in each epoch to assess the difference between color and shape blocks. The rule dependence of activity was observed in one to six epochs (mean of 2.1 epochs;  $t$  test) in each of the 71 cells. The epoch in which the rule dependence was seen varied from cell to cell, and across the cell population all epochs including ITI showed rule dependence of activity (Fig. 3). The percentage of cells that showed significant rule dependence in each epoch ranged from 15 to 20% (18% on average) among the 214 recorded cells and from 24 to 38% (31% on average) among the 71 cells selected by the two-way ANOVA. Across the 71 cells, the average magnitude of the rule-dependent component of activity was 28.4% in terms of the averaged firing rate.

The rule dependence of activity was not associated with particular samples, because many cells showed significant rule dependence in the ITI, cue, and fixation epochs in which a sample had not been shown. Also, when a two-way ANOVA (sample  $\times$  rule) was applied to the sample epoch activity of the 71 cells, a significant interaction between rule and sample was observed in only 3 of 25 cells that showed a significant main effect of rule.



**Figure 1.** Activity difference between color and shape blocks during the sample epoch in a PFC cell. The raster gram indicates activities of a PFC cell in individual correct trials, aligned at the onset of sample presentation, in color and shape blocks. Each row corresponds to a trial, and each dot represents an action potential. The line graphs at the bottom left show the averaged firing rates in color and shape blocks, also aligned at the sample onset. The bin size is 50 ms. The  $p$  value indicates the significance level of interaction between rule and epoch in a two-way ANOVA. The bar graph at the bottom right represents the mean firing rate during the sample epoch in consecutive blocks. Vertical calibration bar, 5 spikes/s. The error bars of the bar graph represent SEMs. Red and black colors of dots, lines, and bars indicate color and shape blocks, respectively. The best-, intermediate-, and worst-performance blocks are marked by B, I, and W, respectively.

In the incongruent condition, one of the three test items matched the sample in color and another matched the sample in shape. In the congruent condition, one of the three test items was identical to the sample, and therefore the correct response for a given sample was identical in color and shape blocks (see Materials and Methods, Behavioral task). For both monkeys, in all three response directions, the response latency was shorter in congruent trials. A two-way ANOVA [congruency (congruent/incongruent)  $\times$  direction (left/right/down)] applied to the response latency showed that, in each of the recording sessions, the main effect of congruency was significant ( $p < 0.05$ ). The difference in response latency between congruent and incongruent trials was  $123 \pm 3.3$  (mean  $\pm$  SEM),  $38 \pm 3$ , and  $125 \pm 3$  ms in M258 and  $70.8 \pm 7$ ,  $56 \pm 5$ , and  $72 \pm 2$  ms in M259 for the leftward, rightward, and downward hand movements, respectively. The rule dependence of activity was observed even in the congruent condition. When the two-way ANOVA (rule  $\times$  epoch) was applied separately to the activity in incongruent and congruent conditions, significant effects of rule were observed in 33 and 25% of the 214 cells, respectively (Table 1). Note that the incongruent and congruent conditions can be defined only after

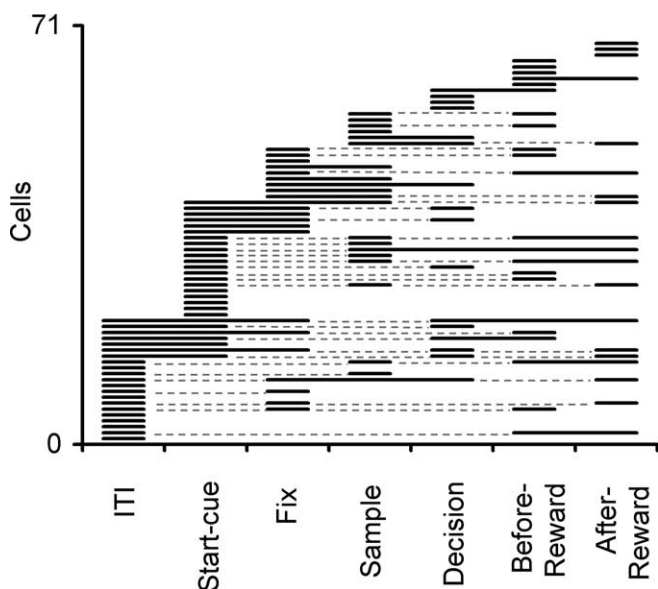


**Figure 2.** Two other PFC cells with significant activity differences between color and shape blocks during the fixation and ITI epochs. The notation is the same as that in Figure 1, except that, for the cell in *b*, the activities were aligned at the start-cue onset. The rule dependence appeared in the fixation epoch and ITI epoch for the cells in *a* and *b*, respectively. There is no raster gram shown for the cell in *b*.

the sample onset. Therefore, the analysis here was applied only to activities in the sample and later epochs.

### Correlation of the rule dependence of activity with monkeys' performance

We then studied whether the magnitude of the rule-dependent component of activity covaried with the performance of monkeys. We used the number of error trials in each block as an index of efficiency of the monkeys' behavior in adapting to rule changes. For each cell, we defined the best-, intermediate-, and worst-performance blocks as those with the smallest, intermediate, and largest number of error trials, respectively. They were determined separately for color and shape blocks. Two cells that were recorded while the monkeys had a very low performance ( $>30$  errors per block after averaging over the best-,



**Figure 3.** The distribution of activity differences between color and shape blocks in different trial epochs. Each line corresponds to one cell. The solid parts of the line indicate the epochs in which the cell showed significant differences between color and shape blocks. The thin broken parts connect the epochs with significant differences when significant differences appeared in discontinued epochs. Of the 71 cells for which the ANOVA showed significant rule dependency, two cells are not shown here because they did not show significant differences when the activity difference was examined by *t* test for each epoch. For 16 of the 71 cells, which were recorded at an early stage of the experiment, we did not store the activity during the ITI.

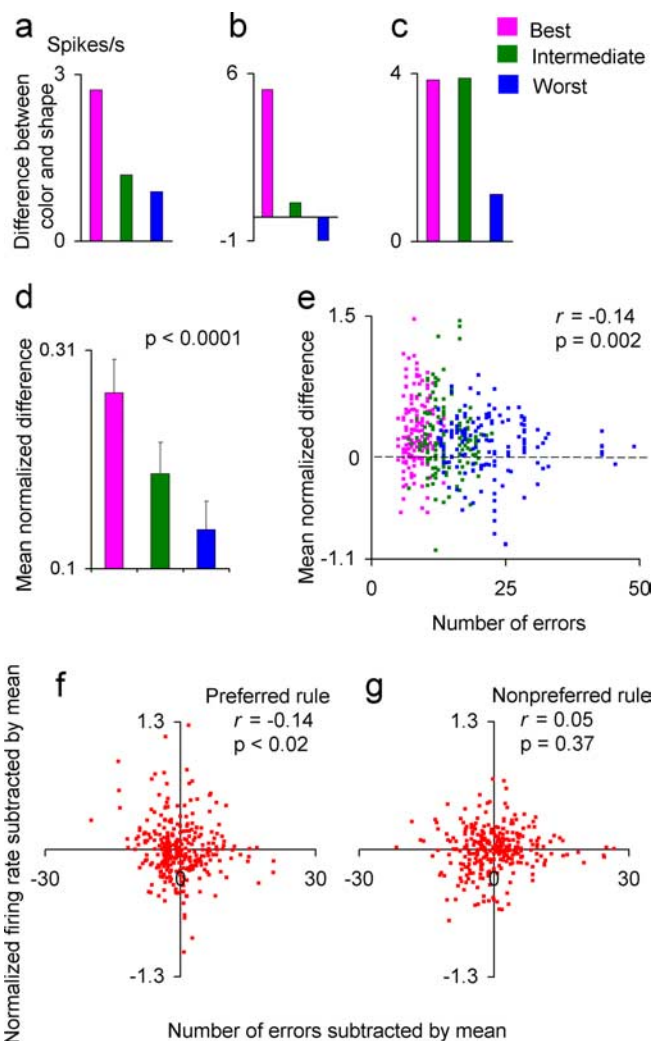
**Table 1:** Numbers of cells that showed different types of activity modulation.

Rule-dependent activity	Dependence on success/failure in current trial	Response to reward	Response to error signal	Dependence on success/failure in previous trial
71 (33%)	32 (15%)	110 (51%)	95 (44%)	ITI epoch 35 (20%)
Rule-related in congruent 54 (25%)	Higher firing in correct 13	Rule-dependent 9 (8%)	Rule-dependent 16 (17%)	Higher firing in After-correct 9
Rule-related in incongruent 71 (33%)	Higher firing in error 19	Excitatory response 11	Inhibitory response 5	Higher firing in After-error 26
		larger in color 7	larger in shape 3	Start-cue epoch 45 (21%)
		4	2	Higher firing in After-correct 26
				Higher firing in After-error 19
				Fixation epoch 29 (14%)
				Higher firing in After-correct 14
				Higher firing in After-error 15

The number of cells with significant modulation is shown for each type of activity modulation. The total number of recorded cells was 214 cells, while activity in ITI epoch was recorded only in 178 cells.

intermediate-, and worst-performance blocks) were excluded from the following analyses. While the 69 cells were recorded, the numbers of errors were  $8.5 \pm 0.2$  (mean  $\pm$  SEM),  $14 \pm 0.3$ , and  $21.8 \pm 0.6$  in the best-, intermediate-, and worst-performance blocks, respectively.

To calculate the activity difference between color and shape blocks at a particular performance level, we paired the best-performance color block with the best-performance shape block. Pairs were also formed for the intermediate- and worst-performance blocks separately. The activity difference was calculated between pairs of the best-performance blocks, between those of the intermediate-performance blocks, and between those of the worst-performance blocks. Figure 4*a–c* represents the magnitude of activity differences between color and shape blocks of the three performance levels for the three example cells shown in Figures 1 and 2, *a* and *b*, respectively. The activity difference in each of the 69 cells with significant rule dependency was normalized by the mean activity of the cell and then averaged across cells to obtain the mean normalized activity differences

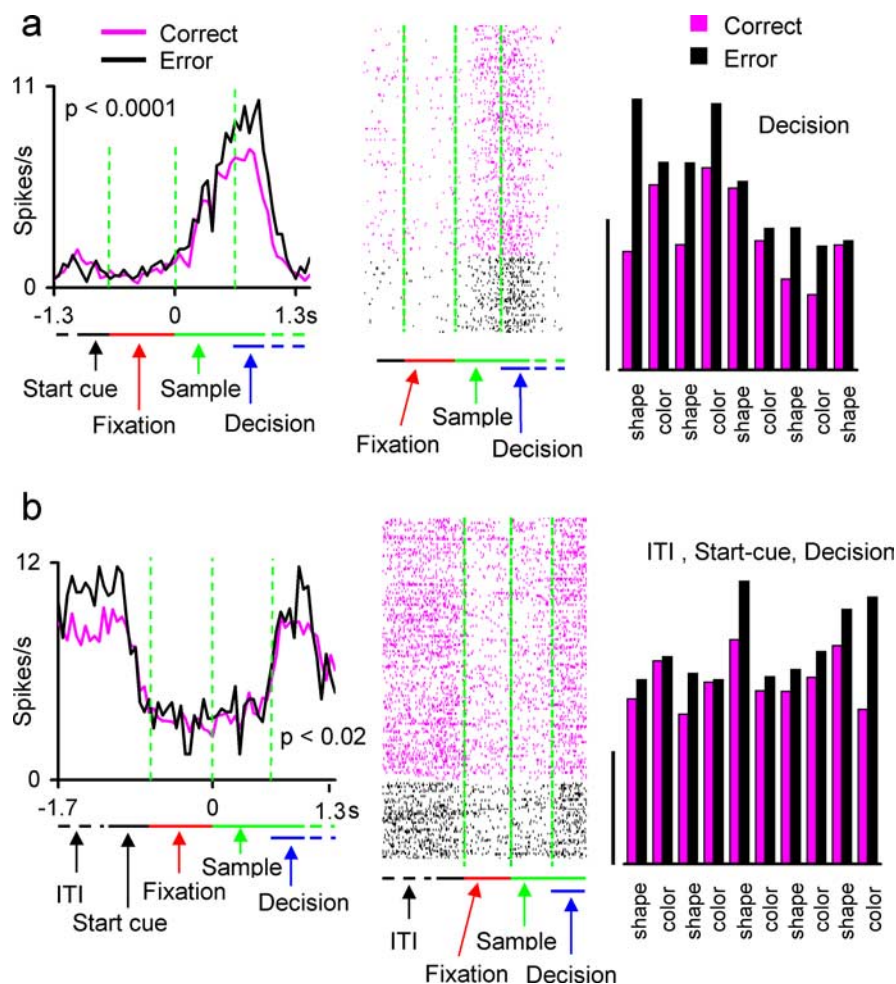


**Figure 4.** Correlation of the magnitude of activity differences between color and shape blocks with monkeys' performance. *a–c*, The averaged activity differences between pairs of the best-, intermediate-, and worst-performance blocks, for the cells shown in Figures 1, 2*a*, and 2*b*, respectively. *d*, The mean normalized activity differences between pairs of the best-, intermediate-, and worst-performance blocks, for the population of 69 cells. The differences were normalized by the mean activity of individual cells and then averaged over the 69 cells. The averaging across cells was made for preferred and nonpreferred rules. *e*, The difference in activity (preferred – nonpreferred) between each pair of the best-, intermediate-, and worst-performance blocks is plotted against the averaged number of errors in the block pair for the 69 cells. *f, g*, The deviation of the mean firing rate in each block from the overall mean of individual cells is plotted against the deviation of the number of errors in the block from the mean number of errors during the recording from the cell for preferred (*f*) and nonpreferred (*g*) rules. The deviation of firing rate was normalized by the mean firing rate. The *r* and *p* values in *e–g* represent the correlation coefficient and the significance of the correlation, respectively.

between preferred (for which each cell showed a higher activity) and nonpreferred rules in the cell population (Fig. 4*d*). The mean normalized activity difference was the largest ( $0.27 \pm 0.03$ ) in the best-performance blocks. It was smaller ( $0.19 \pm 0.03$ ) in the intermediate-performance blocks and the smallest ( $0.14 \pm 0.03$ ) in the worst-performance blocks.

To examine the correlation between rule-related activity and performance in the population of 69 cells, the color and shape rules were replaced by preferred and nonpreferred rules, and a three-way ANOVA [rule (preferred/nonpreferred)  $\times$  performance (best/intermediate/worst)  $\times$  cell (69 cells as levels)] was applied. There was a highly significant interaction between rule and performance ( $p < 0.0001$ ), indicating that the magnitude of





**Figure 5.** Activity differences between correct and error trials in two PFC cells. *a, b*, The line graphs at the left show the averaged firing rates in correct and error trials, aligned at sample onset. The *p* values indicate the significance of effects of response type in a three-way ANOVA (rule  $\times$  response type  $\times$  epoch). The bin size is 50 ms. The raster grams in the middle show activity in individual correct and error trials of the first four blocks, also aligned at sample onset. The bar graph at the right represents the mean firing rate in the epoch noted above individual graphs for correct and error trials in consecutive blocks. The activity differences between correct and error trials were consistently seen in the same direction in consecutive color and shape blocks. Pink and black colors of lines, bars, and dots indicate correct and error trials, respectively. Vertical calibration bar, 5 spikes/s.

the rule-dependent component of activity varied significantly with the monkeys' performance. Also, a significant negative correlation (Pearson's correlation,  $r = -0.14$ ;  $p = 0.002$ ) was seen between the normalized activity differences and the numbers of errors (averaged between two rules), as shown in the scatter plot (Fig. 4*e*), in which each dot represents a pair of blocks of the same performance level. These results indicate that the magnitude of the rule-dependent component of activity correlated with the monkeys' performance. The magnitude of the rule-dependent component of activity was larger when the monkeys performed better and reached a high performance level with fewer errors.

The covariance was also directly examined by correlation analysis between neuronal activity and monkeys' performance in individual blocks. To examine the correlation between the deviations of activity and performance in individual blocks from their means, we subtracted the activity in each block of a rule by the mean activity in all the blocks of that rule, and the number of errors in the block was subtracted by the averaged number of errors in all of the blocks of that rule. These calculations were performed separately for preferred and nonpreferred rules. Figure 4, *f* and *g*, shows the scatter plots for blocks of preferred and

nonpreferred rules, respectively. One dot represents one block of trials, and all of the 69 cells with significant rule dependence were included. There was a significant negative correlation between the magnitude of activity in each block and the number of errors committed in the block for the preferred rule ( $r = -0.14$ ;  $p < 0.02$ ), but no correlation was seen for the nonpreferred rule ( $r = 0.05$ ;  $p = 0.37$ ). The negative correlation between the neuronal activity and the number of errors in each block indicates that, for the preferred rule, when the monkeys committed fewer errors to re-establish the high performance in a block, the cellular activity was higher in that block. These results suggest that the correlation of neuronal activity with the monkeys' performance mainly occurred in blocks of the preferred rule.

Observing the rule dependence of activity in correct trials and the correlation of its magnitude with monkeys' performance, one may suspect that a reduction of the rule-dependent component of activity, or its reversal in direction, caused the monkeys' erroneous selections in error trials. To examine this possibility, we compared the magnitude of the rule-dependent component of activity between correct and error trials. The difference in the magnitude of the rule dependence between correct and error trials was expected to appear in ANOVA analysis as a significant interaction between rule and response type (correct/error). We applied a three-way ANOVA (rule  $\times$  response type (correct/error)  $\times$  cell) to the activity of the cells that showed significant rule dependence of activity (see Materials and Methods, Data analyses). There was no significant interaction between rule and response type ( $p = 0.18$ ), whereas the main effect of rule remained highly significant ( $p < 10^{-6}$ ). These results indicate that the rule dependence of activity in correct trials remained in the same direction and with nearly the same magnitude in error trials.

#### Activity differences between correct and error trials

Despite the similarity of rule representation in correct and error trials, our additional analyses showed that there was rule-independent activity differences between correct and error trials. We applied a three-way ANOVA [rule  $\times$  response type (correct/error)  $\times$  epoch] to each of the 214 recorded cells. The analysis was limited to activities from the ITI to decision epochs, because activities in later epochs could have been modulated by performance feedback (water rewards and error signals). There was a significant effect of response type (main effect of response type or interaction between response type and epoch,  $p < 0.025$ ) in 32 of the 214 cells. In 19 cells, activity was higher in error trials (Fig. 5*a,b*), and, in 13 cells, activity was higher in correct trials (no examples are shown).

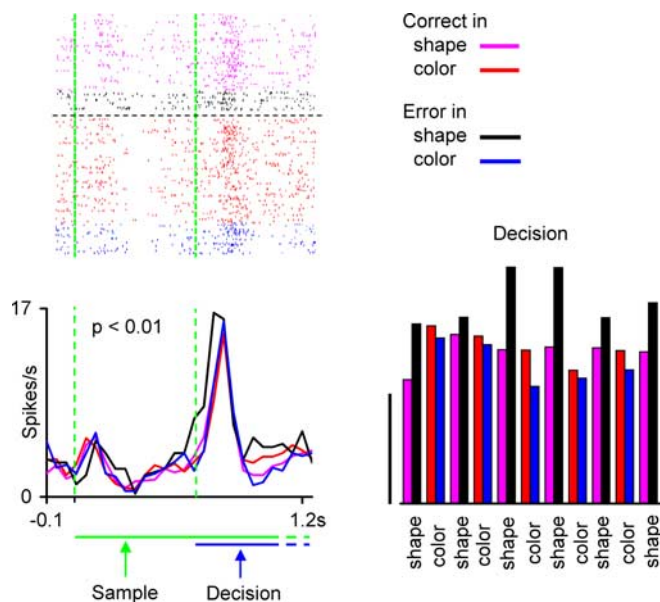
The bar graphs in Figure 5, *a* and *b*, show that the activity

difference between correct and error trials was consistently seen in the same direction in consecutive color and shape blocks. Because the difference between correct and error trials appeared in different epochs among the 32 cells, the activity in the epoch with the largest difference between correct and error trials was used for additional analysis (*t* test was applied for differences between correct and error trials, and the epoch with smallest *p* value was selected). In all but 1 of the 32 cells, the direction of activity differences between correct and error trials was consistent between color and shape blocks. Only 5 of the 32 cells showed significant rule dependence of activity in correct trials. These indicated that the differences between correct and error trials were mostly independent of the relevant rule. The magnitude of activity difference between correct and error trials was on average 22% of the averaged activity.

In the error trials used in the comparison of cellular activity in correct and error trials, the monkeys selected the target based on the irrelevant rule. It is unlikely that these errors were just resulting from the monkeys' lack of interest or inadvertent mistakes, because monkeys started the trial by themselves and had to keep pushing the bar and fixating their eyes until the appearance of test stimuli. The monkeys seldom selected the test item that did not match the sample by either rule. This suggests that the monkeys were attentive even in error trials while performing the task.

Because there were more error trials in early parts of blocks than in late parts of blocks, the observed activity difference between correct and error trials could be related to activity changes from early to late parts of blocks. To examine this possibility, we compared the activity differences between correct and error trials between the first and last 20 trials of individual blocks. A two-way ANOVA (early/late × response type) showed a significant interaction between response type and the early/late factor in only 2 of the 32 cells. We also calculated an index for the first and last 20 trials for each cell:  $(FR_{\text{correct}} - FR_{\text{error}})/(FR_{\text{correct}} + FR_{\text{error}})$ , where  $FR_{\text{correct}}$  and  $FR_{\text{error}}$  represent the firing rates in correct and error trials, respectively. There was no difference in index value between the first and last 20 trials in either the group of 19 cells with higher activity in error trials ( $p = 0.8$ , paired *t* test) or the group of 13 cells with higher activity in correct trials ( $p = 0.67$ ). These results indicate that the observed activity differences between correct and error trials were consistent between the early and late parts of blocks.

The three-way ANOVA (rule × response type × epoch) also showed that there was significant interaction between response type and rule in 23 cells (significant interaction between rule and response type or significant interaction among the three factors,  $p < 0.025$ ). Of these 23 cells, five cells also showed significant main effect of response type or significant interaction between response type and epoch. Figure 6 shows the activity of an example cell. The bar graph in Figure 6 shows that the mean activity in error trials was higher than that of correct trials in shape blocks, but the difference was attenuated and reversed in direction in color blocks. For this cell, there was no significant activity difference in correct trials between color and shape blocks ( $p = 0.44$ , *t* test). Only 7 of the 23 cells showed significant rule dependence of activity in correct trials. This indicates that the significant interaction between rule and response type was mostly attributable to changes in the activity in error trials. Thus, the activity differences between correct and error trials were generally independent from the rule-related modulation of activity in correct trials.



**Figure 6.** Activity difference between correct and error trials in another PFC cell. The notation is the same as that in Figure 5, except that pink and black colors indicate correct and error trials, respectively, in shape blocks, and red and blue colors indicate the same thing in color blocks. The *p* value indicates the significance of interaction between response type and rule in the three-way ANOVA (rule × response type × epoch).

### Neuronal responses to feedback

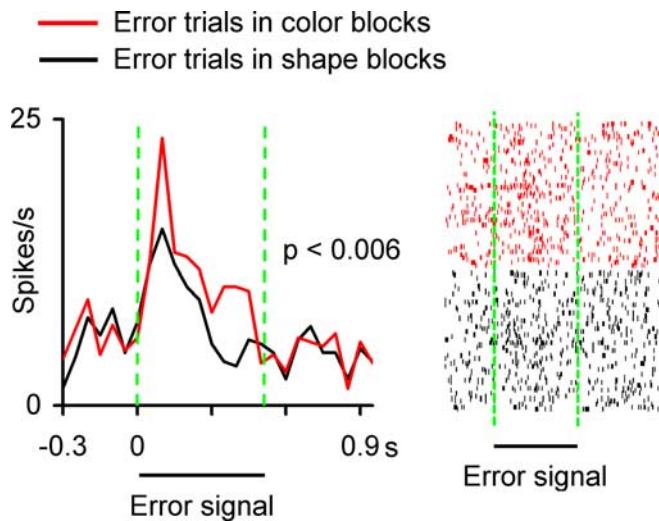
The monkeys had to rely on feedback (error signal or reward) to their selection to find the relevant rule and its changes. Approximately one-half of the 214 cells (95 cells, 44%) significantly (paired *t* test at  $p < 0.025$ ) responded to the visual error signal presented after erroneous target selections in color and/or shape blocks.

The meaning of the error signal as feedback could be different between color and shape blocks. We found that 17% (16 cells) of the 96 cells with significant responses to the error signal also showed a significant (*t* test,  $p < 0.05$ ) difference in error responses between color and shape blocks (Fig. 7). This number of cells is significantly ( $p < 10^{-4}$ , binomial test) different from the number of cells expected by chance. There were no significant differences ( $p > 0.05$ , *t* test) in either horizontal or vertical position of the eyes between color and shape blocks in 14 of the 16 cells with rule-dependent responses to the error signal.

The differences in responses to the error signal between color and shape blocks cannot be explained by the modulation of sensory responses by selective attention to the relevant sensory dimension. If selective attention to a specific sensory dimension had modulated the activity of these cells, responses to other items appearing in other epochs of the error trials should have also been modulated. The firing rate during the sample epoch and that during the start-cue epoch of error trials were significantly different between color and shape blocks (*t* test,  $p < 0.05$ ) in only 4 and 2 cells, respectively, of the 16 cells. Also, we defined “sample response” as the difference between the before-sample period (300 ms before sample onset) and the after-sample period (from 100 ms after sample onset for 400 ms). When we compared the sample response in error trials between color and shape blocks (*t* test,  $p < 0.05$ ), only 3 of 16 cells showed a significant difference between color and shape blocks. These results indicated that the rule-related modulation was seen only in the neuronal responses to error signal.

We also addressed this issue by comparing responses to the





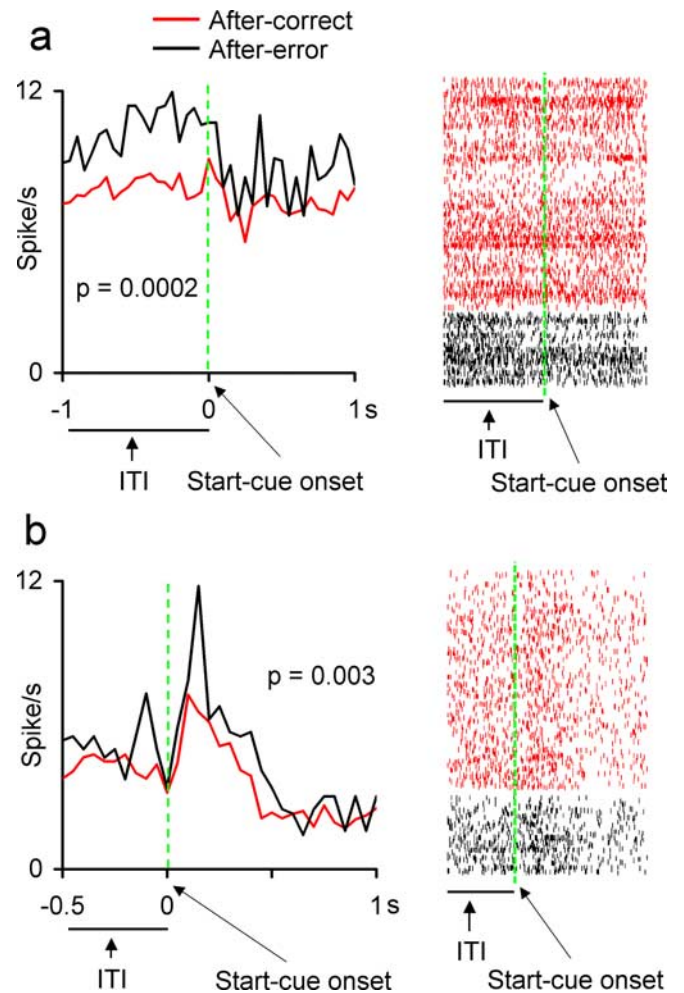
**Figure 7.** Rule-dependent response to error-signal presentation. For a PFC cell, the perievent histogram is aligned at the error-signal onset and shows the average activity in error trials of color and shape blocks. The raster gram indicates activities in individual error trials, aligned at the error-signal onset. Each row corresponds to a trial, and each dot represents an action potential. This cell had a significant (paired *t* test,  $p < 0.01$ ) response to the error-signal presentation in both color and shape blocks. The *p* value on the graph indicates the significance level (*t* test) for the difference in the magnitude of responses between color and shape blocks. The bin size is 50 ms. Red and black dots/traces indicate color and shape trials, respectively.

error signal in error and fixation-break trials. The same error signal was presented after selection errors and fixation breaks. If selective attention had caused the modulation of responses to the error signal, the responses to the error signal after fixation breaks should have also been modulated in the same direction. Of 16 cells that showed significant rule dependence in the magnitude of responses to the error signal, only one showed significant difference in responses to the error signal after fixation breaks. Another cell showed significant difference but in the opposite direction. These results suggest that rule dependence of neuronal responses to the error signal was seen only when the feedback (error signal) was given in error trials.

In our WCST analog, the reward is not only a motivation to perform the task but can also be used to assess the correctness of the behavior. The reward administration evoked significant (paired *t* test at  $p < 0.025$ ) responses in 51% (110 cells) of the 214 cells in color and/or shape blocks. The magnitude of responses to the reward delivery was significantly different between color and shape blocks (*t* test,  $p < 0.05$ ) in only nine of the 110 cells. This proportion is not significantly different from that expected by chance (binomial test,  $p > 0.05$ ).

#### Dependence of activity on the success or failure in the preceding trial

In the WCST analog, the experienced behavioral outcome/feedback should be used to sustain or update the short-term memory of the relevant rule to perform properly in subsequent trials. We compared the neuronal activity between trials after correct trials (after-correct trials) and those after error trials (after-error trials). We applied a two-way ANOVA [trial type (after-correct/after-error)  $\times$  rule (color/shape)] separately to the activity in each of the three epochs before the sample presentation (ITI, start-cue, and fixation epochs). We found that the main effect of trial type was significant ( $p < 0.05$ ) for the ITI epoch in 35 of the recorded cells (20%). In start-cue and fixation epochs, 45 cells (21%) and 29 cells (14%) showed a significant main effect of trial



**Figure 8.** Dependence of activity on the success or failure in the preceding trial. *a, b*, For two PFC cells, the perievent histogram aligned at the start-cue onset shows the average activity in after-correct and after-error trials. The raster gram indicates activities in individual trials, aligned at the start-cue onset. Each row corresponds to a trial, and each dot represents an action potential. The *p* values in *a* and *b* indicate the significance level of the main effect of trial type in two-way ANOVA [trial type (after-correct/after-error)  $\times$  rule] for ITI and start-cue epochs, respectively. The bin size is 50 ms. Red and black dots/traces indicate after-correct and after-error trials, respectively.

type, respectively (Table 1). These numbers were significantly different from the numbers of cells expected by chance ( $p < 10^{-6}$ , binomial test). The numbers of cells showing significant interaction between trial type and rule factors were not significantly different from the numbers expected by chance ( $p > 0.05$ , binomial test). Figure 8 shows the activity of two PFC cells in which the neuronal activity was significantly different between after-correct and after-error trials in the ITI (Fig. 8*a*) and start-cue (Fig. 8*b*) epochs, respectively. These results suggest that, in the period before the appearance of the sample, the activity of PFC cells was modulated by the success/failure in the preceding trial.

The differences in activity between after-correct and after-error trials could not be attributable to the offset of visual stimuli in the previous trial because the ITI epoch began 1800 ms after the latest event in the previous trial, and neither could the differences in activity between after-correct and after-error trials be related to the success or failure (being correct or error) in the current trials (after-correct or after-error). Of those cells that showed activity modulation by the success or failure in the pre-

ceding trial during the ITI, start-cue, and fixation epochs, only 11, 13, and 4%, respectively, showed a significant ( $t$  test,  $p < 0.05$ ) activity difference between correct and error trials in the current trial.

Because there were more error trials in early parts of blocks than in late parts of blocks, the observed activity difference between after-correct and after-error trials could be related to the activity changes from early to late parts of blocks. To examine this possibility, we compared the activity differences between after-correct and after-error trials between the first and last 20 trials of individual blocks. A two-way ANOVA [early/late  $\times$  trial-type (after-correct/after-error)] showed a significant interaction between trial type and the early/late factor in only 3, 0, and 1 of the 35, 45, and 29 cells that showed a significant difference between after-correct and after-error trials in ITI, start-cue, and fixation epochs, respectively. We also calculated an index for the first and last 20 trials for each cell:  $(FR_{A_{correct}} - FR_{A_{error}})/(FR_{A_{correct}} + FR_{A_{error}})$ , where  $FR_{A_{correct}}$  and  $FR_{A_{error}}$  represent the firing rates in after-correct and after-error trials, respectively. There was no difference in the index value between the first and last 20 trials in each of the cell groups that showed a significant difference between after-correct and after-error trials in ITI, start-cue, and fixation epochs (paired  $t$  test,  $p > 0.2$ ). These results indicate that the observed activity differences between after-correct and after-error trials were consistent between the early and late parts of blocks.

## Discussion

### Active maintenance of relevant rule within and across trials

The unique feature of the WCST and its analog used here is the absence of external cues to indicate the relevant rule. In the tasks used in previous single-cell recording studies (Sakagami and Niki, 1994; White and Wise, 1999; Asaad et al., 2000; Wallis et al., 2001), the monkeys were informed of the required task or rule in each trial by an explicit cue or by differences in configurations of trials. In the WCST and our WCST analog, there is no explicit cue or configuration difference to indicate the relevant rule. Therefore, subjects have to maintain the memory of the relevant rule not only within trials but also across trials and sustain/update it by assessing the behavioral outcome/feedback.

The features of the rule-related activity reported here (Figs. 1–3) suggest that this activity might be a neuronal correlate of short-term memory of the relevant rule. First, the rule-related activity in the sample epoch was not specific to particular samples, indicating that it did not originate in the selectivity to particular stimulus attributes (e.g., red) appearing through selective attention to the relevant sensory dimension (e.g., color). Second, the rule-related activity was observed in the congruent condition as well as in the incongruent condition, and therefore it could not reflect differences in sample-response mappings between color and shape blocks. Third and more importantly, the rule-related activity appeared not only in the sample and later epochs (Figs. 1, 3) but also in the ITI, start-cue, and fixation epochs in which the sample had not yet been presented (Figs. 2, 3). Therefore, this activity cannot be ascribed to sensory responses modulated by selective attention to a particular sensory dimension, nor to rule-specific sensorimotor processes on presented samples.

The memory of relevant rule had to bridge across trials. Accordingly, the rule-related activity was found in the ITI, start-cue, and fixation epochs, as well as in the intratrial epochs. The rule-related activity in the sample and later epochs might also reflect rule-related executive processes. The rule-related activity was found in one to six epochs in individual cells. A chain of neurons

might maintain the relevant rule within and across trials (Baeg et al., 2003).

The magnitude of the rule-related activity did not differ between correct and error trials. Therefore, deterioration in the representation of memory of relevant rule could not be a primary cause of individual erroneous responses. It has been reported that some patients with prefrontal damage sometimes committed errors but gave a correct verbal report of the relevant rule (Milner, 1963). They might have problems in the control processes while selecting a target rather than problems in maintaining the memory of relevant rule itself. The monkeys' erroneous responses in WCST analog might originate from malfunction in at least two cognitive processes. An impairment in the short-term memory of the relevant rule would predispose the system to erroneous responses. A malfunction in other control processes, such as a breach in attention, might also lead to erroneous responses even when the relevant rule is fairly maintained in short-term memory.

### Correlation of rule-related activity with monkeys' performance

The magnitude of the rule-related activity was larger when the monkeys committed fewer errors and more efficiently adapted to rule changes (Fig. 4). It is unlikely that this correlation originated in differences in general arousal/attention or motivation (Hasegawa et al., 2000). Because differences in arousal or motivation must have commonly influenced the neuronal activity in both color and shape blocks, they could not increase the activity difference between color and shape blocks.

The observed correlation between rule-related activity and monkeys' performance does not necessarily indicate a causal relationship between the cellular activity and monkeys' behavior. However, it is consistent with the idea that the rule-related activity is the neuronal correlate of the active maintenance of relevant rule. When the representation of the short-term memory of relevant rule was strong, appropriate responses could be evoked more securely based on the maintained rule.

### Activity difference between correct and error trials

The activity differences between correct and error trials (Figs. 5, 6) could not be attributable to the monkeys' knowledge of the correctness of their selections (Watanabe, 1989) because they occurred in the part of trial in which the monkeys had not received any feedback. The interpretation that the differences reflect slow changes in general arousal (Hasegawa et al., 2000) also appears unlikely because correct and error trials were intermingled in each block, and there were approximately equal numbers of cells that showed higher activity in correct and error trials. Finally, because these activity differences were independent from the rule representation in correct trials, they could not represent the rule that the monkey mistakenly applied in error trials.

We raise two different interpretations for these activity differences depending on the epochs in which they appeared. Those differences that appeared before the sample presentation might reflect the control processes that operate as preparatory attention to set up the upcoming rule-based sensorimotor processes. Activity predicting the success or failure of the subject during the period before actual responses has been observed in the dorsolateral PFC by functional magnetic resonance imaging (fMRI) in humans (Hester et al., 2004).

We posit that those activity differences that appeared after sample presentation might be related to the performance monitoring process (Cohen et al., 2000; Gehring and Knight, 2000).

EEG recordings from human subjects have shown activity difference between correct and error trials that appears as a discrete negative potential in error trials around the time of motor response onset [error-related negativity (ERN)]. ERN is thought to originate from the medial PFC, but recent studies also show its dependence on the dorsolateral PFC (Gehring and Knight, 2000). fMRI studies have shown an activity increase in error trials in the dorsolateral PFC as well as in the medial PFC (Carter et al., 1998; Kiehl et al., 2000). Although the performance monitoring process was originally discussed in relation to conflicts in motor responses, a broader definition considers it as the cognitive processes that evaluate the quality of information processing and executive control (Van Veen and Carter, 2002). The activity differences of dorsolateral PFC cells between correct and error trials might reflect the interaction between the dorsolateral and medial PFC networks in the performance-monitoring process. This process might monitor the possibility of error commission regardless of the relevant rule and adjust the allocation of cognitive control in the WCST analog (Botvinick et al., 2001; Cohen et al., 2000; Van Veen and Carter, 2002).

### Representation of behavioral outcome/feedback in cellular activity

The successful performance in WCST requires monitoring and use of feedback information (Fristoe et al., 1997), and activity of PFC time locked to feedback presentation has been observed in humans performing the WCST (Barcelo, 1999; Monchi et al., 2001; Wang et al., 2001). We found that a large population of PFC cells responded to visual error signals, and, in a small but definite proportion of them, the responses were modulated by the relevant rule. The rule dependence of responses to the error signal suggests that the representation of negative feedback in PFC includes the context in which the errors are committed. Although many PFC cells responded to rewards, the responses to rewards were not modulated by the relevant rule. This suggests that the dorsolateral PFC may be more involved in assessing negative feedback rather than positive feedback for correction of behavior.

### History of recent success/failure is encoded in cellular activity

We found that, in the period before the start of a trial, the cellular activity was modulated by the success/failure in the previous trial. This activity modulation might represent the short-term memory of recently experienced outcome/feedback.

In WCST, the process of learning the new rule after rule changes (rule switching) should be directed based on the assessment of behavioral outcome. Our results suggest that the activity of dorsolateral PFC cells conveyed information about the behavioral outcome and its context and the history of recent success/failure. This information might be used to sustain/update the short-term memory of the relevant rule. Transient activities time locked to the motor or rule switching have been observed previously in the medial PFC, inferior PFC convexity, parietal cortex, and caudate nucleus (Shima and Tanji, 1998; Nakahara et al., 2002; Pasupathy and Miller, 2005). We did not observe such transient activities in the dorsolateral PFC, but this was possibly attributable to the considerable block-to-block variability in the speed of learning. However, it is also possible that neuronal activities directly involved in the rule switching mainly occurred in some of these other regions.

### Conclusion

In conclusion, our results suggest that the dorsolateral PFC contributes to the monkeys' flexible rule-based behavior by (1)

maintaining the short-term memory of the relevant rule within and across trials, (2) controlling or monitoring the processes that could lead to success and error in individual trials, (3) assessing the behavioral outcome/feedback, and (4) maintaining the short-term memory of recent success/failure across trials. An impairment of these cognitive processes may underlie the poor performance of prefrontal patients in the WCST.

### References

- Anderson SW, Damasio H, Jones RD, Tranel D (1991) Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *J Clin Exp Neuropsychol* 13:909–922.
- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451–459.
- Baeg EH, Kim YB, Huh K, Mook-Jung I, Kim HT, Jung MW (2003) Dynamics of population code for working memory in the prefrontal cortex. *Neuron* 40:177–188.
- Barcelo F (1999) Electrophysiological evidence of two different types of error in the Wisconsin Card Sorting Test. *NeuroReport* 10:1299–1303.
- Berg EA (1948) A simple objective test for measuring flexibility in thinking. *J Gen Psychol* 39:15–22.
- Berman KF, Ostrem JL, Randolph C, Gold J, Goldberg TE, Coppola R, Carson RE, Herscovitch P, Weinberger DR (1995) Physiological activation of a cortical network during performance of the Wisconsin Card Sorting Test: a positron emission tomography study. *Neuropsychologia* 33:1027–1046.
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. *Psychol Rev* 108:624–652.
- Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280:747–749.
- Cohen JD, Botvinick M, Carter CS (2000) Anterior cingulate and prefrontal cortex: who's in control? *Nat Neurosci* 3:421–423.
- Dias R, Robbins TW, Roberts AC (1996) Primate analogue of the Wisconsin Card Sorting Test: effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behav Neurosci* 110:872–886.
- Drewe EA (1974) The effect of type and area of brain lesion on Wisconsin card sorting test performance. *Cortex* 10:159–170.
- Fristoe NM, Salthouse TA, Woodard JL (1997) Examination of age-related deficits in Wisconsin Card Sorting Test. *Neuropsychology* 11:428–436.
- Funahashi S (2001) Neural mechanisms of executive control by the prefrontal cortex. *Neurosci Res* 39:147–165.
- Gehring WJ, Knight RT (2000) Prefrontal-cingulate interactions in action monitoring. *Nat Neurosci* 3:516–520.
- Goldman-Rakic PS (1990) Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. *Prog Brain Res* 85:325–335.
- Goldstein B, Obrzut JE, John C, Ledakis G, Armstrong CL (2004) The impact of frontal and non-frontal brain tumor lesions on Wisconsin Card Sorting Test performance. *Brain Cogn* 54:110–116.
- Grant DA, Jones OR, Tallantis B (1949) The relative difficulty of the number, form, and color concepts of a Weigle-type problem. *J Exp Psychol* 39:552–557.
- Hasegawa RP, Blitz AM, Geller NL, Goldberg ME (2000) Neurons in monkey prefrontal cortex that track past or predict future performance. *Science* 290:1786–1789.
- Heaton RK (1993) Wisconsin Card Sorting Test Manual. Odessa, FL: PAR.
- Hester RL, Murphy K, Foxe JJ, Foxe DM, Javitt DC, Garavan H (2004) Predicting success: patterns of cortical activation and deactivation prior to response inhibition. *J Cogn Neurosci* 16:776–785.
- Kiehl KA, Liddle PF, Hopfinger JB (2000) Error processing and the rostral anterior cingulate: an event-related fMRI study. *Psychophysiology* 37:216–223.
- Konishi S, Kawazu M, Uchida I, Kikyo H, Asakura I, Miyashita Y (1999) Contribution of working memory to transient activation in human inferior prefrontal cortex during performance of the Wisconsin card sorting test. *Cereb Cortex* 9:745–753.
- Levy R, Goldman-Rakic PS (1999) Association of storage and processing functions in the dorsolateral prefrontal cortex of the nonhuman primate. *J Neurosci* 19:5149–5158.
- Mansouri FA, Tanaka K (2002) Behavioral evidence for working memory of sensory dimension in macaque monkeys. *Behav Brain Res* 136:415–426.



- Mentzel HJ, Gaser C, Volz HP, Rzanny R, Hager F, Sauer H, Kaiser WA (1998) Cognitive stimulation with the Wisconsin Card Sorting Test: functional MR imaging at 1.5 T. *Radiology* 207:399–404.
- Milner B (1963) Effects of different brain lesions on card sorting. *Arch Neurol* 9:90–100.
- Monchi O, Petrides M, Petre V, Worsley K, Dagher A (2001) Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J Neurosci* 21:7733–7741.
- Nakahara K, Hayashi T, Konishi S, Miyashita Y (2002) Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science* 295:1532–1536.
- Nieder A (2005) Counting on neurons: the neurobiology of numerical competence. *Nat Rev Neurosci* 6:177–190.
- Niki H, Watanabe M (1976) Prefrontal unit activity and delayed response: relation to cue location versus direction of response. *Brain Res* 10:79–88.
- Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873–876.
- Pontecorvo MJ, Sahgal A, Steckler T (1996) Further developments in the measurement of working memory in rodents. *Brain Res Cogn Brain Res* 3:205–221.
- Rao SC, Rainer G, Miller EK (1997) Integration of what and where in the primate prefrontal cortex. *Science* 276:821–824.
- Sakagami M, Niki H (1994) Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: relation to relevant task conditions. *Exp Brain Res* 97:423–436.
- Shima K, Tanji J (1998) Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282:1335–1338.
- Stuss DT, Levine B, Alexander MP, Hong J, Palumbo C, Hamer L, Murphy KJ, Izkawa D (2000) Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia* 38:388–402.
- Van Veen V, Carter CS (2002) The timing of action monitoring processes in the anterior cingulate cortex. *J Cogn Neurosci* 14:593–602.
- Volz HP, Gaser C, Hager F, Rzanny R, Mentzel HJ, Kreitschmann-Andermahr I, Kaiser WA, Sauer H (1997) Brain activation during cognitive stimulation with the Wisconsin Card Sorting Test—a functional MRI study on healthy volunteers and schizophrenics. *Psychiatry Res* 75:145–157.
- Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953–956.
- Wang L, Kakigi R, Hoshiyama M (2001) Neural activities during Wisconsin Card Sorting Test—MEG observation. *Brain Res Cogn Brain Res* 12:19–31.
- Watanabe M (1989) The appropriateness of behavioral responses coded in post-trial activity of primate prefrontal units. *Neurosci Lett* 101:113–117.
- White IM, Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126:315–335.