Systems/Circuits

Performance Monitoring in Monkey Frontal Eye Field

Tobias Teichert, Dian Yu, and Vincent P. Ferrera
Department of Neuroscience, Columbia University, New York, New York 10032

The frontal eye fields (FEF) are thought to mediate response selection during oculomotor decision tasks. In addition, many FEF neurons have robust postsaccadic responses, but their role in postchoice evaluative processes (online performance monitoring) is only beginning to become apparent. Here we report error-related neural activity in FEF while monkeys performed a biased speed-categorization task that enticed the animals to make impulsive errors. Twenty-three percent of cells in macaque FEF coded an internally generated error-related signal, and many of the same cells also coded task difficulty. The observed responses are primarily consistent with three related concepts that have been associated with performance monitoring: (1) response conflict; (2) uncertainty; and (3) reward prediction. Overall, our findings suggest a novel role for the FEF as part of the neural network that evaluates the preceding choice to optimize behavior in the future.

Key words: **performance monitoring; conflict monitoring; uncertainty; reward prediction; decision-making; error-detection**

Introduction

Neurons in the frontal eye field (FEF) respond before and after the execution of saccadic eye movements (Goldberg and Bruce, 1990). Presaccadic activity in the FEF has been studied in great detail and is known to carry visual, memory, and motor information (Bruce and Goldberg, 1985). However, it has long been observed that FEF neurons also have robust postsaccadic activity (Goldberg and Bruce, 1990). Because postsaccadic responses have been studied in less detail, our understanding of FEF function remains incomplete. The current experiment addresses this gap and focuses on postsaccadic FEF activity in a biased sensory decision task.

Presaccadic activity in the FEF plays an important role in the selection of appropriate responses in oculomotor decision tasks (Thompson et al., 1997; Kim and Shadlen, 1999; Bichot et al., 2001). A decision process does not end with the execution of a motor response, but the role of the FEF in subsequent evaluative processes (online performance monitoring) is only beginning to become apparent. Recent reports have found evidence of postsaccadic task difficulty signals in near-threshold sensory detection tasks (Ding and Gold, 2012; Seo et al., 2012). Such signals may reflect the signal-to-noise ratio of sensory information that determines task difficulty. Alternatively, these signals may be explained by more cognitive accounts, such as reward expectation, decision uncertainty, or response conflict. Interestingly, all of these options predict that these signals should differ in correct and error trials. Unfortunately, the two studies by Ding and Gold and Seo et al. did not analyze the data by outcome. Hence, an effect of error may have been present but remained undetected.

To date, there is no report in the literature that directly links the FEF to internally generated choice-error signals. Hanes et al. (1998) did not find error-related signals in FEF neurons for errors of commission in the stop-signal reaction time task. The same task reliably elicits error-related responses from neurons in the supplementary eye fields (SEFs; Stuphorn et al., 2000) and anterior cingulate cortex (ACC; Ito et al., 2003; Taylor et al., 2007). Cranial ERP recordings in monkeys performing the same task have confirmed the generators of the error-related negativity (Ne or ERN) in frontocentral regions, including the cingulate cortex but not the FEF (Godlove et al., 2011). Although FEF responses have not been linked to choice errors, recent studies have linked FEF activity to internally generated motor-error signals that may drive corrective saccades toward inferred target position (Murthy et al., 2007; Ferrera and Barborica, 2010).

The current study was designed to determine whether FEF neurons signal choice errors in a paradigm that was specifically designed to examine postdecision activity. To that aim, macaque monkeys performed a reward-biased speed-categorization task with delayed feedback that was optimized to elicit impulsive errors and thus allow for the emergence of internally generated error-related signals. Our results show error-related responses in 23% of FEF cells. Furthermore, the magnitude of these error-related signals is strongly correlated with the difficulty of the decision. These observations suggest that the FEF plays a role in postdecision outcome evaluation and argue against a strict segregation of movement planning in lateral prefrontal cortex and performance monitoring in medial PFC.

Materials and Methods

Subjects. Subjects were two male macaque monkeys (monkeys L and F). At the time of the experiments, the animals were between 9 and 12 years

Received Aug. 29, 2013; revised Nov. 8, 2013; accepted Dec. 10, 2013.

Author contributions: T.T. and V.P.F. designed research; T.T. and D.Y. performed research; T.T. analyzed data; T.T. and V.P.F. wrote the paper.

This work was supported by National Institutes of Health Grant MH059244 (V.P.F.) and German Research Foundation Grant TE819/1-1 (T.T.). We thank Girma Asfaw for help with surgical procedures and animal care. We acknowledge Stephen Dashnaw for his help with the MR imaging. We thank the members of the Ferrera laboratory (Jack Grinband, Marianna Yanike, Franco Pestilli, and Gaurav Patel) and the Mahoney Center for their helpful comments and discussions. Finally, we thank the anonymous reviewers for their helpful and constructive criticisms.

The authors declare no competing financial interests.

Correspondence should be addressed to Tobias Teichert, Columbia University, 1051 Riverside Drive, Unit 87, New York, NY 10032. E-mail: tobias.tshirt@gmail.com.

DOI:10.1523/JNEUROSCI.3694-13.2014

 0270-6474/14/341657-15$15.00/0

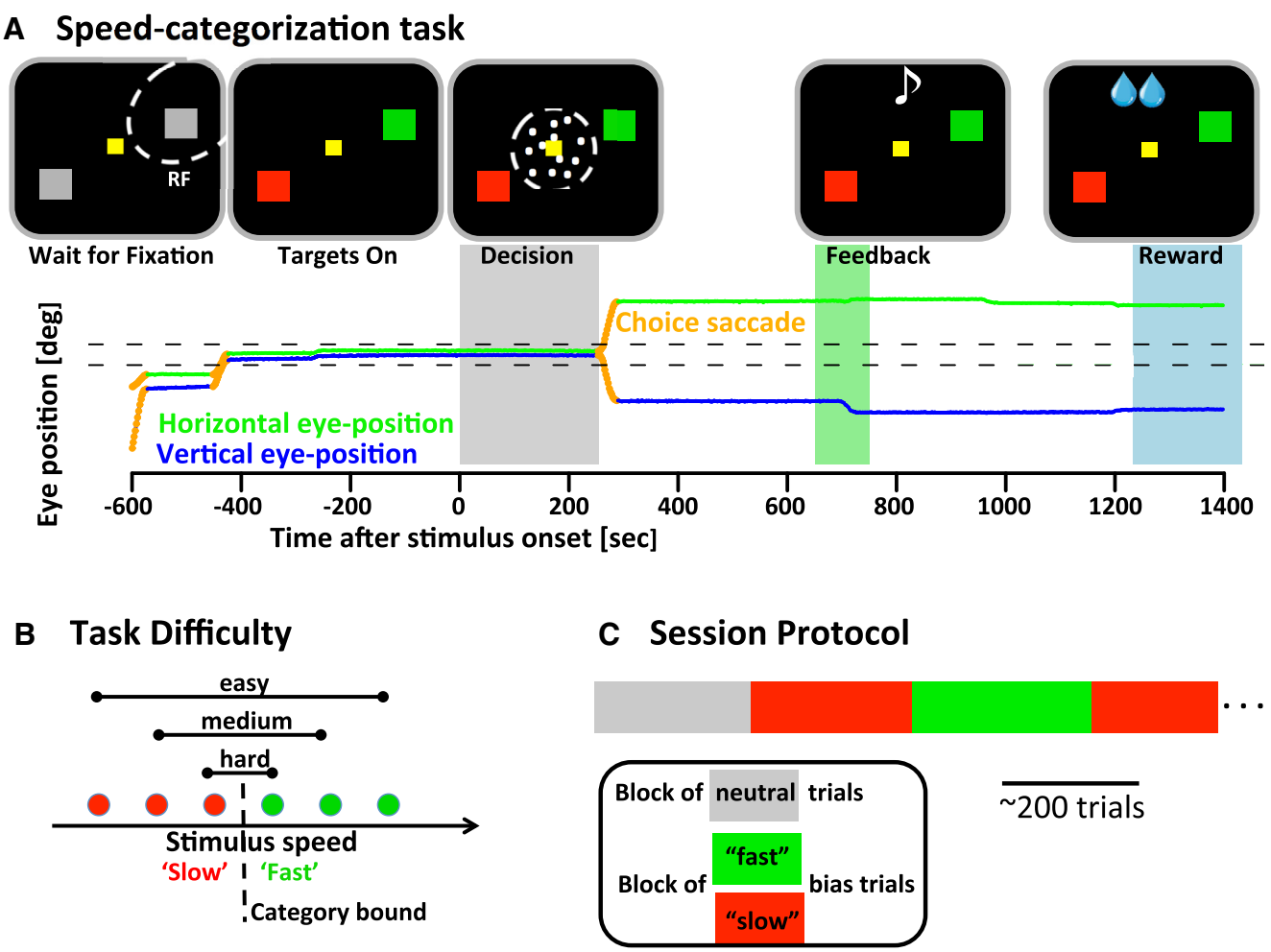

Figure 1. ***A***, Macaque monkeys were trained to categorize the speed of moving random dots as either slow or fast relative to an arbitrary yet constant criterion speed that was learned by trial and error. Monkeys categorized speeds by making a saccade to the red and green targets to signal slow and fast stimulus speeds, respectively. The position of the red/green targets inside/outside the receptive field was randomized to distinguish category choice from saccade direction. Animals were free to indicate their choice any time after the onset of the motion stimulus. Auditory feedback was delivered after 400–1200 ms of continued fixation of the choice target. Reward was delivered after additional 300–600 ms fixation delay. ***B***, Six different stimulus speeds were presented spaced equidistant around the category boundary. Based on the distance from the category bound, speeds were classified as "easy," "intermediate," and "difficult" (***C***). In blocks of 100–300 trials, animals were biased either toward the red or green target by giving more reward for correct saccades to targets of this color. The current reward schedule was never indicated but had to be inferred based on rewards received on previous trials.

old and weighed between 8 and 11 kg. The treatment of the monkeys was in accordance with the guidelines set by the U.S. Department of Health and Human Services (National Institutes of Health) for the care and use of laboratory animals. All methods were approved by the Institutional Animal Care and Use Committee at Columbia University and the New York State Psychiatric Institute. Monkeys were prepared for experiments by surgical implantation of a post for head restraint and a recording chamber to give access to cortex. Eye position was recorded using a monocular scleral search coil (Judge et al., 1980). Subjects had extensive experience (>12 months) with the psychophysical tasks described below.

Experimental setup. The animals were seated in an upright primate chair while head movements were restrained by a surgically implanted head post. Visual stimuli were generated and controlled by a CRS VSG2/3F video frame buffer. Stimuli were displayed on a 60 Hz CRT monitor (1280 × 1024 pixels) that was placed at a distance of ~50 cm. Some of the later sessions of animal L were recorded with a 60 Hz LCD monitor. Eye position was recorded with a scleral search coil (CNC Engineering) and digitized at a rate of 500 Hz.

Visual stimuli. Visual stimuli consisted of random dot kinematograms and small squares of different colors that served as fixation and eye movement targets. The random dot kinematograms comprised 200 dots of 2 × 2 pixels each. The dots moved coherently in one direction. Dots had infinite lifetime but were replaced as soon as they left a 5° circular aperture around the fixation spot in the center of the screen. The fixation spot was a yellow square with a width of 0.5°. The saccade targets were luminance-matched gray, red, and green squares with a width of 3°.

Biased speed-categorization task. Monkeys categorized the speed of the random dot kinematogram as either fast or slow by making a saccade to the green and red targets, respectively (Fig. 1*A*,*B*). In blocks of 100–300 trials, the reward context of the task changed (Fig. 1*C*). The changes were not overtly signaled to the animal but had to be inferred from past trials based on the change in reward contingency. In the neutral condition, animals received the same amount of reward for a correct saccade to either the red or green target. In the "slow/fast bias" condition, the animals received more reward for a correct saccade to the red/green target (Teichert and Ferrera, 2010).

After the subjects directed their gaze to the fixation target in the center of the screen, two gray placeholder targets appeared on diametrically opposed sides of the fixation spot. After a random delay (500 ms + truncated exponential distribution with a time constant of 400 ms and maximum duration of 1200 ms), the two gray placeholders turned red and green, respectively ("target onset"). The position of the targets on each side of the fixation point was chosen randomly. After an additional random delay (200 ms + truncated exponential with a time constant of 500 ms and maximum value of 1500 ms), a random dot kinematogram appeared in a 5° circular window around the fixation target ("stimulus

onset"). On approximately half of the recording sessions, the delay between target and stimulus onset was zero, i.e., the dots appeared at the same time as the targets changed color from gray to either red or green. The subjects were free to signal their choice at any time after stimulus onset with a saccade to the corresponding target. For each subject, the stimulus speed of a particular trial was drawn randomly from a set of six predetermined speeds. The speeds were spaced symmetrically around the cutoff speed of 5.5 pixels/frame. The range of speeds was adjusted for each subject individually to account for differences in performance. The direction of dot motion was aligned with the axis orthogonal to the axis defined by the two response targets. Both directions of motion along the orthogonal axis were used, and the direction was chosen randomly on each trial. The direction of motion was irrelevant for the task and had to be ignored by the animals.

After a valid choice saccade (see below) and a random delay (400–600 ms uniform delay for monkey L, 600–1200 ms uniform delay for monkey F), auditory feedback was delivered. A high tone (880 Hz) indicated a correct response, and a low tone (440 Hz) indicated a wrong response. The subjects were required to keep fixating the target to receive the fluid reward associated with the correct response. The delay between the auditory feedback and the fluid reward was uniformly distributed between 350 and 650 ms.

Eye movements during the task were restricted to a single saccade to one of the targets after stimulus onset (choice saccade). If the eye left the fixation window before stimulus onset, the trial was aborted and marked as fixation break. Similarly, if the eye left the fixation window after stimulus onset and failed to refixate on one of the two saccade targets, the trial was considered incomplete. Fixation break or incomplete trials were repeated with identical parameters on the next trial until it was completed. Fixation break or incomplete trials as well as trials after such trials were not included in any of the analyses.

On a small fraction of trials, the animals made a valid choice saccade followed by another saccade to the other target in an effort to revise the original decision (revised trials). Revised trials were never rewarded and indicated by a very distinctive auditory event: an upward followed by a downward sinusoidal sweep. Revised trials were included in some of the behavioral analyses. However, they were not included in any of the main electrophysiological analyses. One follow-up analysis used firing rate in revised trials.

Single-cell recordings. The animals had plastic (two hemispheres) or metal (one hemisphere) recording chambers implanted over the left and/or right FEF (stereotactic coordinates: ~25 mm anterior, ~15 mm lateral). The locations of the two plastic chambers were verified using MRI with the recording chambers in place. The location of the metal chamber (left hemisphere for monkey L) was estimated based on the relation to the contralateral plastic chamber and stereotactic coordinates during implantation of the chamber.

Single-cell activity was recorded with glass-coated tungsten electrodes (Alpha Omega) with impedances between 0.5 and 2 MΩ measured at a frequency of 1 kHz. Raw signals were amplified, digitized, and high-pass filtered. For different recording sessions, we used either a Plexon or FHC preamplifier, in combination with an FHC or an Alpha Omega recording system. Potential waveforms were identified by threshold crossings and a time–amplitude window discriminator. Extracellular action potentials were converted into digital pulses that were sampled by the computer with 0.01 ms temporal resolution. In most cases, waveforms were stored for offline spike sorting using wavClus. However, offline spike sorting was necessary only in very few cases because recording was mostly restricted to cells that could be clearly isolated online. Later recording sessions used multiple simultaneous electrodes in combination with offline spike sorting. In this case, spike sorting was performed with WaveSorter (sourceforgenet/projects/wavesorter).

Several methods were used to reconstruct the locations of the recording sites using structural MRI, recording grid coordinates, and microdrive readings. Figure 2 shows the reconstructions of the recording locations relative to the arcuate and principal sulci. The anatomical landmarks were reconstructed by rotating the MR image using FSL (for FMRIB Software Library; Smith et al., 2004) such that the *z*-axis corresponded to the direction of the electrode penetration. Furthermore, the image was aligned such that *x*- and *y*-axis corresponded to the axes of the electrode recording grid. The openings and the fundi of the sulci were traced in the aligned MR image and projected onto the *x–y* plane of the grid (Fig. 2). In the case of the two plastic chambers, we estimate that the reconstructions are accurate within ±1 mm. In the case of the metal chamber, the accuracy is lower because of the lack of an MR image with the chamber in place. Given that similar targeting routines and coordinates were used to place the metal chamber, we estimate the accuracy at ±2 mm. In addition to the anatomical reconstructions, electrical stimulation (300 Hz, biphasic pulses, current at <50 μA) was used to verify recording locations in the FEF by evoking involuntary saccades. Physiological properties, including the presence of high-frequency presaccadic bursts of action potentials, spatially selective visual, memory-delay and presaccadic activity, postsaccadic activity with spatial tuning opposite to the presaccadic activity, and, finally, the presence of cells with purely movement-related activity. All of these properties are characteristic of the FEF. Based on all of these criteria, most of the recording sites could be attributed to the FEF. However, it is possible that some sites were not in the FEF but adjacent periarcuate cortex, including area 46. Four recording sites in monkey F were posterior and medial to the arcuate sulcus. These cells were recorded in ventral premotor cortex (vPM) that is also know to have oculomotor responses similar to the FEF (Fujii et al., 1998). By default, these cells were included in all analyses. However, excluding these four cells did not have a meaningful effect on the main results.

In addition, it is important to note that we did not preselect neurons to match specific criteria, such as strong and spatially selective visual and/or memory activity. Hence, although most cells were recorded in the FEF, not all of them exhibit textbook-example direction-selective response properties.

Overall, we recorded from >200 cells during the course of the experiment, but a large fraction of cells was never considered for analysis because the isolation was lost before a reasonable number of trials could be collected. A total of 174 cells entered a prescreening process. Fifty-nine of these cells were not fit for additional analysis because of bad/nonstationary isolation or low number of trials (<200) after excluding periods of nonstationary isolation. Note that prescreening occurred before response properties of the cells were analyzed. Hence, the prescreening did not bias the fraction of significant cells for the different predictors.

Analysis of neural data. Neural data were analyzed using multiple linear regression models (Draper and Smith, 1966; Kim and Shadlen, 1999) with spike times aligned to the onset of the choice saccade. The main analysis focused on spike rate in a window from 0 to 300 ms after saccade onset. Spike count was modeled as a function of several independent/experimental and dependent/behavioral variables (Table 1). Because of the inclusion of behavioral variables that cannot be controlled experimentally, the resulting design was inherently unbalanced. The unbalanced design was addressed using type II sums of squares. The analysis was implemented in R (R Development Core Team, 2009) using the *lm* function and the *Anova* function from the *Car* package (Weisberg and Fox, 2010).

The linear regression models included up to 24 predictors and some of their interaction terms. Most importantly, the model included two predictors related to online performance monitoring (difficulty, error), as well as their interaction. Difficulty was coded as a numeric regressor and could take on one of three values (−1, 0, or +1, for easy, medium, and hard trials), and error was coded as a binary regressor, with 1 corresponding to an error and 0 to a correct trial. Another set of variables accounted for task-related changes in firing rate (target configuration, dot direction, dot speed, stimulus direction, reward context, action value, chosen value). A number of predictors modeled the effect of the choice saccade (saccade direction, reaction time, peak saccade velocity, saccade duration, saccade amplitude, velocity/amplitude), as well as their interaction with saccade direction. Finally, we added seven regressors to account for different types of subsequent eye movements [type 1 saccades: (1) number, (2) amplitude, (3) peak velocity, (4) horizontal displacement, (5) vertical displacement; type 2a saccades: (6) latency within error condition; type3 saccades: (7) number within direction]. Unless otherwise mentioned, statistical significance was assessed using an α value of 0.05. For $\alpha = 0.05$, the probability of observing >11, 13, and 15 significant

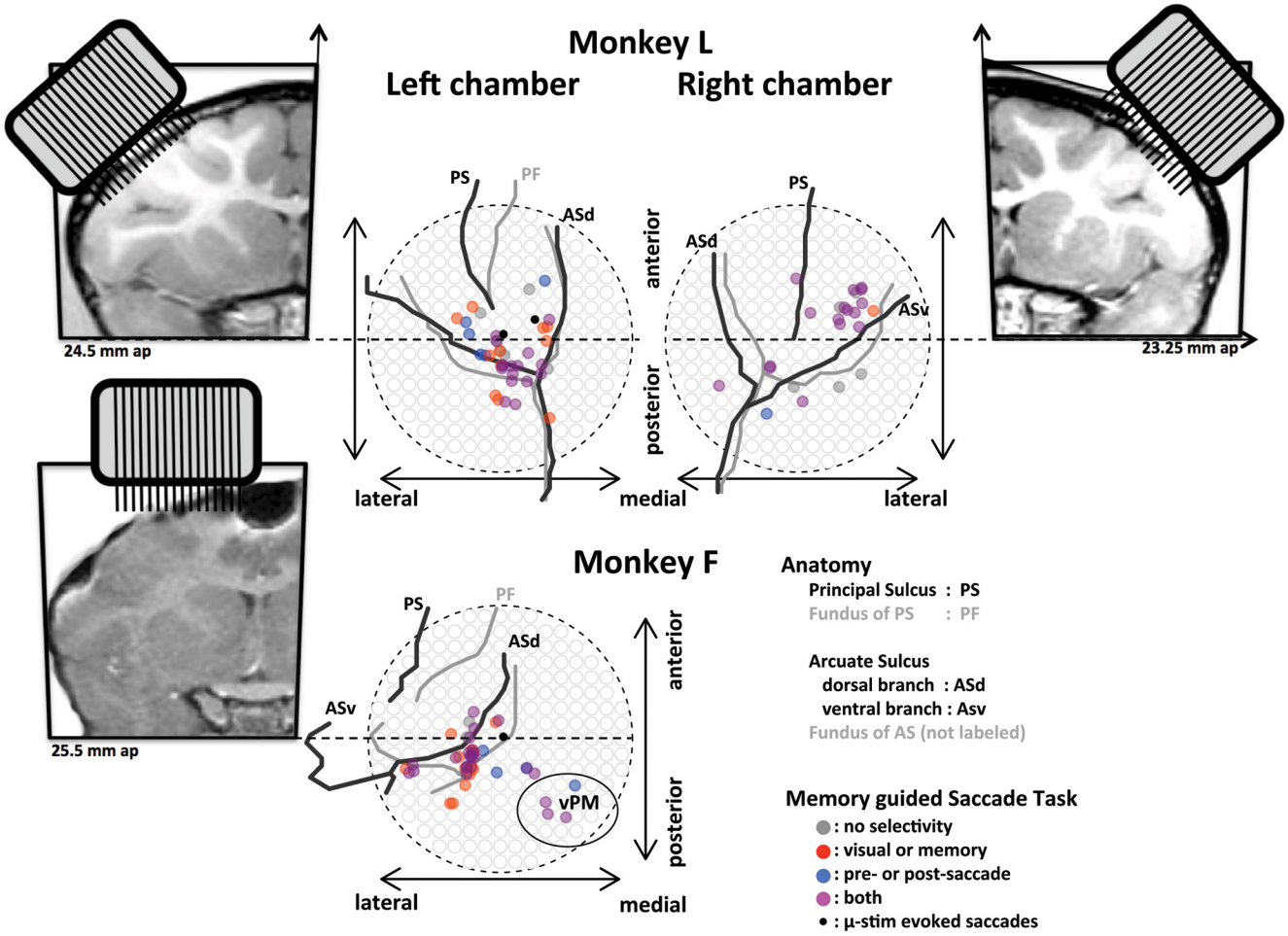

Figure 2. Recording sites were reconstructed based on stereotaxic MR images, chamber implantation coordinates and angles, MR images with plastic chambers in place (monkey L, right chamber; monkey F, left chamber only), and grid coordinates of the recording sites. Stereotaxically aligned coronal images were taken at the approximate anteroposterior position corresponding to the middle of the chamber. The diagram inset visualizes the tilt of the recording chamber. Sulci were reconstructed to resemble a view point along the electrode penetration pathway. The two chambers of monkey L were implanted at an angle approximately perpendicular to the cortical surface. As a result, the openings (black) and the fundi of the sulci (gray) are approximately overlapping, allowing a continuous penetration along the banks of the arcuate sulcus. The chamber in monkey F was implanted vertically. Hence, the openings of the sulci are more lateral than the fundi. MR image of monkey F with the chamber in place had a low signal-to-noise ratio and suffered from artifacts attributable to cortical screws. Hence, the reconstruction of the openings of the sulci is less accurate and was partially based on a extrapolation from a previous MR image without the chamber in place. In two of the chambers, microstimulation was used to verify the FEF by evoking microsaccades with a stimulation current of <50 μA (black dots). The response properties of neurons in a standard memory-guided saccade task were visualized using different semitransparent colors (see legend). Similar profiles of selectivity toward visual/memory (red) and motor (blue) were found in each recording location. The majority of the recording sites could be localized to the FEF. However, some sites were positioned in the surrounding periarcuate cortex, including area 46. Four recording sites in monkey F were reconstructed to vPM. Excluding these cells from the analysis had no qualitative influence on our results.

neurons of a total of 115 by chance alone was estimated at <0.05, 0.01, and 0.001, respectively.

The analysis excluded outlier trials defined on the properties of the choice saccade and the subsequent fixational eye movements. A trial was excluded if the duration of the choice saccade exceeded a fixed criterion value (60 ms for monkey L, 120 ms for monkey F). A trial was excluded if we detected more than one type 1 fixational saccade. Furthermore, we excluded trials if amplitude or peak velocity of the type 1 saccade exceeded 1.25 times the value of the 95th percentile.

The main results of the study were replicated using four different sets of linear models that included different sets of predictors (Table 1). The "full model" included all regressors, and the "base model" dropped all eye movement-related predictors except saccade direction and reaction time. The "reduced model" included only the variables related to error and difficulty, as well as saccade direction. Finally, we used a dynamic model ("drop model") that included all regressors in the full model if their p value was <0.4. In addition, the drop model always included the core variables, i.e., the variables in the reduced model. The drop model seemed to provide the best compromise between overfitting on the one hand and not correcting for potentially important confounds on the other. Hence, the body of the paper reports these values. Results of the other three models are consistent and are reported in Table 1.

To visualize mean firing rate as a function of error and difficulty, we normalized the raw firing rate. To normalize firing rates, we (1) subtracted out the effects of all predictors in the drop model, except the variables to be visualized, i.e., error and difficulty. (2) We divided the resulting values by the SD. In unbalanced designs, this normalization process is important to reduce the effect of confounding variables.

Test for balanced error/difficulty signals. To study how error signals depend on saccade direction, we calculated a regression model that allowed us to extract independent parameter estimates for ipsilateral and contralateral errors. A cell was classified as having a balanced error signal if p values for ipsilateral and contralateral error both passed a particular statistical criterion and the two parameter estimates had the same sign. The statistical criterion of the individual tests was chosen such that the combined test maintained a statistical criterion of 0.05. For example, if both ipsilateral and contralateral error signals were significant at an in-

Table 1. List of all predictors in the four different types of the model

Predictor	Reduced model	Base model	Drop model	Full model
Online performance monitoring				
Error or error*direction	√*	√*	√*	√*
Difficulty or difficulty*direction	√	√	√	√
Error*difficulty	√*	√*	√*	√*
Task-related confounds				
Target configuration		√	(√)	√
Dot direction		√	(√)	√
Dot speed		√	(√)	√
Reward context		√	(√)	√
Action value		√	(√)	√
Chosen value		√	(√)	√
Choice saccade				
Saccade direction	√	√	(√)	√
Reaction time		√	(√)	√
Peak velocity*direction			(√)	√
Amplitude*direction			(√)	√
Duration*direction			(√)	√
Main sequence (velocity/amplitude)*direction			(√)	√
Type 1 saccade				
Number of Type 1 saccades			(√)	√
Amplitude			(√)	√
Peak velocity			(√)	√
Horizontal displacement			(√)	√
Vertical displacement			(√)	√
Type 2a saccade				
Latency of error saccade			(√)	√
Latency of correct saccade				
β (memsac)				
Type 2b saccade				
NA				
Type 3 saccade				
(Number of Type 3 saccades)*direction			(√)	√

A check mark indicates that a regressor is part of the model. A check mark with an asterisk indicates that this regressor is only present in models that include error trials. A check mark in parentheses indicates that the predictor was included only if it had a p values of <0.4 in the full model. An asterisk between to predictors indicates that both predictors as well as their interaction was included in the model.

dividual criterion of 0.05, then the combined test would be significant at a level of $0.05^2 = 0.0025$. This is the case as long as the two tests are independent. Taking into account the fact that the test only includes cells if the their β values have the same sign, this reduces the effective statistical criterion to $0.05^2/2 = 0.00125$. To maintain the desired α value of 0.05, we defined the p value of our new balanced error test as the square root of the maximum of the p values for the ipsilateral and contralateral errors divided by 2. To ensure that only cells with equal sign are included, we set p values to 1 if the two parameter estimates had different signs.

In summary, this new p value can be compared with any α value to provide a test at this statistical criterion. Note that this test is stricter than the standard (generally accepted) test in that it only pulls out cells that respond similarly to both ipsilateral and contralateral errors. A corresponding test for balanced difficulty was defined along the same lines.

Results

Behavioral evidence of online performance monitoring

We recorded choice behavior and neural activity while monkeys categorized the speed of a moving random dot pattern as either slow or fast by making a saccade to a red or green target, respectively (Fig. 1*A*). In blocks of 100–300 trials, the reward context of the task changed, favoring either the slow or the fast choice category (Fig. 1*B*). The biased reward schedule enticed animals on some trials to choose the category with the larger reward even if this led to an incorrect, and hence unrewarded, response. The error rate approximately doubled when the reward schedule favored the wrong category (Fig. 3*A*). On each trial, the animals were allowed to make a single saccade to one of the two choice targets and were required to maintain fixation to receive auditory feedback >400 ms after the choice saccade and reward in the event of a correct response. Trials were classified as "revised" if animals broke fixation before receiving auditory feedback and if the second saccade was directed at the other choice target.

Revised trials were rare but occurred reliably in both animals (monkey L, 5.6 ± 3.6%; monkey F, 0.8 ± 0.5%, mean ± SD over recording sessions). Revisions were observed almost exclusively after an initially wrong choice (Fig. 3*B*; monkey L, 92.6 ± 8.4%; monkey F, 95.4 ± 12.6%). This implies that animals have access to a neural signal that identifies response errors independent of and before external feedback. It is possible that revisions occur on trials when the error signal is particularly strong. However, because the task actively discourages animals from revising their choices, it is also possible that revisions occur when control mechanisms are weak. Hence, it is possible that an internally generated error signal is present even on trials in which the animals did not revise the error. This is important, because it suggests that we can use unrevised error trials to look for the neural correlate of error detection.

Neural activity in the FEF during online performance monitoring

Figure 4*A* shows the responses of a single FEF example neuron for correct and (unrevised) error trials during the biased speed-categorization task. Responses were significantly stronger after error-saccades in the preferred or nonpreferred direction. Importantly, the neural error responses were observed before and thus independent of auditory feedback that indicated whether a response was correct. The same neuron also responded more strongly after difficult (correct) trials (Fig. 4*B*). Note that, in our task, difficulty is not correlated with stimulus strength (Ding and Gold, 2012; Seo et al., 2012) but is a highly processed signal that arises from comparison of stimulus speed to an arbitrary and abstract category bound (Fig. 1).

Before analyzing error-related and difficulty responses of the population, we outline the controls that were performed to distinguish cognitive choice error signals from less interesting accounts related to eye-movement planning or execution.

Test for balanced error/difficulty signals

The standard test to detect error cells compares mean firing rates of all error trials to the firing rates in all correct trials. This is an adequate test if all error trials are of the same type. However, in the current study, there are at least two distinct types: (1) errors after ipsilateral saccades and (2) errors after contralateral saccades. These types need to be considered separately because error signals in an oculomotor structure such as the FEF may be related to the planning of corrective saccades. Thus, we would expect the error responses to depend on the direction of the corrective saccade (opposite to the direction of the choice saccade). Hence, the standard test may pull out cells that respond exclusively after ipsilateral errors but have no or even weak opposite responses for contralateral errors. The classification of such cells as choice error cells would be misleading. In the following, we test whether or not error signals in the FEF depend on saccade direction and, based on this analysis, define a better, i.e., more stringent, test that labels error cells only if error responses are present for both saccade directions.

To study how error signals depend on saccade direction, we calculated a regression model that allowed us to extract independent parameter estimates for ipsilateral and contralateral errors. Figure 5*A* plots the parameter estimates for contralateral errors as

a function of the corresponding value for ipsilateral errors. A significant fraction of neurons responded to ipsilateral (30 of 115 = 26%) as well as contralateral (18 of 115 = 16%) errors. We observed a robust and significant correlation between ipsilateral and contralateral error signals (Pearson's test for correlation coefficients, $p < 10^{-5}$, $t = 5.08$, df = 113, $r = 0.43$). This correlation was present in both animals and highlights the overall bilateral nature of the error signal. Significant signals after both ipsilateral and contralateral errors were found in 11% of the neurons (13 of 115). This number is significantly higher than expected by chance (binomial test, df = 114, $p < 10^{-17}$) and higher than expected if ipsilateral and contralateral error signals were independent (χ^2 test for independence, $p < 10^{-5}$).

We took advantage of the lateralized error analysis to define a more rigorous criterion for error, which we call the "balanced" error signal (see Materials and Methods and Fig. 5*B*). Briefly, a cell was classified as having a balanced error signal if *p* values for ipsilateral and contralateral error both passed a particular statistical criterion and the two parameter estimates had the same sign. A corresponding test for balanced difficulty was defined along the same lines. For all the main analysis of the paper, we use this test for balanced error and difficulty signals. The use of balanced tests reduces the likelihood that the error-related signal is related to motor planning. However, it does not address certain confounds related to the eye-movement execution.

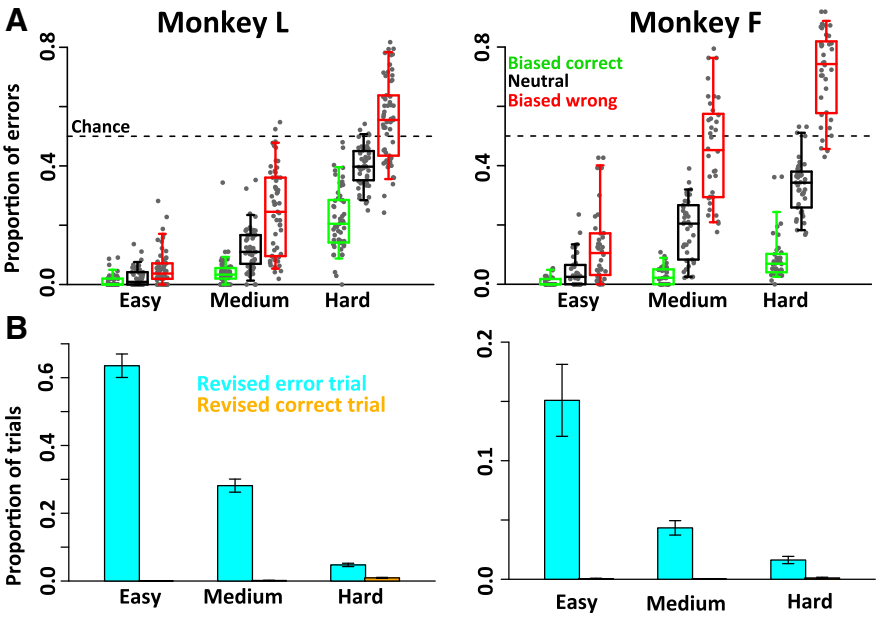

Figure 3. ***A***, Proportion of errors as a function of difficulty and reward bias (box-and-whisker plots indicate the 0.05, 0.25, 0.75, and 0.95 quantiles over recording sessions). Relative to the neutral condition (black), error rate approximately doubled when the reward schedule favored the wrong target (red). In contrast, error rate dropped when reward and stimulus speed favored the same response. ***B***, Proportion of revised trials as a function of difficulty and accuracy of the initial choice (mean ± SE over recording sessions). Both animals almost exclusively revised wrong choices and were more likely to revise if the trial was easy.

Potential artifacts I: dynamics of the choice saccade

The FEFs are primarily recognized as a motor area that is involved in the execution of saccadic eye movements. Hence, it is important to rule out that saccade execution may affect the observed neural responses to error and difficulty. Figure 6 plots raw and normalized reaction time as well as peak saccade velocity as a function of error and difficulty. These analyses show that reaction times are significantly shorter for error trials. Furthermore, they show that, for one of the animals, harder trials go along with shorter reaction times. This finding is unexpected because reaction times in many tasks and species increase with task difficulty. The inverse findings—at least for one of the animals—cannot be attributed to either species or task alone. Human subjects performing the same task show longer reaction times for difficult trials (our unpublished observations). Similarly, some but not all macaques show a moderate increase of reaction time with difficulty (our unpublished observations). Macaque monkeys performing the same task with manual rather than oculomotor responses exhibit the standard effect of longer reaction times for difficult trials (our unpublished observations).

Our results further show that normalized peak saccade velocity, i.e., after accounting for saccade amplitude and other potential confounds, depends on task difficulty. In both animals, we found higher peak velocities for more difficult trials. These findings suggest that it is important to include variables such as saccade amplitude, peak velocity, and reaction time (Table 1) as potential confounds into the model. The inclusion of these regressors in combination with the type II sums-of-squares minimizes any potential effect of these factors on the variables of interest.

Potential artifacts II: postdecision eye movements

To better understand and account for additional potential confounds that might explain the effects of error, we performed a detailed analysis of eye movements that occur after the choice saccade (also labeled first saccade). To that aim, we registered and classified all eye movements that occurred after the choice saccade. Based on the order of occurrence, we termed them second, third, and fourth saccade, etc.). A saccade was defined by a minimum amplitude of 1° and a minimum peak velocity of 40°/s. Figure 7 shows the latencies of these additional saccades as a function of time from the onset of the choice saccade.

For both animals, we observe the second saccade between 200 and 400 ms after the choice saccade. Overall, we detected more fixational eye movements for monkey L, indicating that animal F tended to exhibit more stable fixations. The set of third saccades is typically observed after the feedback (gray sigmoid line). The set of fourth saccades is mainly detected for monkey L.

A more detailed analysis reveals three types of postdecision saccades. (1) The first is small fixational eye movements that do not lead to a break in fixation. (2) Break saccades (>8°) that take the eye outside of the fixation window can be subdivided into two categories: (A) breaks that occur after the feedback; and (B) breaks that occur before feedback. Trials with saccades of type 2B were classified as "revised" if the second saccade brought the subject into the fixation window around the other fixation target. (3) We observed a third type of saccade for one of the animals: correction for initially hypometric choice saccades.

Type 1

The prevalence, amplitude, direction, and latency of type 1 saccades is very similar between correct and error trials. Nevertheless, we ob-

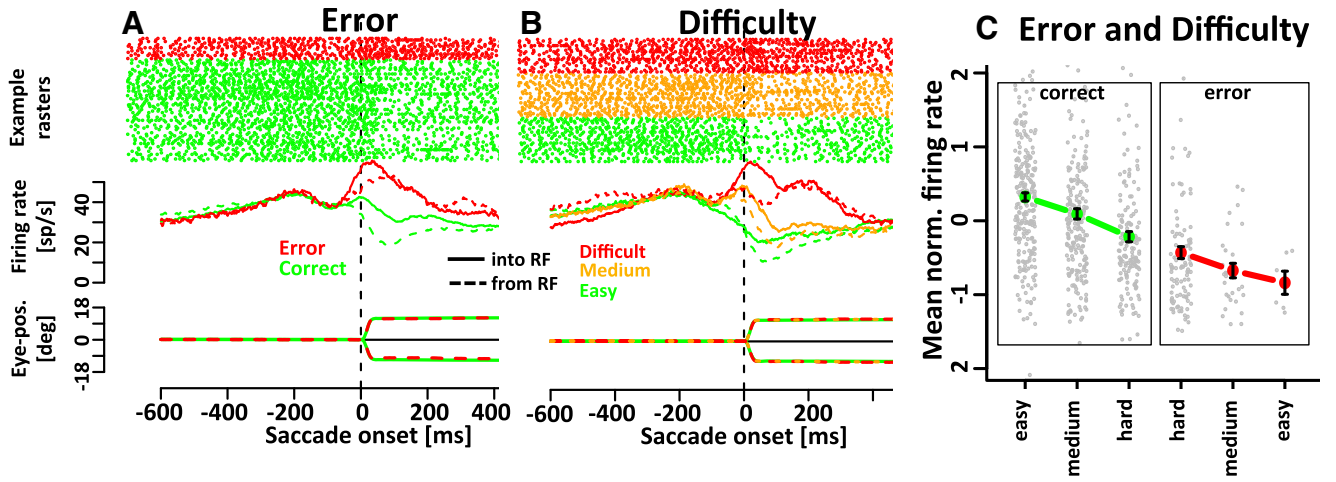

Figure 4. Spike raster (top), PSTH (medium), and horizontal eye position (bottom) aligned to saccade onset. An FEF example neuron increased firing rate after choice errors (***A***) and difficult (correct) trials (***B***) independent of saccade direction. ***C***, Normalized firing rate of a second example neuron as a function of error and difficulty in the time window from 0 to 300 ms after saccade onset (mean ± SE). RF, Receptive field.

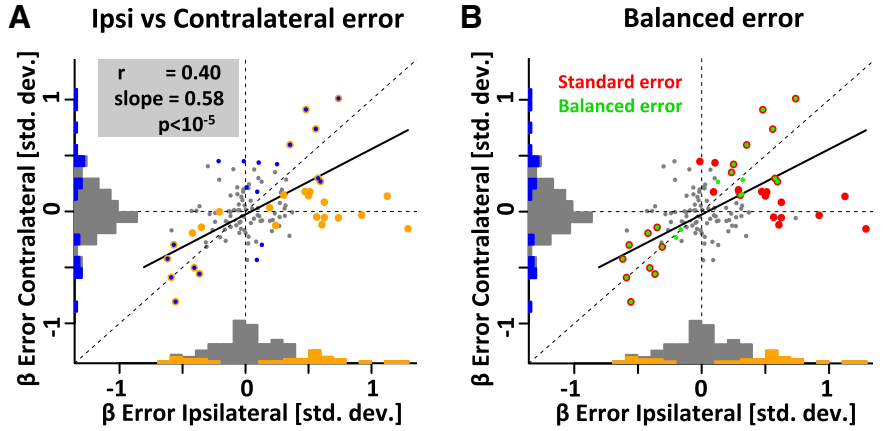

Figure 5. ***A***, Parameter estimates for ipsilateral and contralateral errors have a significant and robust correlation. ***B***, Comparison of error cells identified with the standard test (red) and the balanced error test (green). Note that the balanced test is more rigorous and excludes many cells that pass the standard test for error cells.

served small differences that needed to be addressed in the main analysis. To that aim, we detected the number of type 1 saccades in a window from 50 to 350 ms after onset of the choice saccade. As we analyzed spikes in a window from 0 to 300 ms, this accounted for type 1 saccades that occurred up to 50 ms after the spike window. Most trials exhibited zero or one type 1 saccade in this time period. To facilitate the analysis, we excluded trials that had more than one type 1 saccade in this window. We then used the latency, amplitude, peak velocity, displacement in *x*-direction, as well as displacement in *y*-direction as additional regressors in the model.

Type 2a (break saccades)

Type 2a saccades occur by definition after the auditory feedback that is presented at least 400 ms after the choice saccade. Because animals are required to maintain fixation after a correct response to collect the reward, the latency of type 2a saccades differs substantially between error (750 ms for monkey L; 1125 ms for monkey F) and correct trials (1750 ms for monkey L; 2150 ms for monkey F) (Fig. 7*E*,*F*). Because of these long latencies, we address type 2a saccades in the section about saccade planning below.

Type 2b

Type 2b saccades occur by definition only in revised trials. Because the main analysis focuses on error and correct trials, type 2b saccades were of no concern.

Type 3

Type 3 saccades occurred only in one animal as the result of a hypometric choice saccade that was then correct with the second saccade. Type 3 saccades occurred on error and correct trials with similar probabilities. However, to correct for their potential influence, we added a predictor to the model that included the occurrence of a type 3 saccade for each of the two choice directions separately.

Population analyses

To quantify the prevalence of error and difficulty signals in the population, we analyzed 115 neurons from three hemispheres of two animals using multiple regression analyses (210 to 1534 trials per cell, average of 609 trials per cell). Analysis was restricted to the time window from 0 to 300 ms after onset of the choice saccade and included regressors for potential confounds related to task parameters and oculomotor response (Table 1). Normalized firing rates after regressing out all potential confounds are shown for a second example cell as a function of task difficulty and error (Fig. 3*C*). Note that all analyses excluded revised trials to rule out the execution of corrective saccades as a potential confound. Furthermore, the effect of error cannot be explained in terms of task difficulty because its effects are regressed out before assessing the effect of error. Finally, the effect of difficulty itself was assessed in a model that only included correct trials to avoid the potential confound of difficulty and error rate.

Linear regression analysis revealed that 28% of cells modulated their activity in response to error (32 of 115). The more stringent criterion that required balanced error signals for both ipsilateral and contralateral errors revealed 27 of 115 significant cells (23%). Forty-one of 115 neurons (36%) modulated their activity as a function of task difficulty (27% using the more strin-

gent balanced difficulty criterion). For both balanced error and balanced difficulty, the number of significant neurons was significantly higher than expected by chance (Table 2). We also observed a robust correlation between the parameter estimates for error and difficulty (Pearson's test for correlation coefficients, $p < 10^{-7}$, $t = 6.05$, df $= 113$, $r = 0.49$; Fig. 8*A*), i.e., cells with stronger/weaker responses after choice errors tend to have stronger/weaker responses after difficult trials. Fifteen cells (13%) coded both balanced error and difficulty. This number is significantly higher than expected by chance (Table 1) or if error and difficulty were independent (χ^2 test for independence, $p < 0.001$).

Similar results were observed when the analysis window was restricted to the time window from 0 to 200 ms after saccade onset (Table 3). In contrast, none of the population results were significant when the analysis window ranged from −400 to −200 ms before saccade onset. This negative finding confirms that the analysis itself does not introduce artificially significant results. Furthermore, we confirmed the main results using three additional models that used either fewer or more regressors (see Materials and Methods and Tables 1, 3). The replication of the main results in models with fewer predictors shows that our results are not attributable to an overfitting of the data. The fact that we can replicate the findings in the full model with a larger number of potential confounds suggests that the neural responses are indeed correlated to the variables of interest and not a correlated confound.

The behavioral analysis showed that animals were more likely to try to revise an error in easy compared with difficult trials (Fig. 3*B*). One explanation is that neural error signals were stronger in easy trials. This was confirmed by showing larger firing-rate modulations for easier trials when mean normalized firing rate was calculated as a function of error and difficulty (Fig. 8*B*). A receiver operating characteristic (ROC) analysis confirmed that error and correct trials can be decoded more accurately on medium and easy trials (Fig. 8*D*). To enable the comparison of ROC values across cells that increase and decrease firing rate as a function of error, we recoded the ROC values such that an ROC value >0.5 corresponded to a firing-rate modulation in line with the sign of the parameter estimate of the error predictor β_{error}.

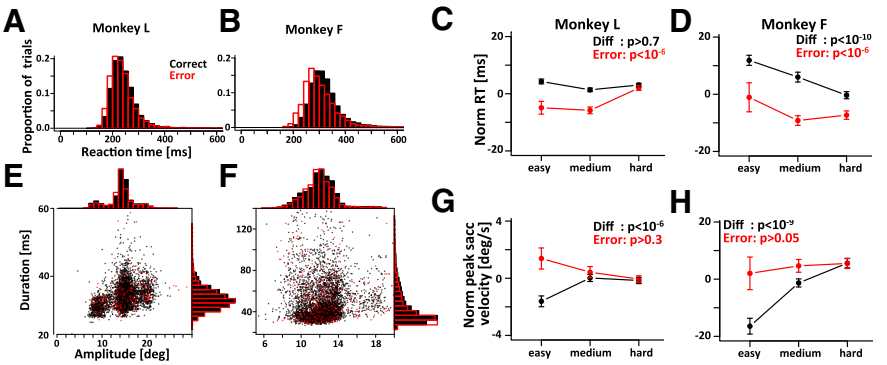

Figure 6. Reaction time and saccade velocity. ***A***, ***B***, Raw reaction time distributions for error (red) and correct (black) responses. ***C***, ***D***, Normalized reaction time (mean ± SE across recording sessions) as a function of task difficulty (*x*-axis) and outcome (error, red; correct, black). For each session, we fit reaction time as a linear function of task difficulty and outcome. One model included only main effects of task difficulty (β_d) and error (β_e). A second model included their interaction ($\beta_{d,e}$). To account for potentially unbalanced effects of time within the session, saccade direction, category choice, and the value of the chosen option, we included these variables as predictors in the model. For each session, normalized reaction time was defined as the parameter estimates for outcome and difficulty ($\beta_{d,e}$). Note that this is a purely additive normalization that subtracts mean and other potential confounds. For a balanced design, these values correspond to the corresponding group means minus the overall mean. The parameter estimates of error (β_e) and difficulty (β_d) were then fed into a two-tailed *t* test. The results of these tests is indicated in the top of each panel (difficulty, black; error, red). ***E***, ***F***, Raw main sequence of error and correct trials. ***G***, ***H***, Normalized peak saccade velocity (mean ± SE across recording sessions). Conventions as in ***C*** and ***D***. Because of the different nature of the dependent variable, we added saccade amplitude and reaction time as predictors to the linear model.

Relation to functional specificity

The FEF contains neurons with directionally selective visual, memory, saccade, or fixation-related activity. Cells with different response properties have been suggested to play different roles during oculomotor behavior (Hanes et al., 1998; Purcell et al., 2012). Hence, it is possible that difficulty and error-related responses are preferentially found in one specific functionally defined subtype. To test this assumption, we plotted the occurrence of error-related and difficulty signals as a function of the behavior of the cells in the memory-guided saccade task (Fig. 8*C*). Our results show that difficulty and error-related responses are observed with similar prevalence in all subtypes.

Relation to anatomical specificity

Cells in the FEF are arranged in a topographic continuum with the ventrolateral branch preferentially coding small-amplitude saccades and the dorsomedial branch large-amplitude saccades. Our recordings were centered on the FEF, and different recording chambers focused on different parts of the FEF anatomy. The right chamber in monkey L preferentially targeted the ventrolateral branch, and the left chamber in monkey F focused on the dorsomedial branch. Finally, the left chamber in monkey L focused on the most posterior part that defines the midpoint between the two branches. This variability within the FEF anatomy may also give rise to different response properties in the decision task. In addition, some of the recordings were made from the fringes of the FEF and potentially included other regions of the dorsolateral prefrontal cortex as well as vPM. Hence, it is important to test whether the difficulty and/or error-related signals were localized to different subregions within the FEF/dorsolateral prefrontal cortex. To that aim, we plotted the location of cells with difficulty and error-related signals onto the anatomical reconstructions (Fig. 8*F*). Cells of both types can be found in each anatomical subregion that was sampled. Hence, we conclude that such responses can be found in all of the FEF and speculate that they are also present in other regions of the dorsolateral prefrontal cortex, such as area 46, and even vPM.

Potential artifacts III: eye-movement planning

As outlined above, many of the potential eye-movement confounds were accounted for by adding them as predictors to the linear model. However, in the case of type 2a break saccades, this approach was not feasible. The predictor in question (latency of the break saccade) was highly correlated to our independent variable of interest (error). Because of the tight correlation between error and latency, the two predictors would soak up each other's variance in our linear model that uses type II sums-of-squares. Hence, we used a different approach. If the latency of the break

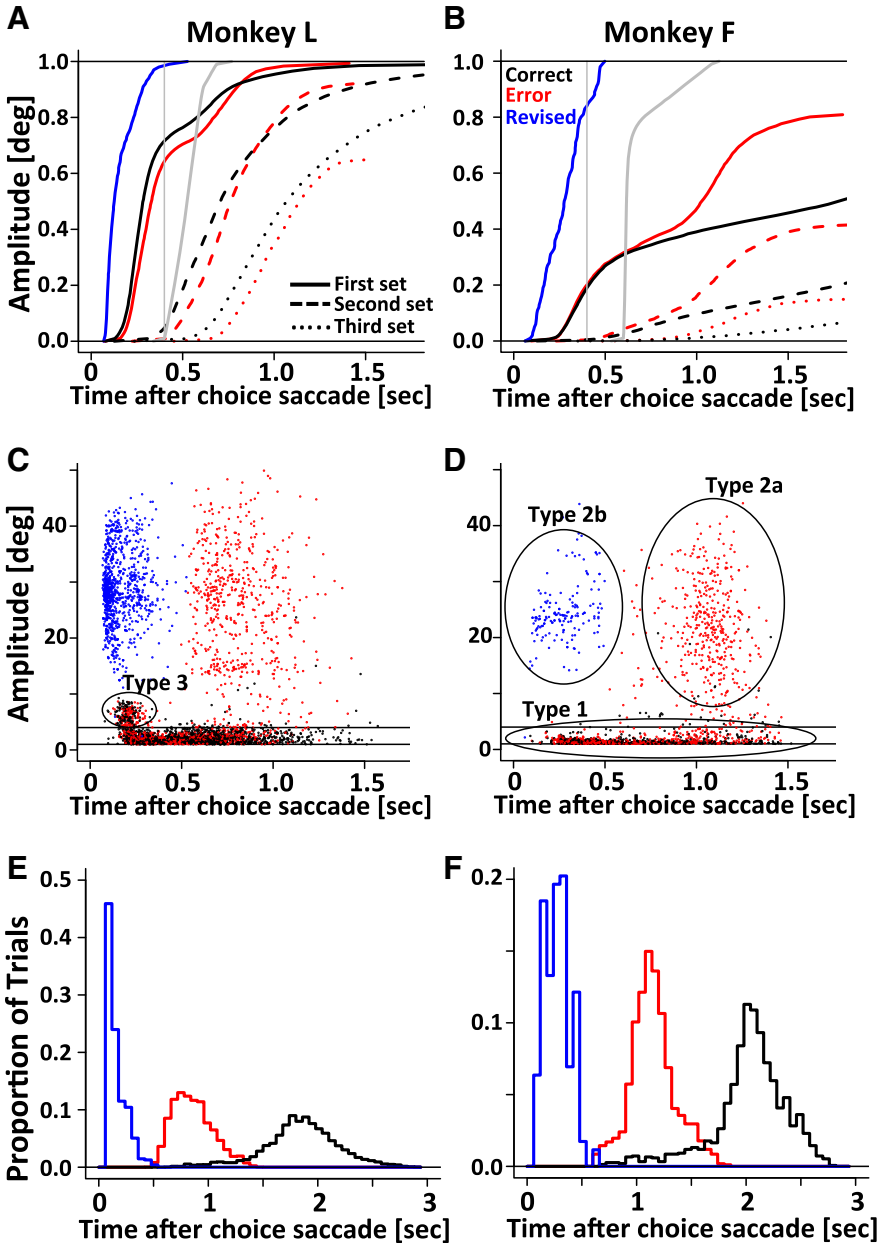

Figure 7. Analysis of postdecision eye movements. ***A***, ***B***, Timing of subsequent saccades as a function of time after the execution of the choice saccade separated by outcome (error, red; correct, black; revised, blue). The cumulative distribution plots reveal subsequent waves of eye movements separated by ~300 ms. ***C***, ***D***, Timing and amplitude of the first two sets of saccades for a random sample of trials. This figure reveals three different types of postdecision saccades: (1) type 1, small fixational eye movements; (2) type 2a and 2b, large saccades that move the eye out of the fixation window; and (3) type 3, compensation for initially hypometric choice saccades (only present in animal L). ***E***, ***F***, Density distribution of the latencies of type 2 break saccades indicate clear differences between all trial 3 types (blue, revised; red, error; black, correct). The differences between error and correct trials reflect the fact that subjects were required to maintain fixation to collect the reward after the delivery of positive auditory feedback. After negative feedback, the animals could not be encouraged to maintain fixation. Revised trials (type 2b) are by definition earlier than regular break saccades (type 2a).

saccade has an effect on firing rate, then this effect should be present within error trials whose latencies differ over a wide range. Hence, we added two regressors that coded the latency of the break saccade for error and correct trials separately. If the latency of the corrective saccade were to explain the effect of error, then we would expect the cells in question to have two properties: (1) the cells should have a significant effect of latency of the break saccade in error trials; and (2) the sign of the latency regressor should be compatible with the sign of the error regressor.

After thus defining cells whose error responses might be attributable to the latency confound, we excluded these cells and ran the population-analysis on the reduced sample. This did not change the main findings and suggests that saccade latency is not a likely explanation of the observed error signals (Table 4). Furthermore, we tested whether the parameter estimates of the latency regressor and error regressor were correlated. The notable absence of a significant correlation (Pearson's test for correlation coefficients, $t = -0.3285$, $df = 113$, $p = 0.7431$) provides additional evidence against the assumption that latency of the break saccade may explain the observed error signals. In fact, it suggests that even the cells that were excluded as potential confounds were excluded unnecessarily.

Finally, it is important to keep in mind that many cells code not only error but do so in conjunction with difficulty. For difficulty, the difference in latency of the break saccade is substantially smaller. Hence, the latency of the break saccade cannot account for the effect of difficulty on firing rate, which is on the same order of magnitude than the effect of error. The strong correlation between error and difficulty suggest that the two effects may be caused by a common mechanism (be it eye movements or something more interesting). If this is the case, the latency of the break saccades is a poor candidate because it may only explain the effect of error, not difficulty.

However, the entire effect of the break saccade might go beyond latency. In error trials, the latency of the break saccade was determined by the timing of the feedback. If the feedback occurred early, the break saccade occurred early. On correct trials, the latency of the break saccade depended on the time of reward delivery. Hence, the latency might not account for all aspects of saccade planning. It is possible that the subjects prepared the break saccade after an error and withheld its execution until the feedback confirmed the choice error. In this case, the latency of the break saccade cannot account for the saccade planning activity. To test this assumption, we turned to the memory-guided saccade task. We argue that the planning of the break saccade should be comparable with the saccade planning during the memory period of the memory-guided saccade task. Hence, we tested whether memory-period activity may explain the error-related activity. To that aim, we followed a similar approach as outlined for latency. First, we extracted parameter estimates for activity in the memory compared with the fixation period of the memory-guided saccade task. We then excluded all cells in which this parameter was significant and had the same sign as the parameter of the error regressor.

Table 2. Population statistics for both animals jointly and separately

	Both animals			Monkey L			Monkey F		
	n	%	*p*	*n*	%	*p*	*n*	%	*p*
Error	32 of 115	28	$<10^{-14}$	19 of 69	28	$<10^{-9}$	13 of 46	28	$<10^{-6}$
Difficulty	38 of 115	33	$<10^{-19}$	28 of 69	41	$<10^{-18}$	10 of 46	22	$<10^{-4}$
Balanced error	27 of 115	23	$<10^{-10}$	17 of 69	25	$<10^{-7}$	10 of 46	22	$<10^{-4}$
Balanced difficulty	31 of 115	27	$<10^{-14}$	22 of 69	32	$<10^{-11}$	9 of 46	20	$<10^{-3}$
Congruent balanced error and difficulty	13 of 115	10	$<10^{-21}$	9 of 69	13	$<10^{-15}$	4 of 46	9	$<10^{-6}$
Interaction of error and difficulty	18 of 115	16	$<10^{-13}$	13 of 69	13	$<10^{-5}$	5 of 46	11	$<10^{-1}$

Based on $\alpha = 0.05$, shown are the number of significant cells (*n*), the percentage of these cells in the population (percentage), and the probability (*p*) of observing this number by chance alone (binomial test, $p_0 = \alpha$ for the main effects, $p_0 = \alpha^2/2$ for the conjunction analysis of congruent balanced error and difficulty). Excluding the four vPM cells has no meaningful effect on the population statistics. Two of the vPM cells have a significant balanced error signal, one of them in conjunction with balanced difficulty.

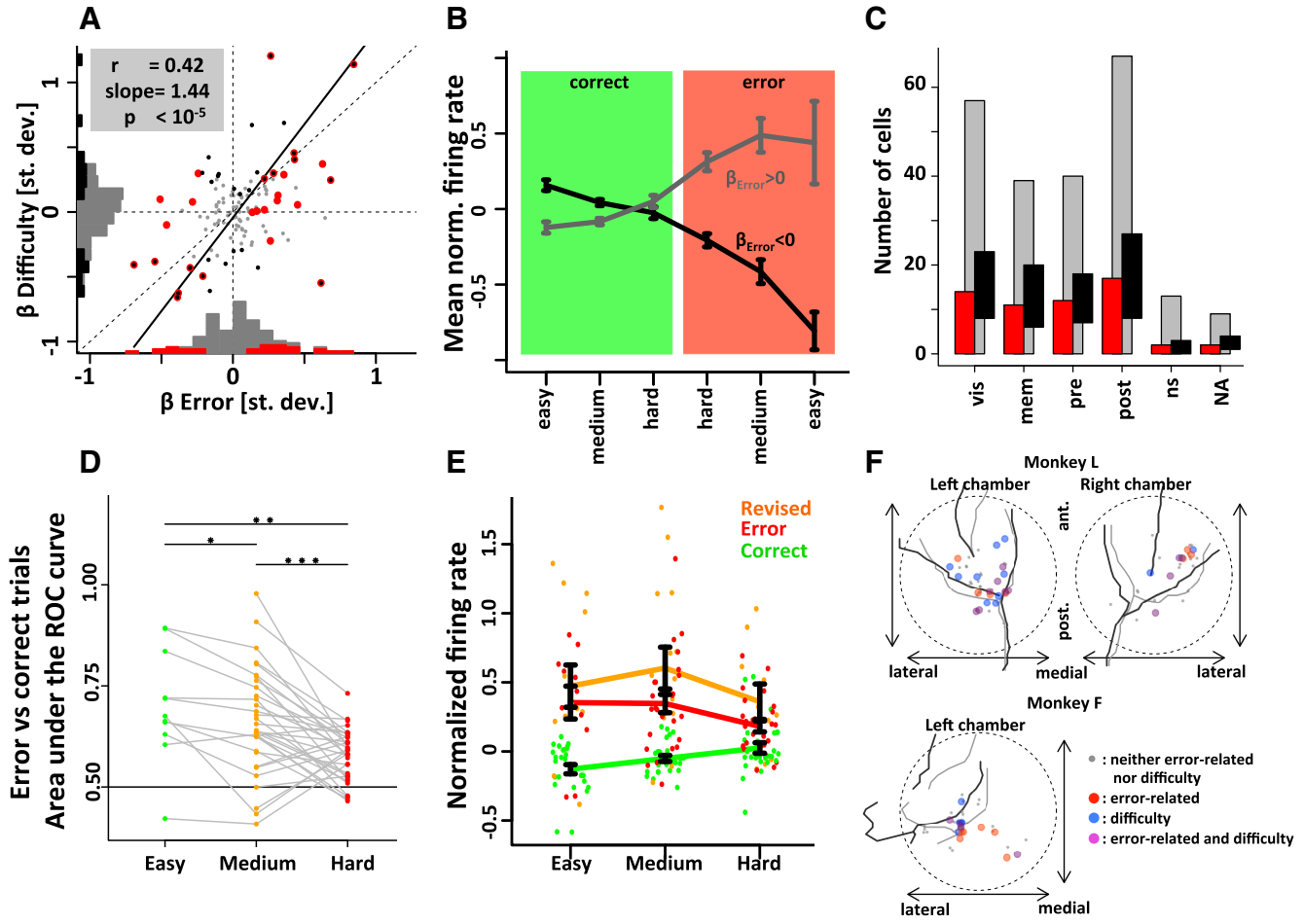

Figure 8. ***A***, Across the population, parameter estimates for error and task difficulty exhibit a significant positive correlation. Positive parameters indicate stronger firing after errors and difficult trials. ***B***, Normalized firing rates as a function of difficulty and error for 32 neurons with a main effect error or an interaction of error and difficulty (mean ± SE). The data were split in two groups depending on whether the parameter estimate for error was positive (gray) or negative (black). ***C***, Number of cells with error-related and difficulty signals as a function of direction selectivity in the memory-guided saccade task. Note that many cells had significant directionally tuned activity in more than one epoch. Hence, the numbers in the different columns do not add up to the total number of cells. Note that error-related and difficulty signals are present in all cell types, regardless of directional tuning. ***D***, Choice probabilities (area under the curve of ROC) as a function of difficulty for all neurons with a significant main effect of balanced error or an interaction of balanced error and difficulty. Firing rate was normalized by subtracting out all predictors of the "drop" model except difficulty, error, and their interaction. The residual firing rate was then divisively normalized to an SD of 1. In addition, normalization involved multiplication with the sign of the parameter of the error regressor to enable the averaging of cells with firing-rate increase and decrease in the same plot. Choice probabilities were computed within each level of difficulty. Errors on easy trials were rare, and here we included the ROC values only if more than four errors were observed for a particular difficulty. The bars in the top part of the panel indicate the outcome of a paired *t* test between the ROC values observed for different difficulty levels. *$p = 0.05$, **$p = 0.01$, and ***$p = 0.001$. Our results show that the separation of firing rates is stronger for easy compared with medium and difficult trials. ***E***, Normalized population responses for correct, error, and revised trials as a function of task difficulty. The sample includes cells that had a significant effect of balanced error or an interaction of error and difficulty and at least one revised trial. The analysis window included times from −200 to 0 ms before the execution of a revised saccade. For trials without a revised saccade, we randomly picked a time from one of the revised trials. A population-based one-sided one-sample *t* test showed that responses were significantly stronger before revised than unrevised errors ($t = 2.1395$, df $= 25$, $p = 0.02117$). Very similar results were found in the standard saccade-locked window from 0 to 300 ms after onset of the choice saccade. However, the effect did not reach significance in this window ($p < 0.1$). ***F***, Cells with error-related and difficulty signals were observed in all three animals without specific clustering to a particular part of the dorsolateral PFC. Two error/difficulty cells were reconstructed in the vPM. Excluding the vPM cells did not affect the main findings. ant, Anterior; post, posterior.

Table 3. The main results were replicated using four different sets of models that included different sets of predictors

Model	Error	Difference	Balanced error	Balanced difference	Congruent balanced error and difference	Interaction of error and difficulty	Slope (β_e/β_d)
0–300 ms							
Full	**28**	**34**	**22**	**22**	**10**	**16**	**1.43**
	24%	**30%**	**19%**	**19%**	**9%**	**14%**	$\mathbf{p < 10^{-5}}$
Drop	**32**	**38**	**27**	**31**	**13**	**18**	**1.44**
	28%	**33%**	**23%**	**27%**	**11%**	**16%**	$\mathbf{p < 10^{-5}}$
Base	**33**	**37**	**21**	**28**	**10**	**18**	**1.49**
	29%	**32%**	**18%**	**24%**	**9%**	**15%**	$\mathbf{p < 10^{-6}}$
Reduced	**35**	**37**	**19**	**33**	**18**	**20**	**1.53**
	30%	**32%**	**17%**	**29%**	**16%**	**17%**	$\mathbf{p < 10^{-5}}$
0–200 ms							
Full	**27**	**36**	**18**	**24**	**7**	**13**	**1.63**
	23%	**31%**	**16%**	**21%**	**6%**	**11%**	$\mathbf{p < 10^{-4}}$
Drop	**28**	**38**	**19**	**28**	**8**	**16**	**1.62**
	24%	**33%**	**17%**	**24%**	**7%**	**14%**	$\mathbf{p < 10^{-5}}$
Base	**29**	**38**	**20**	**28**	**8**	**21**	**1.61**
	25%	**33%**	**17%**	**24%**	**7%**	**18%**	$\mathbf{p < 10^{-5}}$
Reduced	**29**	**36**	**18**	**24**	**7**	**16**	**1.74**
	25%	**31%**	**16%**	**21%**	**6%**	**14%**	$\mathbf{p < 10^{-5}}$
−400 to −200 ms							
Full	6	7	3	8	0	8	−1.73
	5%	6%	3%	7%	0%	7%	Ns
Drop	7	4	4	8	0	9	−1.68
	6%	3%	3%	7%	0%	8%	Ns
Base	6	5	4	7	0	8	**−0.90**
	5%	4%	3%	6%	0%	7%	$\mathbf{p < 0.05}$
Reduced	4	7	5	6	0	7	**−1.05**
	3%	6%	4%	5%	0%	6%	$\mathbf{p < 0.05}$

All numbers indicate how many cells passed a corresponding statistical test at $\alpha = 0.05$. Bold indicate that this value is significantly higher than expected by chance (binomial test, $p < 0.01$, $p_0 = \alpha$; df = 114; $p_0 = \alpha^2/2$ for the conjunction analysis of error and difficulty). The same analyses were conducted using two additional time epochs from 0 to 200 ms and from −400 to −200 ms. The 0–200 ms epoch replicates the main findings and shows that similar signals can also be found closer to saccade onset. The −400 to −200 ms epoch provides a negative control and shows that, as expected, none of the findings can be explained by elevated false-positive rate of any of the tests. Ns, Not significant.

Table 4. The main results were replicated using two different criteria to exclude potential cells whose error-related activity may be an artifact of type 2a latency or saccade planning

Exclude potential artifact cells	Error	Difference	Balanced error	Balanced difference	Congruent balanced error and difference	Interaction error/difference	Slope (β^e/β_d)
None	32/115	38/115	27/115	31/115	13/115	18/115	1.44
	28%	33%	23%	27%	11%	16%	$p < 10^{-5}$
Type 2a latency	30/110	36/110	25/110	29/110	11/110	18/110	1.56
	27%	33%	23%	26%	10%	16%	$p < 10^{-4}$
Memory period activity	25/87	28/87	20/87	22/87	8/87	14/87	1.26
	29%	32%	25%	25%	9%	16%	$p < 10^{-4}$
Both types	23/83	26/83	18/83	20/83	6/83	14/83	1.39
	28%	31%	22%	24%	7%	17%	$p < 10^{-3}$

The main findings are not affected by excluding either or both of these potential artifact cells from the Drop model in the time-window from 0–300 ms. All effects were significant at the population level ($p < 0.05$).

As before, we excluded all potential artifact cells from the main analysis (Table 4). Again, the main results were not affected by the exclusion of these potential artifact cells. In addition, we calculated the correlation between the parameter estimate of the memory period and the error regressor. If saccade planning were to explain the differences between error and correct trials, we would expect the memory and error parameters to be correlated. The notable absence of such a correlation (Pearson's test for correlation coefficients, $t = -1.0654$, df = 113, $p = 0.289$) provides additional evidence that error responses cannot be explained in terms of saccade planning.

Revised trials

It is possible that the error-related signals drive the execution of corrective saccades. In this case, we would expect stronger error-related signals before the execution of a corrective saccade compared with error trials that were not revised. Hence, we compared firing rates in a 200 ms window before the onset of a revision saccade (type 2b saccade) with a time-matched period from correct and unrevised error trials. A comparison of correct and error trials in this time window yields results similar to the standard time window from 0 to 300 ms after onset of the choice saccade. Furthermore, we find an overall similar pattern of firing rates for revised and unrevised trials (Fig. 8*D*). As for the ROC analysis, we recode firing rates to combine data from cells that increase and decrease firing rates in response to errors. Collapsing over all levels of difficulty, we find significantly stronger firing-rate modulation before revised compared with unrevised errors (paired t test, $p < 0.05$).

Timing of the error-related signals

We further examined the timing of the difficulty and error-related signals. To that aim, spike times were aligned to stimulus and saccade onset. For both datasets, the number of spikes was counted in a 200 ms sliding boxcar window that was moved in

steps of 10 ms. Time-resolved spike count was modeled using the same base model that was also used for the stationary analysis. Figure 9 shows the fraction of cells with significant standard (*A–C*) and balanced (*D*, *E*) difficulty and error-related signals as a function of time from the onset of the stimulus and the choice saccade. *A* and *D* suggest that difficulty signals emerge earlier than error signals in the stimulus-locked analyses. This effect is weaker for the saccade-locked analyses. From the saccade-locked analyses, we extracted the earliest time point at which three successive time bins had a significant effect for difficulty and error. The average onset of the difficulty signals was 92 ± 124 ms after saccade onset when using the main effect of difficulty. This value increased to 157 ± 125 for the balanced difficulty criterion. The onset times for all cells with a main effect error was 125 ± 125 ms after saccade onset. The more rigorous balanced error criterion increased this value to 136 ± 140 ms. The onset times for all cells that coded both (balanced) error and difficulty are shown in Figure 9, *D* and *F*. The data based on standard error and difficulty regressors support the hypothesis that difficulty signals precede error signals. However, when using the more rigorous balanced criterion, there is no evidence that supports the earlier onset of the difficulty signals.

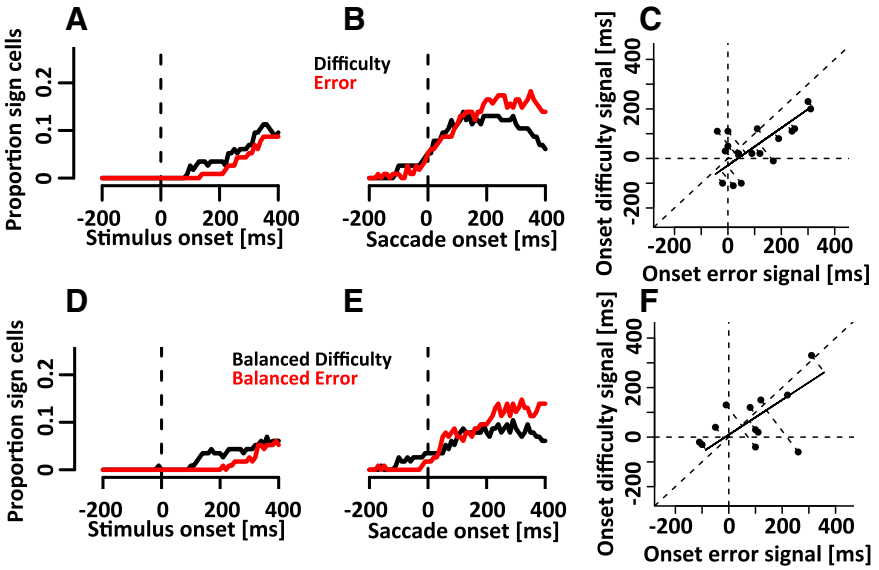

Figure 9. Relative timing of error and difficulty signal in the FEF. Time point *t* includes spikes from *t* − 100ms to *t* + 100 ms. ***A***, ***B***, Proportion of cells with a significant effect of error (red) and difficulty (black) as a function of time from stimulus and saccade onset. ***C***, Earliest onset of the error and difficulty signal relative to saccade onset for all cells with both effects. The two variables are significantly correlated. Difficulty signals emerge 50 ms earlier than error signals. ***D–F***, Same conventions as in ***A–C*** but for balanced error/difficulty. The timing differences observed for the unbalanced tests are no longer significant.

Error-related versus true error signals

Kepecs et al. (2008) have recently introduced a definition of uncertainty that may be related to the difficulty and error-related signals reported here. In their definition, uncertainty reflects the conviction of the subject about having guessed (strong uncertainty signals are associated with a 50/50 chance of being correct). In contrast, they argue that a true error signal should reflect a subject's awareness/conviction about having made a mistake (strong error signals are associated with 0% chance of being correct). In the main body of the paper, we use the term "error-related signal" or "balanced error-related signal" if firing rate is significantly modulated by the observable outcome. This definition does not differentiate a "true error signal" from an uncertainty signal because we would expect a modulation of firing rate with error in both situations.

In the following, we test whether the joint error-related/difficulty signal reported here may be an uncertainty signal. In principle, a true error signal can be dissociated from an uncertainty signal by calculating error likelihood for all trials exceeding a certain firing-rate threshold: a true error signal should lead to error rates above chance for the most extreme firing-rate thresholds (i.e., 100% probability of error for trials with the most extreme firing rates). In contrast, the error rates should never rise above chance for an uncertainty signal (50% probability of error even for the most extreme firing rates; for details, see Kepecs et al. 2008). Based on this approach, we plotted error likelihood as a function of normalized firing-rate cutoff for all neurons with a main effect of error. Averaged across all neurons, error likelihood did not rise above chance for any cutoff quantile (Fig. 10*A*). This finding is consistent with the hypothesis that the difficulty/error-related activity may reflect uncertainty or its inverse, confidence.

However, it is important to point out that the test described above inherently favors the uncertainty account: the test is based on the fact that, for a true error signal, it is always possible to find a firing-rate quantile beyond which error rate increases above 50%. However, if the error signal is small relative to the baseline firing-rate variability, the quantile for which this is true may be very large. For example, given a Poisson-like neuron with a baseline firing rate of 25 Hz, an increase in firing rate for error trials of 5 Hz, an overall rate of correct responses of 77% (similar to monkeys in our study), analyzed in a 200 ms time bin would be expected to have >50% errors only for trials with firing rates in the top 1%. Given that we collected ~500 trials per cell on average, we would expect approximately five trials in this range. Even if all five of these trials were error trials, a binomial test still would not reject the null hypothesis at a statistical threshold of 0.05. Hence, such neurons could easily be mistaken for uncertainty neurons.

Chosen value and reward expectation

Reward expectation is defined as the product of reward magnitude and reward probability. As outlined in the previous paragraph, the difficulty/error-related signals reported here may reflect error likelihood or the uncertainty/confidence of the animals in their choice and hence their probability of reward. This quantity could be converted to a proxy for reward expectation by multiplication with the value of the chosen target. It is also possible that this operation has already been performed and that the cells signal expected reward rather than error likelihood or confidence. To test this assumption, we split the data by the value of the chosen option.

Cells that decreased firing rates with increasing difficulty showed stronger responses when the animal chose a large reward. Conversely, cells that increased their firing rate with difficulty had weaker responses when the animal chose the large reward. To increase statistical power, we combined both types of cells into a single analysis. To account for the different types of responses (increase vs decrease of firing rate with difficulty), we multiplied

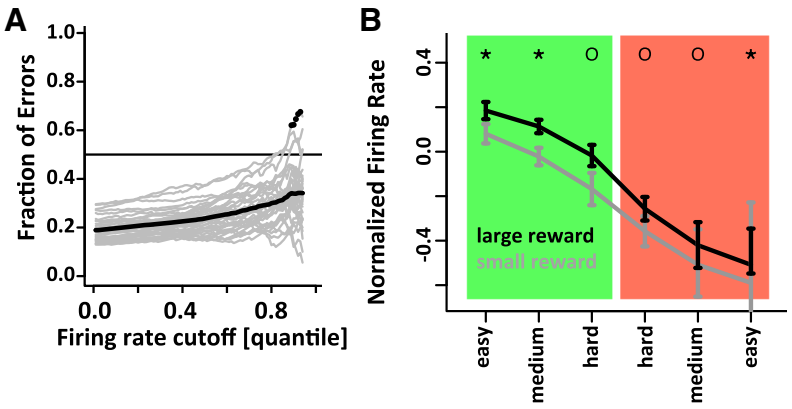

Figure 10. ***A***, Error rate for subsets of trials defined by normalized firing rate. Normalization was performed in the same way as for the ROC analysis, i.e., based on the residuals of a model that excluded error, difficulty, and their interaction. Note that, as for the ROC analysis, the residuals were multiplied with the sign of the parameter of the error regressor to combine cells with firing-rate increase and decrease in the same analysis. The values on the *y*-axis indicate the fraction of errors in the trials that exceeded the firing-rate criterion defined on the *x*-axis. For example, when the *x*-axis equals zero, the data displayed on the *y*-axis include all trials. Hence, the value of ~0.2 on the *y*-axis indicates that, across all trials, mean error rate for the included recording sessions was ~20%. For trials in the top 10 percentiles (*x*-axis equal to 0.9), mean error rate increased to ~35%. The analysis was performed for all cells that had a significant main effect of error or a significant interaction of error and difficulty. Results of individual cells are depicted in gray, and the grand average is indicated in black. Instances in which the error rate is significantly larger than chance are indicated by black dots. The *x*-axis ends at 0.95 because, for higher values, the number of eligible trials is too small to yield reliable estimates. However, the trend continues and the mean error rate keeps rising if the *x*-axis is extended beyond 0.95 (data not shown). ***B***, Effect of the chosen value. Normalized firing rate as a function of error and difficulty split by the value of the chosen target (black, large reward; gray, small reward). The traces from individual neurons were multiplied by −1 if the parameter estimate for error was positive. This way, all normalized firing rates are larger for correct trials. An open circle indicates a nonsignificant outcome. $^*p = 0.05$.

the responses of each cell with −1 times the sign of the difficulty parameter β_{diff}. This way, all cells were artificially converted into cells that decrease firing rate with difficulty, i.e., potential reward expectation cells. We find significantly stronger responses when the animal chose the large reward. The sign of this change is consistent with the assumption that these cells code reward expectation (Fig. 10*B*). However, the overall size of this effect is weaker than would be expected if the cells actually computed expected reward: on average, the large-reward target was approximately twice as large as the small target. Hence, we would expect firing rates to approximately double in the large compared with the small reward condition. This is clearly not the case. However, this may be attributable to a nonlinear relationship between expected reward and firing rate, yet we also fail to find evidence for the prediction that the slope should be steeper for the larger compared with the smaller chosen value. Together, the data are more suggestive of an additive effect of reward magnitude rather than a multiplicative one.

Discussion

We studied single-neuron responses in FEF while macaque monkeys performed a speed-categorization task with variable asymmetric reward contingencies that enticed animals to make impulsive errors. We focused on neural responses immediately after the animals conveyed their choice but before feedback indicated whether or not they picked the correct target. We found that approximately one-quarter of the cells in FEF modulated their firing rate in response to choice errors in the 300 ms after saccade initiation. A similar fraction of cells modulated their firing rate in line with task difficulty. The coding of these two variables in the population was highly correlated, and neurons that increased/decreased their activity after difficult correct trials also tended to increase/decrease their activity after error trials. Our findings show that postsaccadic responses in FEF are correlated with cognitive constructs, such as response conflict, uncertainty, and reward prediction that are typically associated with performance monitoring (Wessel, 2012). Most importantly, they question the segregation of brain regions into actor (response selection in dorsolateral PFC) and critic (response evaluation in medial PFC) and argue in favor of a more integrated architecture.

This is not the first report that links FEF responses to task difficulty. Findings from Bichot et al. (2001) suggest that higher firing rates of FEF cells that code a more salient distractor are the underlying cause that makes such trials more difficult. In contrast, our findings are more consistent with the idea that FEF neurons signal task difficulty, because (1) the difficulty responses were independent of saccade direction, (2) could even be observed in neurons without directional tuning, and (3) persisted well after the execution of the motor response. Hence, our findings are more similar to the difficulty responses reported in two more recent studies (Ding and Gold, 2012; Seo et al., 2012). However, the joint error-related/difficulty signals reported here provide an important extension of these previous findings. First, task difficulty in our paradigm is defined by the similarity of the current speed with an arbitrary category bound. Hence, task difficulty is not directly related to the signal-to-noise ratio as in the other two studies. Second, the difficulty signal was observed before reward delivery, i.e., early enough to put the receipt of the reward in perspective with the difficulty of the trial. Most importantly, our study includes an analysis of error trials to show that cells code a very tightly regulated joint error-related/difficulty signal.

Our detailed analyses allow us to make a strong point that difficulty and error-related responses do not simply reflect the activation of corrective saccades or other motor-related confounds. We also argue against the notion that difficulty signals can be explained by neurons with Gaussian speed tuning centered around the boundary speed. First, we observe linear speed tuning in a number of cells (data not shown), and these linear speed-tuning signals emerged earlier in the trial than the difficulty signals. Second, the Gaussian speed-tuning account cannot explain the observed correlation between difficulty and error-related signals. After rejecting motor and sensory accounts, we consider three related cognitive constructs that are closely linked to performance monitoring: (1) response conflict (Yeung et al., 2004); (2) uncertainty (Kepecs et al., 2008); or (3) reward expectation (Holroyd and Coles, 2002).

Response conflict

The conflict-monitoring theory (Botvinick et al., 2001) states that dedicated neurons monitor the simultaneous coactivation of competing response options (the correct and the wrong response). Based on simulations (Yeung et al., 2004), these conflict-monitoring cells should code task difficultly during the decision process and an error-related signal immediately after the execution of the response. Our analysis supports the prediction that

difficulty signals precede error signals when using the standard error and difficulty regressors (Fig. 8*A–C*). However, the more rigorous balanced regressors do not support this notion because the time course for the error and difficulty signals was not significantly different (Fig. 9*D–F*). These mixed results regarding the sequential progression from difficulty to error only partially support the conflict-monitoring account.

Reward expectation

In their groundbreaking paper, Holroyd and Coles (2002) linked error signals to reward prediction error. The main features of our data closely resemble one component of this theory: reward expectation (Fig. 8*B*). However, when splitting the data based on the value of the chosen target, we find evidence both in favor and against reward expectation. Cells that increased firing rates with putative reward likelihood showed stronger responses when the animal chose a large reward (Fig. 10*B*). Conversely, cells that decreased their firing rate with putative reward likelihood had lower responses when the animal chose the large reward. However, the overall size of this effect is weaker than would be expected if the cells actually computed expected reward. Also, we fail to find evidence for the prediction that the slope of the regression changes with chosen value. Overall, these findings provide some evidence in favor as well as some evidence against the interpretation of the signal as a correlate of reward expectation.

Uncertainty

Performance monitoring has also been closely linked to the assessment of error likelihood (Brown and Braver, 2005) and uncertainty (Kepecs et al., 2008). Using the method suggested by Kepecs et al. (2008), we tested whether the observed error-related signals were more consistent with a true error signal or uncertainty. Overall, the analysis supports the predictions of the uncertainty account over a true error signal (Fig. 10*A*). However, it is important to keep in mind that the analysis inherently favors the uncertainty account, especially when the effect size of error is small.

If we assume that the observed signals are uncertainty related, then they could not be involved in mediating the corrective saccades that occurred almost exclusively after wrong choices. Hence, we tested whether the error-related signal is stronger after revised errors. Our results show that this is the case (Fig. 8*E*). This finding is in line with the original observation that ERN amplitude is stronger after revised errors (Gehring et al., 1993). However, subsequent single-cell studies did not find a correlation between error-related activity and subsequent error correction. Although the effect size of revised versus unrevised errors is small, it suggests that the observed error-related signal may be involved in the detection and subsequent correction of the error. If this were the case, it would argue against the uncertainty account, because an uncertainty signal would not be able to explain the almost exclusive revisions of wrong choices. However, other explanations for the more extreme firing rates in revised trials are possible. In particular, it is possible that the signal is related to the preparation of the corrective saccade rather than the signaling of an error.

In summary, we find some evidence in favor of all three of the cognitive accounts. However, we also find evidence that is inconsistent with these theories. Consequently, the available electrophysiological data do not clearly distinguish between the three accounts, in part because they make very similar predictions. Evidence against the uncertainty account is the least convincing and hence provides the most parsimonious account of the electrophysiological data by itself. This is in part because of the fact that the uncertainty account makes the least specific additional predictions and is therefore more difficult to refute.

However, we believe that the uncertainty account is less parsimonious, if we take the behavior of the animals into account. Revisions occurred almost exclusively after error trials, which implies that the animals had access to a true error signal. If a true error signal is indeed present and triggers corrective saccades in the FEF, we would have to assume that there are two types of error-related activity: (1) the uncertainty signal that we measured; and (2) the true error signal that is present and drives saccades but did not show up in our recordings. Although this is possible, it seems unlikely. In the end, we lean in favor of the conflict-monitoring account because it explains many aspects of both the neural recordings and behavior.

Diverse performance monitoring signals in prefrontal cortex

Performance monitoring is typically associated with medial prefrontal structures, such as the SEFs and ACC. In a saccadic countermanding task, distinct groups of neurons in the SEF have been suggested to code either errors of commission (15%) or response conflict (13%) during successful response inhibition (Stuphorn et al., 2000). In the same task, neurons in the ACC (~14%) were found to respond after errors but not as a function of response conflict (Ito et al., 2003). Because many of the same error cells also respond to the omission of reward, it was speculated that error-related responses in the ACC may be related to reward-prediction error. In our speed-categorization task, the FEF exhibits a distinct pattern that is consistent with neither the SEF or ACC. In contrast to the ACC, the FEF encodes error-related signals (23%) and difficulty (27%) in the categorization task.

In contrast to the SEF, in which there is separate coding of error and difficulty in distinct neural populations, the FEF multiplexes both error and difficulty onto the same neurons. The latency of the balanced error-related signals (136 ± 140 ms) was between the latencies reported for error signals in the SEF (111 ± 13 ms, mean ± SEM) and ACC (180 ± 13, mean ± SEM) (Ito et al., 2003). However, note that different analysis methods were used, and it is necessary to subtract ~21 ms from the values reported by Ito et al., (2003) and to keep in mind that the current study used wider bins.

Our findings show that FEF cells can either increase (18 cells) or decrease (13 cells) firing rate with task difficulty (Fig. 7). This mix of positive and negative firing-rate changes is consistent with previous findings in FEF (Ding and Gold, 2012; Seo et al., 2012). Our data also show that FEF can increase (17) as well as decrease (10) firing rate after wrong choices (Fig. 7). This is the first report of error-related signals in the FEF and is in contrast to established results in the SEF and ACC in which all putative error cells increased firing rates after failed inhibitions.

In summary, we show that a significant fraction of cells in the FEF modulate their firing rate as a function of choice error and difficulty in the aftermath of an oculomotor decision. This finding extends previous studies and highlights that the role of the FEF does not end with the selection and execution of a motor response. Most importantly, our findings question the segregation of brain regions into actor (response selection) and critic (response evaluation) and argue in favor of a more integrated architecture. The subsequent evaluation of responses in the same structures that mediate their selection may be particularly important for the iterative adjustment of selection strategies to changing circumstances.

References

Bichot NP, Chenchal Rao S, Schall JD (2001) Continuous processing in macaque frontal cortex during visual search. Neuropsychologia 39:972–982. CrossRef Medline

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624–652. CrossRef Medline

Brown JW, Braver TS (2005) Learned predictions of error likelihood in the anterior cingulate cortex. Science 307:1118–1121. CrossRef Medline

Bruce CJ, Goldberg ME (1985) Primate frontal eye fields. I. Single neurons discharging before saccades. J Neurophysiol 53:603–635. Medline

Ding L, Gold JI (2012) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. Cereb Cortex 22:1052–1067. CrossRef Medline

Draper NR, Smith H (1966) Applied regression analysis, Ed 2. New York: Wiley.

Ferrera VP, Barborica A (2010) Internally generated error signals in monkey frontal eye field during an inferred motion task. J Neurosci 30:11612–11623. CrossRef Medline

Fujii N, Mushiake H, Tanji J (1998) An oculomotor representation area within the ventral premotor cortex. Proc Natl Acad Sci U S A 95:12034–12037. CrossRef Medline

Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. Psychol Sci 4:385–390. CrossRef

Godlove DC, Emeric EE, Segovis CM, Young MS, Schall JD, Woodman GF (2011) Event-related potentials elicited by errors during the stop-signal task. I. Macaque monkeys. J Neurosci 31:15640–15649. CrossRef Medline

Goldberg ME, Bruce CJ (1990) Primate frontal eye fields. III. Maintenance of a spatially accurate saccade signal. J Neurophysiol 64:489–508. Medline

Hanes DP, Patterson WF 2nd, Schall JD (1998) Role of frontal eye fields in countermanding saccades: visual, movement, and fixation activity. J Neurophysiol 79:817–834. Medline

Holroyd CB, Coles MGH (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 109:679–709. CrossRef Medline

Ito S, Stuphorn V, Brown JW, Schall JD (2003) Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science 302:120–122. CrossRef Medline

Judge SJ, Richmond BJ, Chu FC (1980) Implantation of magnetic search coils for measurement of eye position: an improved method. Vision Res 20:535–538. CrossRef Medline

Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455:227–231. CrossRef Medline

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

Murthy A, Ray S, Shorter SM, Priddy EG, Schall JD, Thompson KG (2007) Frontal eye field contributions to rapid corrective saccades. J Neurophysiol 97:1457–1469. Medline

Purcell BA, Schall JD, Logan GD, Palmeri TJ (2012) From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search. J Neurosci 32:3433–3446. CrossRef Medline

R Development Core Team (2009) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Seo M, Lee E, Averbeck BB (2012) Action selection and action value in frontal-striatal circuits. Neuron 74:947–960. CrossRef Medline

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 [Suppl 1]:S208–S219.

Stuphorn V, Taylor TL, Schall JD (2000) Performance monitoring by the supplementary eye field. Nature 408:857–860. Medline

Taylor SF, Stern ER, Gehring WJ (2007) Neural systems for error monitoring: recent findings and theoretical perspectives. Neuroscientist 13:160–172. CrossRef Medline

Teichert T, Ferrera VP (2010) Suboptimal integration of reward magnitude and prior reward likelihood in categorical decisions by monkeys. Front Neurosci 4:186. CrossRef Medline

Thompson KG, Bichot NP, Schall JD (1997) Dissociation of visual discrimination from saccade programming in macaque frontal eye field. J Neurophysiol 77:1046–1050. Medline

Weisberg S, Fox J (2010) An R companion to applied regression. Thousand Oaks, CA: Sage Publications.

Wessel JR (2012) Error awareness and the error-related negativity: evaluating the first decade of evidence. Front Hum Neurosci 6:88. Medline

Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Rev 111:931–959. CrossRef Medline