

UC Davis

UC Davis Previously Published Works

Title

Frontal eye field neurons selectively signal the reward value of prior actions.

Permalink

<https://escholarship.org/uc/item/1vj6b4q0>

Authors

Zirnsak, Marc

Vega, Gabriel

Moore, Tirin

et al.

Publication Date

2020-12-01

DOI

10.1016/j.pneurobio.2020.101881

Peer reviewed



HHS Public Access

Author manuscript

Prog Neurobiol. Author manuscript; available in PMC 2021 December 01.

Published in final edited form as:

Prog Neurobiol. 2020 December ; 195: 101881. doi:10.1016/j.pneurobio.2020.101881.

Primate Frontal Eye Field Neurons Selectively Signal the Reward Value of Prior Actions

Xiaomo Chen, Marc Zirnsak, Gabriel M. Vega, Tirin Moore

Department of Neurobiology and Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA.

Abstract

The consequences of individual actions are typically unknown until well after they are executed. This fact necessitates a mechanism that bridges delays between specific actions and reward outcomes. We looked for the presence of such a mechanism in the post-movement activity of neurons in the frontal eye field (FEF), a visuomotor area in prefrontal cortex. Monkeys performed an oculomotor gamble task in which they made eye movements to different locations associated with dynamically varying reward outcomes. Behavioral data showed that monkeys tracked reward history and made choices according to their own risk preferences. Consistent with previous studies, we observed that the activity of FEF neurons is correlated with the expected reward value of different eye movements before a target appears. Moreover, we observed that the activity of FEF neurons continued to signal the direction of eye movements, the expected reward value, and their interaction well after the movements were completed and when targets were no longer within the neuronal response field. In addition, this post-movement information was also observed in local field potentials, particularly in low-frequency bands. These results show that neural signals of prior actions and expected reward value persist across delays between those actions and their experienced outcomes. These memory traces may serve a role in reward-based learning in which subjects need to learn actions predicting delayed reward.

Keywords

Prefrontal cortex; Decision-making; Reward; Eligibility trace; Reinforcement learning

1. Introduction

A basic fact of learned behaviors is that they are generally shaped by their experienced consequences; behaviors preceding aversive events tend to diminish in frequency, whereas those preceding reward tend to be repeated. The neural mechanisms of reinforcement

Lead Contact: Tirin Moore, Dept. of Neurobiology, Fairchild Bldg. 299 Campus Drive, Stanford University School of Medicine, Stanford CA 94305.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Competing Interest

The authors declare no competing financial interests.

learning have been extensively studied at multiple levels for several decades (Lee et al., 2012; Neftci and Averbeck, 2019; Soltani et al., 2017), yet among the more significant lingering questions is how behaviors are linked to their ensuing consequences when the latter does not immediately follow the former. Neural signals that command specific motor behaviors, such as eye movements, generally operate on a timescale of tens of milliseconds (Yarbus, 1967), yet in most circumstances, the rewarding or aversive consequences of those behaviors happen on a much longer timescale. This temporal discrepancy has been referred to as the ‘distal reward problem’ (Hull, 1943; Izhikevich, 2007) or the ‘credit assignment problem’ (Barto et al., 1983; Minsky, 1961; Sutton and Barto, 1998). For example, upon scoring a point in a tennis match, the brain needs to associate that type of reward with the specific movements that preceded it. Among the mechanisms proposed to address this problem is a signal that persists long enough to bridge the time between specific behaviors and their consequences (Drew and Abbott, 2006; Sutton and Barto, 1998). Yet, evidence of such a mechanism has thus far been limited (Gerstner et al., 2018; Lee et al., 2012), particularly at the level of spiking neuronal activity.

In humans and other primates, visually guided behavior typically begins with the selection of visual stimuli within a cluttered visual scene and the serial foveation of particular items via saccadic eye movements. Visually guided eye movements are the primary means by which information is gathered from the environment. Each movement transforms previously unresolvable details in the visual periphery into resolvable percepts at the fovea. In this behavior, the fovea is a limited resource, and the value of the information obtained from eye movements is typically unknown until well after they are performed. Thus, one might consider neurons involved in visually guided eye movements as candidates for conveying information about the value of movements even after those movements are carried out. A number of previous studies have demonstrated that neurons within the cortical eye fields of primates, including the frontal eye field (FEF), signal the reward value of upcoming saccadic eye movements (Chen and Stuphorn, 2015; Ding and Hikosaka, 2006; Glaser et al., 2016; Roesch and Olson, 2003; So and Stuphorn, 2010). Yet, it remains unclear if the value of prior movements is likewise encoded.

To address this question, we studied the activity of FEF neurons in a behavioral task that allowed us to measure value signals well after targeting eye movements. We first use prospect theory (Chen and Stuphorn, 2018; Glimcher and Fehr, 2013; Tversky and Kahneman, 1979) to estimate the subjective value of each reward option. We next show that the activity of FEF neurons conveys information both about the direction of prior movements and their corresponding reward value. This post-saccadic activity encodes the conjunction of prior movements and their subjective value and is different from the reward value representation before targeting eye movements. Lastly, we show that post-movement information was also observed in simultaneously recorded FEF LFPs, particularly in low-frequency bands. These results suggest a potential role of the FEF in bridging the delays between specific eye movements and reward outcomes and in reinforcing the eye movements leading to rewarding consequences.

2. Methods

All experimental procedures were in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals, the Society for Neuroscience Guidelines and Policies, and Stanford University Animal Care and Use Committee. Two healthy male rhesus monkeys (*Macaca mulatta*, 17 and 16 kg), monkey J and monkey O, were used in these experiments.

2.1. General and Surgical Procedures

Surgery was conducted using aseptic techniques under general anesthesia (isoflurane) and analgesics were provided during postsurgical recovery. Each animal was surgically implanted with a titanium head post and a cylindrical titanium recording chamber (20 mm diameter) overlaying the arcuate sulcus. A craniotomy was then performed in the chambers on each animal, allowing access to the FEF.

2.2. Neurophysiological techniques

Recording sites within the FEF were identified by eliciting short-latency, fixed vector saccadic eye movements with trains (50–100ms) of biphasic current pulses (50 μ A; 250 Hz; 0.25 ms duration) as in previous studies (Bruce et al., 1985). Single-neuron and local field potential (LFP) recordings were obtained with 16 or 32-channel linear array electrodes with contacts spaced 150 μ m apart (V and S-Probes, Plexon, Inc). Electrodes were lowered into the cortex using a hydraulic microdrive (Narishige International). Neural activity was measured against a local reference: a stainless guide tube located close to the electrode contacts. At the preamplifier stage, signals were processed with 0.5 Hz 1-pole high-pass and 8 kHz 4-pole low-pass anti-aliasing Bessel filters, and then divided into two streams for the recording of LFPs and spiking activity. The stream used ultimately for LFP analysis was additionally amplified (\times 500–2000), processed by a 4-pole 200 Hz low-pass Bessel filter and sampled at 1000 Hz. No other filters were used in the analyses. The stream used for spike detection was processed by a 4-pole Bessel high-pass filter (300 Hz) a 2-pole Bessel low-pass filter (6000 Hz), and was sampled at 40 kHz. Extracellular waveforms were classified as single neurons or multi-units using online-template-matching and subsequently confirmed using offline sorting (Plexon).

2.3. General behavioral techniques

During all behavioral measurements, eye position was monitored and stored at 1000 Hz, and a spatial resolution of \sim 0.05 degrees of visual angle (dva) (Eyelink 1000, SR Research). Task stimuli were presented on a display (Samsung 2233RZ, 120 Hz refresh rate, 1680 \times 1050 pixel resolution) positioned 28–30 cm in front of the animal.

2.4. Behavior task

Monkeys were trained on an oculomotor gamble task in which they made saccadic eye movements to targets with differing, delayed reward outcomes (Figure 1). The monkey first fixated on a central fixation point (0.5 dva diameter) on gray background (60 cd/m²) for a setinterval (900 ms). Fixation was enforced within a \pm 1 dva error window. During this period, on 67% of trials, a texture background (\sim 4% luminance contrast) appeared across

the entire display after 300 ms. Following the fixation interval, one target (force-choice trials) or two targets (free-choice trials) appeared on the display. The targets are 3 dva diameter white circles. They appeared at locations in the right, left or both hemifields, and target eccentricities varied between sessions from 5°–12°. Following a target presentation period of 800 ms, the central fixation point was turned off, and monkeys were free to make saccadic movements to one of the targets. After the saccade, fixation on the target was required for an additional 400–800ms (400 ms: four sessions; 600 ms: one session; 800 ms: six sessions), after which the upcoming reward amount was cued by the number of red dots appearing around the fixation (result onset). Juice reward was delivered following an additional 300 ms after reward onset.

Each behavioral block started with 20 *forced-choice* trials. In these trials, only one target appeared on the display. Monkeys made saccadic eye movements and learned different movement-reward contingencies at two different target locations. After the forced-choice block, monkeys performed 20–40 free-choice trials. These trials were used to assess the extent to which monkeys learned the movement-reward contingencies and to measure their preferences between different reward conditions. Overall, there were three possible reward conditions with varying quantities of a juice reward, a *High-risk*, a *Low-risk*, and a *Sure* condition. In the High-risk condition, monkeys had a 50% chance of receiving 4 drops of juice and a 50% chance of receiving 0 drops. In the Low-risk condition, monkeys had a 50% chance of receiving 3 drops of juice and a 50% chance of receiving 1 drop. In the Sure condition, monkeys always received 2 drops of juice. For all these reward outcomes, the average reward amounts were equal, but the variance in the amount differed between conditions. We used free-choice trials to analyze the monkeys' subjective value for each reward condition. Forced-choice trials, in which all of the risk conditions and target directions were sufficiently sampled, were used for the neurophysiological analysis. Each experimental session consisted of a complete pairing of Sure, Low-risk, and High-risk trials across the two target locations without pairing identical risk conditions. This resulted in 6 different pairings. Overall, there were 18 blocks, as each pair was repeated three times in a pseudo-random order.

2.3 Behavioral Analysis

We assumed a softmax decision function, where the probability of selecting the gamble was indicated by the difference of the subjective values of the two options:

$$h_{\lambda}(\Delta U) = \frac{1}{1 + e^{(-\lambda(\Delta U))}} Eq \quad (1)$$

where $U = U_1 - U_2$ gives the utility difference between gamble options, and λ is the softmax parameter that determines stochasticity in selection between the two options.

We used prospect theory to estimate subjective values of the two gamble options (Chen and Stuphorn, 2018; Hsu et al., 2009). Prospect theory is derived from classical expected value theory in economics and assumes that the subjective value of a gamble depends on the utility of the reward amount that can be earned, weighted by the probability of the particular outcome. The subjective value of the choice option o was calculated as follows:

$$U_o = u_\rho(V_{win_o}) \times p_{win_o} + u_\rho(V_{loss_o}) \times p_{loss_o} \quad (2)$$

where $u_\rho(V)$ is a power function to model the utility function, following previous research (Hsu et al., 2009; Lattimore et al., 1992):

$$u_\rho(V) = V^\rho \quad (3)$$

Both humans and monkeys exhibit relatively accurate estimations of mid-range probability around 0.5 (Farashahi et al., 2018; Stauffer et al., 2015). We thus used p_{win_i} and p_{loss_i} as an estimate of subjective probability.

2.5 Analysis of the Neural Activity

Data were analyzed using custom scripts written in MATLAB (The Mathworks, Inc), unless otherwise indicated. Spike times were converted to firing rate estimates by convolution with a causal 50 ms boxcar filter (Chandrasekaran et al., 2017) and then normalized across all trials using z-score normalization. Overall, we recorded 293 multi and single units with visual activity and 264 LFP channels. We focused our analyses on average neuronal activity and LFP power measured in the early target period [−300 to 49 ms] (pre-target period) and prior to the result onset, following the targeting eye movement [−300 to 49 ms] (post-movement period). During both periods, the monkeys fixated on either the fixation point or the chosen target throughout the analysis period. These analysis windows were chosen to exclude or minimize the influence of transient visual responses. In addition, as we did not observe significant differences between effects measured with low contrast textured backgrounds and homogenous backgrounds during these time windows, all trials were combined in the analysis.

2.5.1 Linear regression analysis—Linear regression analysis was performed to quantify the influence of reward value on neuronal activity and LFP power. Trials were first separated according to the target locations. Next, average neural activity was regressed against the subjective value of reward options.

$$A_i = \beta_0 + \beta_1 \times U_i \quad (4)$$

Where A_i is the average firing rate or the mean energy in a given LFP band on the i th trial, U_i is the subjective value of the reward option, β_0 and β_1 are regression coefficients.

2.5.2 Demixed principal component analysis (dPCA)—We used dPCA analysis to decompose population activity into different task components: condition-independent components, target direction components, subjective value components, and interactions components between target direction and subjective value. For the dPCA analysis, all neuronal activity across all sessions was used. Firing rates were down sampled into 10ms time bins. We performed regularized dPCA and decoding of classes as described in Kobak et al., 2016. Time periods in which classification accuracy exceeded all 100 shuffled decoding accuracies in at least 10 consecutive time bins were considered significant.

2.5.3 Support Vector Machine (SVM) linear classifier—We used a linear support vector machine (SVM) classifiers (Chang and Lin, 2011) to quantify the information about movement direction and subjective value contained in the population of FEF neurons or LFPs in the pre-target period and the post-movement period. In decoding subjective value, for example, a classifier was trained on neural activity (pre-target or post-movement) to discriminate between reward outcome conditions (High, Low, and Sure) per direction. Before training, spike counts for each neuronal recording were normalized across all stimulus conditions. All reported discrimination accuracies are based on five-fold cross-validation. Permutation tests (1000 repetitions) were used to determine whether the discrimination accuracy of a given neuronal recording was significantly greater than that expected by chance (discrimination performance of the classifier after label shuffling).

2.5.4 Power spectral density (PSD)—PSDs were calculated by Thomson's multitaper method (Gregoriou et al., 2009; Jarvis and Mitra, 2001; Pesaran et al., 2008). Four orthogonal discrete prolate spheroidal (Slepian) sequences were used in the analysis. For both pre-target and post-movement period, we examined the 350 ms LFPs with 1000 Hz sampling rate. This resulted in a ~2.9 Hz frequency resolution. On each trial, spectra were converted to decibels, and were normalized across trials for each frequency using min-max normalization. Our analysis of LFPs focused on four frequency bands known to contain task-relevant information in the FEF, specifically the alpha band (8–12 Hz), the beta band (12–30 Hz), the low-gamma band (30–80 Hz), and the high-gamma band (80–150 Hz).

3. Results

3.1 Monkeys preferred risky choices

We trained two monkeys (O and J) to perform an oculomotor gamble task (Fig. 1). In the task, monkeys made saccadic eye movements to targets with differing, delayed reward outcomes. The task consisted of multiple blocks of two types of trials: forced-choice trials and free-choice trials. Each session started with a set of forced-choice trials in which monkeys learned the reward outcomes associated with each of the two target locations, one of which coincided with the response fields (RFs) of recorded FEF neurons. Three possible sets of reward conditions were associated with each location: a High-risk, a Low-risk, and a Sure condition. In the High-risk case, the monkey received 4 drops or 0 drops of juice with equal probability. In the Low-risk case, it received 3 drops or 1 drop of juice with equal probability, and in the Sure case, it always received 2 drops of juice. Thus, in each forced-choice block, the outcomes always had the same average reward, but different reward variances (Fig. 1B). Forced-choice trials were followed by a block of free-choice trials in which the two choice locations had the same reward outcomes as in the preceding forced-choice trials. The free-choice trials were used to measure the monkey's choice behavior, and always consisted of unequal pairings of risk (e.g. Sure and High-risk). Each experimental session consisted of a complete pairing of Sure, Low-risk and High-risk trials across the two target locations (Methods).

During free-choice trials, each monkey's choice behavior was influenced both by reward history and the monkey's risk preference. Consistent with previous work (Barraclough et al.,

2004; Sugrue et al., 2004), reward history reliably influenced choice behavior. Both monkeys were more likely to repeat choosing the same option if they had won the gamble in the previous Low or High-risk trial (Fig. 2A) (2×2 repeated measures ANOVA, Monkey O: $p = 10^{-3}$; Monkey J: $p < 10^{-5}$). In addition, the monkey's risk preference also had a significant effect on choice behavior. Rather than exhibiting a win-stay-lose-switch strategy, both monkeys were more likely to continue choosing the High-risk option, regardless of the previous reward outcome (Fig. 2A) (2×2 repeated measures ANOVA, Monkey O: $p < 10^{-4}$; Monkey J: $p < 10^{-4}$). We further quantified the monkeys' choice behavior using prospect theory (Tversky and Kahneman, 1979) to obtain utility functions (Chen and Stuphorn, 2018; Hsu et al., 2009). We found that the utility functions for both monkeys were convex across experimental sessions (Monkey O: $\rho = 1.76$, $p < 10^{-7}$; Monkey J: $\rho = 1.99$, $p < 10^{-9}$) (Fig. 2B). The convexity of the utility functions reveals that both monkeys were risk-seeking, consistent with previous studies (Chen and Stuphorn, 2018; Farashahi et al., 2018; Kim et al., 2012; McCoy and Platt, 2005). Lastly, we estimated the subjective value of each choice option using the utility functions from each experimental session. For both monkeys, the higher risk options had larger subjective values (Fig. 2C). We next measured the representation of subjective value by FEF neurons during both the pre-target and post-movement periods.

3.2 Neuronal activity in FEF correlates with subjective value during pre-target period

We measured the spiking activity from a total of 293 single and multi-unit recordings in the FEF using linear array micro-electrodes (Methods). We focused our analyses on neural activity measured during forced-choice trials in which all of the risk conditions and target directions were sufficiently sampled. We first examined whether FEF neuronal activity represented the subjective value associated with different target locations in the pre-target period. We found that activity differed across the different risk conditions; that is, FEF activity was modulated by subjective value. This modulation emerged prior to the appearance of the target and was present both when targets appeared inside and outside of the neuronal RF (Fig. 3A). Of the total 293 neuronal recordings, 28% ($n = 82$) exhibited individually significant effects of the target's subjective value (Fig. 3B). Among those units, 59% ($n = 48$) were significantly modulated by the subjective value of targets located inside of the RF, 70% ($n = 57$) were significantly modulated by the subjective value of targets located outside of the RF, and 28% ($n = 23$) were modulated by the subjective value of targets located outside of the RF, and 28% ($n = 23$) were modulated by subjective value of both. In addition, across the full population of recordings, activity was on average positively correlated with subjective value prior to the appearance of the target in the neuron's RF (Fig. 3C) ($\beta_{in} = 0.02$, $p < 10^{-4}$). In contrast, for target locations outside of the neuronal RF, activity was on average negatively correlated with subjective value ($\beta_{out} = -0.04$, $p < 10^{-9}$).

Correlation coefficients obtained from locations inside the RF (β_{in}) were negatively correlated with coefficients obtained from locations outside the RF field (β_{out}) ($p < 10^{-8}$). The pattern of observed modulation suggests that neuronal activity reflects the subjective value of the target. However, determining the influence of subjective value is difficult during the pre-target period because neuronal activity may also reflect other behavioral variables that covary with subjective value, variables such as the planning of eye movements (Bruce

and Goldberg, 1985; Glaser et al., 2016) and/or attentional deployment toward or away from the RF (Thompson et al., 2005). Thus, we turned the focus of our analysis to the post-movement period, when only the anticipation of the reward outcome was a factor.

3.3 FEF neurons selectively signal subjective value of prior movements during the post-movement period.

In our task, eye movements to each target were followed by a post-movement period (400 – 800 ms), after which monkeys received feedback about the impending reward amount (Fig. 1). During this period, maintenance of fixation on the target was highly stable, and neither monkey aborted any trials by breaking fixation (0% abort rate). This period allowed us to measure FEF activity after the monkey had shifted the target to the center of gaze and outside of the neuronal RF. During the post-movement period, the location of neuronal RFs no longer coincided with a behaviorally relevant stimulus, a relevant location, nor a movement plan regardless of whether the target had previously been inside or outside of the RF. Thus, this period allowed us to measure the extent to which FEF activity continued to signal the subjective value of the fixated target. Indeed, we found that during the post-movement period, while the monkey fixated on the target, FEF activity continued to vary according to the target's subjective value (Fig. 4A). The modulation by subjective value was evident both for trials in which targets had previously appeared inside and outside of the RF.

We examined the subjective value modulation during the post-movement period in the full population of 293 neuronal recordings (Fig. 4B). During this period, 26% ($n = 77$) exhibited individually significant effects of the target's subjective value. Of those neurons, 53% ($n = 41$) were significantly modulated by the subjective value of targets that had previously appeared inside of the RF. A similar proportion of neurons, 61% ($n = 47$), was significantly modulated by the subjective value of targets that had appeared outside of the RF, and 14% ($n = 11$) were modulated by subjective value of targets appearing at either location. Similar to the pre-target period, across the full population of recordings, activity was on average positively correlated with subjective value when the target had appeared inside of the RF ($\beta_{in} = 0.02, p < 10^{-3}$). For targets that had appeared outside of the RF, however, activity was also positively correlated with subjective value ($\beta_{in} = 0.03, p < 10^{-8}$). Thus, subjective value was signaled by FEF neurons well after targeting eye movements, whether or not the target had previously been inside of the RF. Yet, the pattern of modulation differed from that observed during the pre-target period, when activity was negatively correlated with subjective value of non-RF targets (Fig. 3B). Moreover, in contrast to the pre-target period, where subjective value correlations during inside and outside RF trials were negatively correlated with one another (Fig. 3B), they were positively correlated during the post-movement period ($p < 10^{-3}$)(Fig. 4B).

In spite of the overall positive correlation between activity and subjective value, individual FEF neurons appeared to exhibit a heterogeneous pattern of modulation both by subjective value and eye-movement direction (Fig. 4A). That is, FEF neurons showed mixed selectivity to the different task components (Fusi et al., 2016; Rigotti et al., 2013). In the post-movement period, many neurons were selective to the saccadic direction, the target's subjective value, or both. We therefore applied dimensionality reduction to reduce the

population activity to task feature-dependent components that summarize most of the neuronal activity patterns (Cunningham and Byron, 2014; Kobak et al., 2016) (Methods). These components revealed a simpler population-level structure underlying individual neuronal firing rates. Specifically, we found that the population activity was captured by a single condition-invariant feature and was significantly modulated by 3 condition-variant features (Fig. 5A). Among these components, the condition-invariant component captured the temporal profile of activity largely identically across the different conditions and explained 55% of the total variance of the population firing rate. Among the condition-variant components, the risk condition component (High, Low, and Sure) correlated with subjective value, and explained 12% of the variance. In addition, direction of prior movement explained 22% of the variance in the post-movement activity. Lastly, an interaction component captured the interaction between the target's subjective value and the prior movement direction and explained 11% of the variance in population activity. This is in contrast with the peri-target period (Fig. S1), where population activity was significantly modulated by target direction (38% of variance) and the interaction between the target's subjective value and target direction (11% of variance), but not subjective value (6% of variance).

To quantify the information provided by each of the task components, we measured the performance of a linear classifier in discriminating between different subjective target values and prior movement directions using the trial-by-trial responses from all 293 neuronal recordings (Fig. 5B). Consistent with the dPCA analysis, the classifier accurately decoded both the prior saccade direction and the risk condition, exceeding the level expected by chance (Mean accuracy across 6 conditions: 61%; chance level = 17%). In addition, the classifier made more errors when decoding different risk conditions within a certain movement direction than between movement directions. We further assessed the classifier performance in 20-ms intervals to examine how task component information evolved throughout the post-movement period. We found that accuracy in decoding the direction of the prior movement was robust and stable throughout the post-movement period (Mean accuracy: 93%). Decoding performance for risk conditions exceeding chance levels (~ 33%) both for trials on which the target had appeared inside the RF (Mean accuracy: 55%) or outside the RF (Mean accuracy: 51%). Furthermore, for both conditions, decoding accuracy increased toward the result onset (Fig. 5C). This pattern was consistent for both monkeys (Fig. S2A and B)

3.4 FEF local field potentials (LFPs) selectively signal subjective value of prior movement during post-movement period.

Lastly, we examined the modulation of LFP by subjective value modulation on LFPs in recorded FEF LFPs ($n = 264$) during the post-movement period. First, we found that LFP responses in all frequency bands represented the direction of prior movements at both the population level and in individual recordings (Fig. 6A and B). Among all frequency bands, activity in the beta band from 58% of recording sites significantly signaled the direction of the prior movement as quantified by a linear classifier (see methods). As a control, LFP responses during the pre-target period showed almost no selectivity to the upcoming movement direction (Fig. S3). In addition, we observed that both low-frequency (alpha and

beta bands) and high-gamma band LFP responses were modulated by the subjective value of targets that had appeared inside and outside of the RF (Fig. 6A and B, see Methods). These effects shared the same polarity across either target locations (Fig. 6C). Specifically, low-frequency LFP responses were negatively correlated with subjective value of the reward condition while the high-gamma band LFP responses were positively correlated with subjective value of targets (Fig. 6C). Notably, across all frequency bands, the majority of beta band LFP responses were significantly modulated by subjective value (In RF: 56%; Out RF: 52%) (Fig. 6B).

Discussion

Our results show that the activity of FEF neurons continues to signal the subjective value of targeted stimuli as well as the direction of prior eye movements even after those movements are completed and when the targets of movements are no longer positioned inside of the RF. Monkeys made eye movements to targets of varying risk magnitude and exhibited a preference for the higher-risk-targets. Consistent with previous studies, during the pre-target period, FEF activity varied with the subjective value of target locations. However, FEF activity during this period is known to depend heavily on the preparation of eye movements (Goldberg and Bushnell, 1981), as well as on attentional deployment (Armstrong et al., 2009; Kastner et al., 1999; Thompson et al., 2005), potentially accounting for the apparent modulation by subjective value. Thus, we examined neuronal activity during the period after movement completion in a task that delayed the reward outcomes of those movements. We found that the activity of individual FEF neurons continued to signal the subjective value of target stimuli during the post-movement period when the targets had appeared inside or never appeared in the neuronal RF. Population level analyses revealed that information about subjective value and prior movement direction was robust throughout the post-movement period in FEF neuronal activities. Furthermore, we found that this conjunction of prior movement direction and subjective value information was also abundant in low-frequency LFPs in around 55% of the recording sites. These results suggest a potential role of FEF in reinforcing the eye movements leading to rewarding consequences.

A collection of past studies shows that signals related to an animal's previous choices have been observed in a number of primate brain areas. These areas include the prefrontal cortex, posterior parietal, and cingulate cortex (Akrami et al., 2018; Barraclough et al., 2004; Fecteau and Munoz, 2003; Genovesio et al., 2006; Hwang et al., 2017; Sugrue et al., 2004). However, it is unclear how history of choice and reward can conjointly shape the neuronal activity and behavior in the next trial, especially when the reward does not immediately follow the action that leads to it. In reinforcement learning theory, this temporal credit assignment problem can be resolved in at least two ways. The first is to broaden the temporal footprint of spike-timing-dependent plasticity (STDP) by neuromodulators such as dopamine, acetylcholine, serotonin, or norepinephrine (Gerstner et al., 2018). Evidence suggests that the 10-msec timescale of conventional STDP may be moderately extended by neuromodulators (Bittner et al., 2017; Gerstner et al., 2018; He et al., 2015; Pawlak et al., 2010; Yagishita et al., 2014). A complementary solution to this problem is to prolong the neural signals, in the form of elevated firing rates, that need to be associated by STDP. These neural signals are a form of memory trace activity that bridges the temporal gaps that

separate action and reward under typical behavioral conditions. Specifically, they represent information about the immediately preceding movement that can be used as part of a feedback-driven learning process to adjust future decisions based on a comparison between the expected and actual outcome of prior movements (Sutton and Barto, 1998). Such signals have been identified in several previous studies in frontal areas (Ding and Gold, 2012; Lee et al., 2012; So and Stuphorn, 2012; Tsujimoto et al., 2010; Tsunada et al., 2019). Our results provide an important complement to previous studies and demonstrate the existence of such signals in the FEF in learning movement-reward contingencies. FEF is an important interface cortical area between frontal regions and posterior visual cortex (Stanton et al., 1995) and oculomotor structures (Stanton et al., 1988). Its unique role in both saccadic target selection and in the deployment of visual spatial attention suggest that the memory representation we report here may also be useful in updating other forms of sensorimotor control such as altering attentional priorities or movement probabilities. Future studies will be needed to test this hypothesis directly.

We observed that post-movement memory information exists not only in the activity of FEF neurons but also in the majority of low-frequency band LFP responses; particularly in the beta band activity. During the post-movement period, the distributions of both prior-movement and subjective value information across the FEF LFP spectrum were similar to one another. In contrast, the distribution of reward-movement information across LFP spectra observed here was notably different from the distributions containing visual spatial information across LFP spectrum observed previously (Chen et al., 2018, 2020). Beta band FEF LFPs have been shown to contain the lowest visual spatial information compared to other frequency bands. Low-frequency frontal LFPs (alpha and beta bands) have been associated both with working memory (Bastos et al., 2018; Lundqvist et al., 2016, 2018; Salazar et al., 2012) and motor planning (Chen et al., 2010; Donoghue et al., 1998; Feingold et al., 2015). This low-frequency activity may be generated by the interaction between pyramidal neurons and local interneurons within the FEF with an excitatory drive provided by thalamocortical (Ketzer et al., 2015) and/or basal ganglia (Chatham and Badre, 2015) loops. It may also reflect a differential dopaminergic modulation of NMDA currents in excitatory and inhibitory neurons (Brunel and Wang, 2001; Durstewitz et al., 2000). Previous studies suggest that this low-frequency frontal activity may play an important role in modulating working memory, and in forming neural ensembles (Kopell et al., 2011; Miller et al., 2018). Consistent with that, the reduced beta power observed in our study during the high subjective value condition may reflect a reduced inhibition that facilitates the persistence of post-movement memory delay spiking activity in the FEF.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to William T. Newsome and Veit Stuphorn for invaluable scientific discussion; and Shellie Hyde and Danielle Abreu Lopes for assistance with animal care and husbandry.

Funding

This work was supported by NEI training fellowship to X.C. (K99EY029759) and by NEI RO1EY014924 to T.M.

References

- Akrami A, Kopec CD, Diamond ME, and Brody CD (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature* 554, 368–372. [PubMed: 29414944]
- Armstrong KM, Chang MH, and Moore T (2009). Selection and maintenance of spatial information by frontal eye field neurons. *J. Neurosci* 29, 15621–15629. [PubMed: 20016076]
- Barracough DJ, Conroy ML, and Lee D (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci* 7, 404. [PubMed: 15004564]
- Barto AG, Sutton RS, and Anderson CW (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans. Syst. Man Cybern* 834–846.
- Bastos AM, Loonis R, Kornblith S, Lundqvist M, and Miller EK (2018). Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory. *Proc. Natl. Acad. Sci* 115, 1117–1122. [PubMed: 29339471]
- Bittner KC, Milstein AD, Grienberger C, Romani S, and Magee JC (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science* 357, 1033–1036. [PubMed: 28883072]
- Brincat SL, and Miller EK (2016). Prefrontal cortex networks shift from external to internal modes during learning. *J. Neurosci* 36, 9739–9754. [PubMed: 27629722]
- Bruce CJ, and Goldberg ME (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *J. Neurophysiol* 53, 603–635. [PubMed: 3981231]
- Bruce CJ, Goldberg ME, Bushnell MC, and Stanton GB (1985). Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *J. Neurophysiol* 54, 714–734. [PubMed: 4045546]
- Brunel N, Wang X-J, 2001 Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of computational neuroscience* 11, 63–85. [PubMed: 11524578]
- Chandrasekaran C, Peixoto D, Newsome WT, and Shenoy KV (2017). Laminar differences in decision-related neural activity in dorsal premotor cortex. *Nat. Commun* 8, 1–16. [PubMed: 28232747]
- Chang C-C, and Lin C-J (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. TIST* 2, 27.
- Chatham CH, and Badre D (2015). Multiple gates on working memory. *Curr. Opin. Behav. Sci* 1, 23–31. [PubMed: 26719851]
- Chen X, and Stuphorn V (2015). Sequential selection of economic good and action in medial frontal cortex of macaques during value-based decisions. *Elife* 4, e09418. [PubMed: 26613409]
- Chen X, and Stuphorn V (2018). Inactivation of Medial Frontal Cortex Changes Risk Preference. *Curr. Biol*
- Chen X, Scangos KW, and Stuphorn V (2010). Supplementary motor area exerts proactive and reactive control of arm movements. *J. Neurosci. Off. J. Soc. Neurosci* 30, 14657–14675.
- Chen X, Zirnsak M, and Moore T (2018). Dissonant representations of visual space in prefrontal cortex during eye movements. *Cell Rep.* 22, 2039–2052. [PubMed: 29466732]
- Chen X, Zirnsak M, Vega GM, Govil E, Lomber SG, and Moore T (2019). The Contribution of Parietal Cortex to Visual Saliency. *BioRxiv* 619643.
- Chen X, Zirnsak M, Vega GM, Govil E, Lomber SG, and Moore T (2020). Parietal Cortex Regulates Visual Saliency and Saliency-Driven Behavior. *Neuron*.
- Cunningham JP, and Byron MY (2014). Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci* 17, 1500. [PubMed: 25151264]
- Ding L, and 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. [PubMed: 21765183]
- Ding L, and Hikosaka O (2006). Comparison of reward modulation in the frontal eye field and caudate of the macaque. *J. Neurosci* 26, 6695–6703. [PubMed: 16793877]

- Donoghue JP, Sanes JN, Hatsopoulos NG, and Gaál G (1998). Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements. *J. Neurophysiol* 79, 159–173. [PubMed: 9425187]
- Drew PJ, and Abbott LF (2006). Extending the effects of spike-timing-dependent plasticity to behavioral timescales. *Proc. Natl. Acad. Sci* 103, 8876–8881. [PubMed: 16731625]
- Durstewitz D, Seamans JK, Sejnowski TJ, 2000 Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of neurophysiology* 83, 1733–1750. [PubMed: 10712493]
- Farashahi S, Azab H, Hayden B, and Soltani A (2018). On the flexibility of basic risk attitudes in monkeys. *J. Neurosci* 38, 4383–4398. [PubMed: 29626169]
- Fecteau JH, and Munoz DP (2003). Exploring the consequences of the previous trial. *Nat. Rev. Neurosci* 4, 435–443. [PubMed: 12778116]
- Feingold J, Gibson DJ, DePasquale B, and Graybiel AM (2015). Bursts of beta oscillation differentiate postperformance activity in the striatum and motor cortex of monkeys performing movement tasks. *Proc. Natl. Acad. Sci* 112, 13687–13692. [PubMed: 26460033]
- Fusi S, Miller EK, and Rigotti M (2016). Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol* 37, 66–74. [PubMed: 26851755]
- Genovesio A, Brasted PJ, and Wise SP (2006). Representation of future and previous spatial goals by separate neural populations in prefrontal cortex. *J. Neurosci* 26, 7305–7316. [PubMed: 16822988]
- Gerstner W, Lehmann M, Liakoni V, Corneil D, and Brea J (2018). Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Front. Neural Circuits* 12, 53. [PubMed: 30108488]
- Glaser JI, Wood DK, Lawlor PN, Ramkumar P, Kording KP, and Segraves MA (2016). Role of expected reward in frontal eye field during natural scene search. *J. Neurophysiol* 116, 645–657. [PubMed: 27169506]
- Glimcher PW, and Fehr E (2013). *Neuroeconomics: Decision making and the brain* (Academic Press).
- Goldberg ME, and Bushnell MC (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. II. Modulation in frontal eye fields specifically related to saccades. *J. Neurophysiol* 46, 773–787. [PubMed: 7288464]
- He K, Huertas M, Hong SZ, Tie X, Hell JW, Shouval H, and Kirkwood A (2015). Distinct eligibility traces for LTP and LTD in cortical synapses. *Neuron* 88, 528–538. [PubMed: 26593091]
- Hsu M, Krajbich I, Zhao C, and Camerer CF (2009). Neural response to reward anticipation under risk is nonlinear in probabilities. *J. Neurosci* 29, 2231–2237. [PubMed: 19228976]
- Hull CL (1943). *Principles of behavior: An introduction to behavior theory*.
- Hwang EJ, Dahlen JE, Mukundan M, and Komiyama T (2017). History-based action selection bias in posterior parietal cortex. *Nat. Commun* 8, 1–14. [PubMed: 28232747]
- Izhikevich EM (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb. Cortex* 17, 2443–2452. [PubMed: 17220510]
- Kastner S, Pinsk MA, De Weerd P, Desimone R, and Ungerleider LG (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22, 751–761. [PubMed: 10230795]
- Ketz NA, Jensen O, and O'Reilly RC (2015). Thalamic pathways underlying prefrontal cortex–medial temporal lobe oscillatory interactions. *Trends Neurosci.* 38, 3–12. [PubMed: 25455705]
- Kim S, Bobeica I, Gamo NJ, Arnsten AF, and Lee D (2012). Effects of α -2A adrenergic receptor agonist on time and risk preference in primates. *Psychopharmacology (Berl.)* 219, 363–375. [PubMed: 21979441]
- Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi X-L, Romo R, Uchida N, and Machens CK (2016). Demixed principal component analysis of neural population data. *Elife* 5, e10989. [PubMed: 27067378]
- Kopell N, Whittington MA, and Kramer MA (2011). Neuronal assembly dynamics in the beta1 frequency range permits short-term memory. *Proc. Natl. Acad. Sci* 108, 3779–3784. [PubMed: 21321198]

- Lee D, Seo H, and Jung MW (2012). Neural basis of reinforcement learning and decision making. *Annu. Rev. Neurosci* 35, 287–308. [PubMed: 22462543]
- Lundqvist M, Rose J, Herman P, Brincat SL, Buschman TJ, and Miller EK (2016). Gamma and beta bursts underlie working memory. *Neuron* 90, 152–164. [PubMed: 26996084]
- Lundqvist M, Herman P, Warden MR, Brincat SL, and Miller EK (2018). Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun* 9, 394. [PubMed: 29374153]
- McCoy AN, and Platt ML (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. *Nat. Neurosci* 8, 1220. [PubMed: 16116449]
- Miller EK, Lundqvist M, and Bastos AM (2018). Working Memory 2.0. *Neuron* 100, 463–475. [PubMed: 30359609]
- Minsky M (1961). Steps toward artificial intelligence. *Proc. IRE* 49, 8–30.
- Neftci EO, and Averbek BB (2019). Reinforcement learning in artificial and biological systems. *Nat. Mach. Intell* 1, 133–143.
- Pawlak V, Wickens JR, Kirkwood A, and Kerr JN (2010). Timing is not everything: neuromodulation opens the STDP gate. *Front. Synaptic Neurosci* 2, 146. [PubMed: 21423532]
- Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, and Fusi S (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585. [PubMed: 23685452]
- Roesch MR, and Olson CR (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J. Neurophysiol* 90, 1766–1789. [PubMed: 12801905]
- Salazar RF, Dotson NM, Bressler SL, and Gray CM (2012). Content-specific fronto parietal synchronization during visual working memory. *Science* 338, 1097–1100. [PubMed: 23118014]
- Sendhilnathan N, Basu D, and Murthy A (2017). Simultaneous analysis of the LFP and spiking activity reveals essential components of a visuomotor transformation in the frontal eye field. *Proc. Natl. Acad. Sci* 114, 6370–6375. [PubMed: 28572407]
- So N, and Stuphorn V (2010). Supplementary eye field encodes option and action value for saccades with variable reward. *J. Neurophysiol* 104, 2634–2653. [PubMed: 20739596]
- So N, and Stuphorn V (2012). Supplementary eye field encodes reward prediction error. *J. Neurosci* 32, 2950–2963. [PubMed: 22378869]
- Soltani A, Chaisangmongkon W, and Wang X-J (2017). Neural circuit mechanisms of value-based decision-making and reinforcement learning. In *Decision Neuroscience*, (Elsevier), pp. 163–176.
- Stanton GB, Goldberg ME, and Bruce CJ (1988). Frontal eye field efferents in the macaque monkey: II. Topography of terminal fields in midbrain and pons. *J. Comp. Neurol* 271, 493–506. [PubMed: 2454971]
- Stanton GB, Bruce CJ, and Goldberg ME (1995). Topography of projections to posterior cortical areas from the macaque frontal eye fields. *J. Comp. Neurol* 353, 291–305. [PubMed: 7745137]
- Stauffer WR, Lak A, Bossaerts P, and Schultz W (2015). Economic choices reveal probability distortion in macaque monkeys. *J. Neurosci* 35, 3146–3154. [PubMed: 25698750]
- Sugrue LP, Corrado GS, and Newsome WT (2004). Matching behavior and the representation of value in the parietal cortex. *Science* 304, 1782–1787. [PubMed: 15205529]
- Sutton RS, and Barto AG (1998). *Introduction to reinforcement learning* (MIT press Cambridge).
- Thompson KG, Biscoe KL, and Sato TR (2005). Neuronal basis of covert spatial attention in the frontal eye field. *J. Neurosci* 25, 9479–9487. [PubMed: 16221858]
- Tsujimoto S, Genovesio A, and Wise SP (2010). Evaluating self-generated decisions in frontal pole cortex of monkeys. *Nat. Neurosci* 13, 120. [PubMed: 19966838]
- Tsunada J, Cohen Y, and Gold JI (2019). Post-decision processing in primate prefrontal cortex influences subsequent choices on an auditory decision-making task. *ELife* 8, e46770. [PubMed: 31169495]
- Tversky A, and Kahneman D (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–291.

Yagishita S, Hayashi-Takagi A, Ellis-Davies GC, Urakubo H, Ishii S, and Kasai H (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* 345, 1616–1620. [PubMed: 25258080]
Yarbus AL (1967). *Eye movements and vision* (Plenum Press).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- FEF neuronal activity correlated with expected reward value of different movements in an oculomotor gamble task.
- Movement direction and expected reward signals persisted after movements were completed.
- Post-movement signals were also present in low-frequency LFPs.

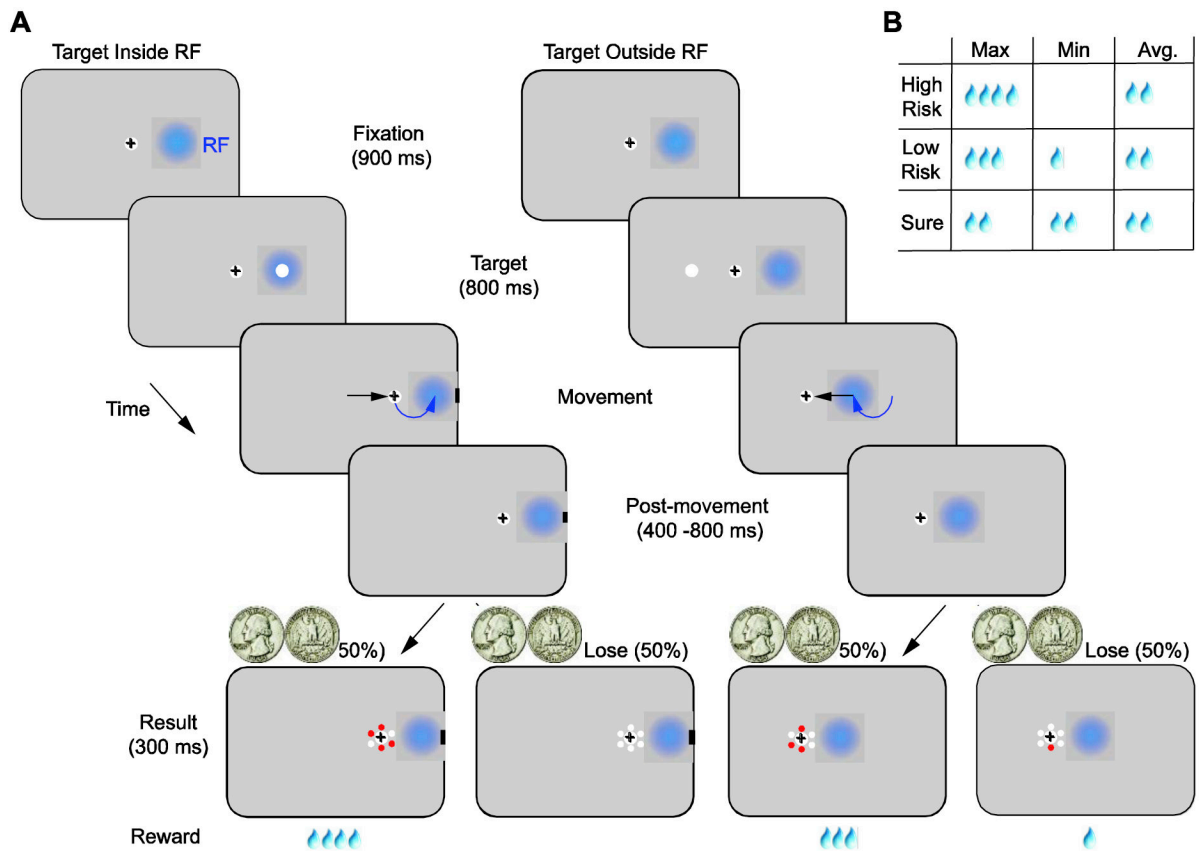


Fig. 1. An oculomotor gamble task.

(A) In the task, monkeys made saccadic eye movements to targets with differing, delayed reward outcomes. The visual targets (white dot) were identical, while the reward outcomes of movements to different target locations differed between blocks. In each trial, following fixation, one target appeared either inside (e.g left) or outside of (e.g right) the RFs of FEF neurons (blue disks). After each movement, monkeys maintained fixation on the target for an additional 400–800 ms (post-movement period), after which the reward amount was cued by the number of red dots appearing around the fixated target (result onset). Juice reward (drops) was delivered 300 ms after the result onset. Each block started with 20 forced-choice trials and was followed by 20–40 free-choice trials. In the forced-choice trials, two visual targets appeared on the display, and monkeys were free to choose one of them. Black crosses denote gaze position. Black arrows on the display show the direction of the saccade and curved, blue arrows show the changes in RF position caused by the saccade. (B) Minimum, maximum, and average reward amounts in the High-risk, Low-risk and Sure conditions. In all conditions, average reward outcomes were the same, but the variance of the reward outcome varied.

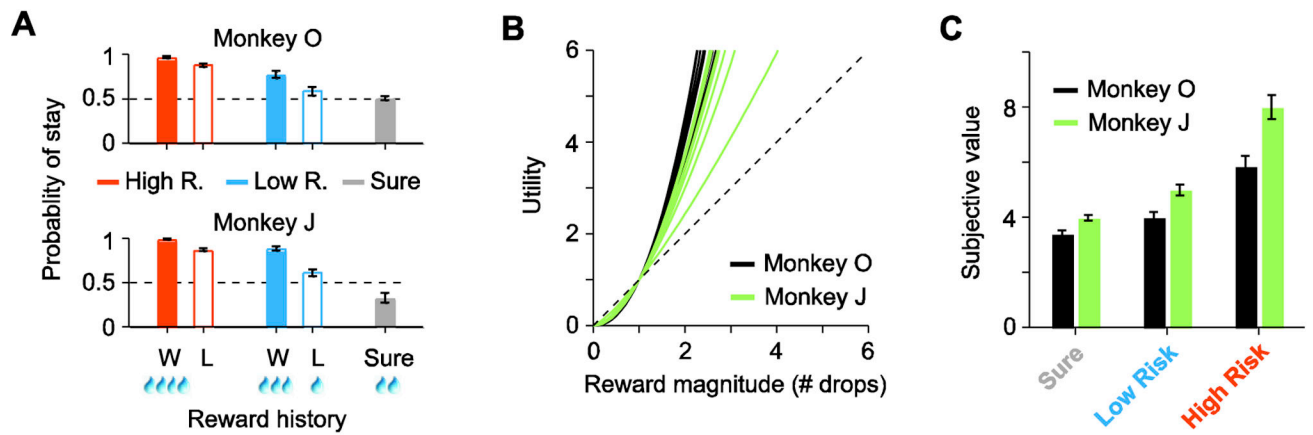


Fig. 2. Risk-seeking behavior during free-choice trials.

(A) Frequency of repeating the same choice (stay) for different juice outcomes of the previous trial in different risk conditions. (B) Estimated power utility functions for both monkeys. Each line shows individual session estimates for the two monkeys. (C) Estimated subjective value for each risk condition for both monkeys. Error bars denote \pm SEM.

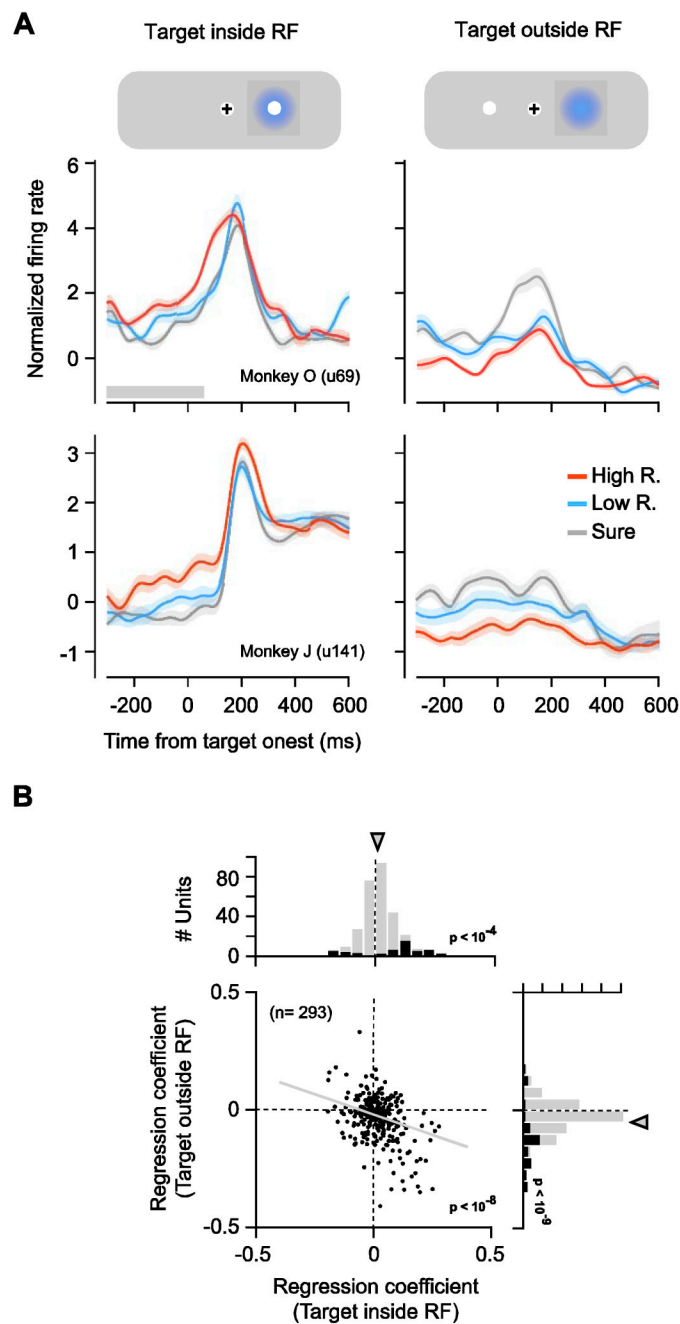


Fig. 3. Neuronal responses were correlated with subjective value during the pre-target period. (A) Mean neuronal responses from two example recordings shown for different subjective value conditions for targets appearing inside (left) and outside of (right) the RF. The shaded region denotes \pm SEM. Gray bar at the bottom indicates the time epoch used in the analyses. (B) Regression coefficients for the subjective value of the targets appearing inside and outside the RF across all recordings. Black bars in marginal distributions denote individually significant coefficients. Arrows indicate means.

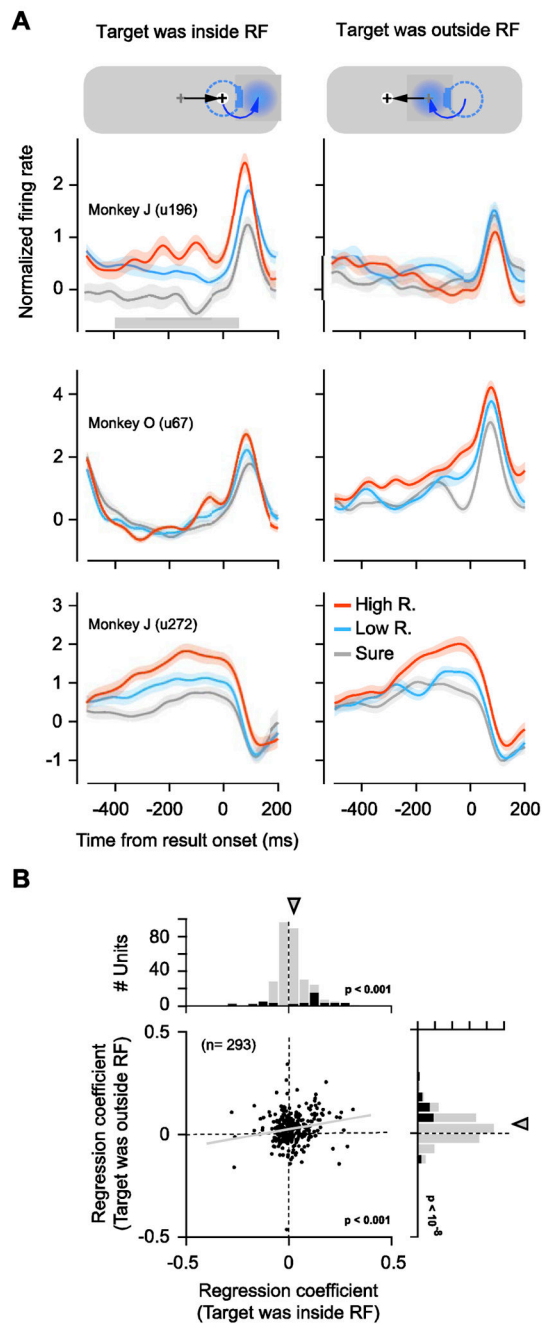


Fig. 4. Neuronal responses remained correlated with subjective value during the post-movement period.

(A) Mean neuronal responses from three example recordings shown for different subjective value conditions for targets appearing inside (left) and outside of (right) the RF. In the above diagram, the gray cross denotes the prior fixation location and the black cross denotes the current fixation location. The dotted circle indicates the prior RF location, before the saccade, and the blue disk indicates the current neuronal RF. Black arrows on the screen represent the direction of the saccade and blue arrows represent the changes of the RF

location by saccade. (B) Regression coefficients for the subjective value of the prior targets during the post-movement period. Same conventions as in Fig. 5.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

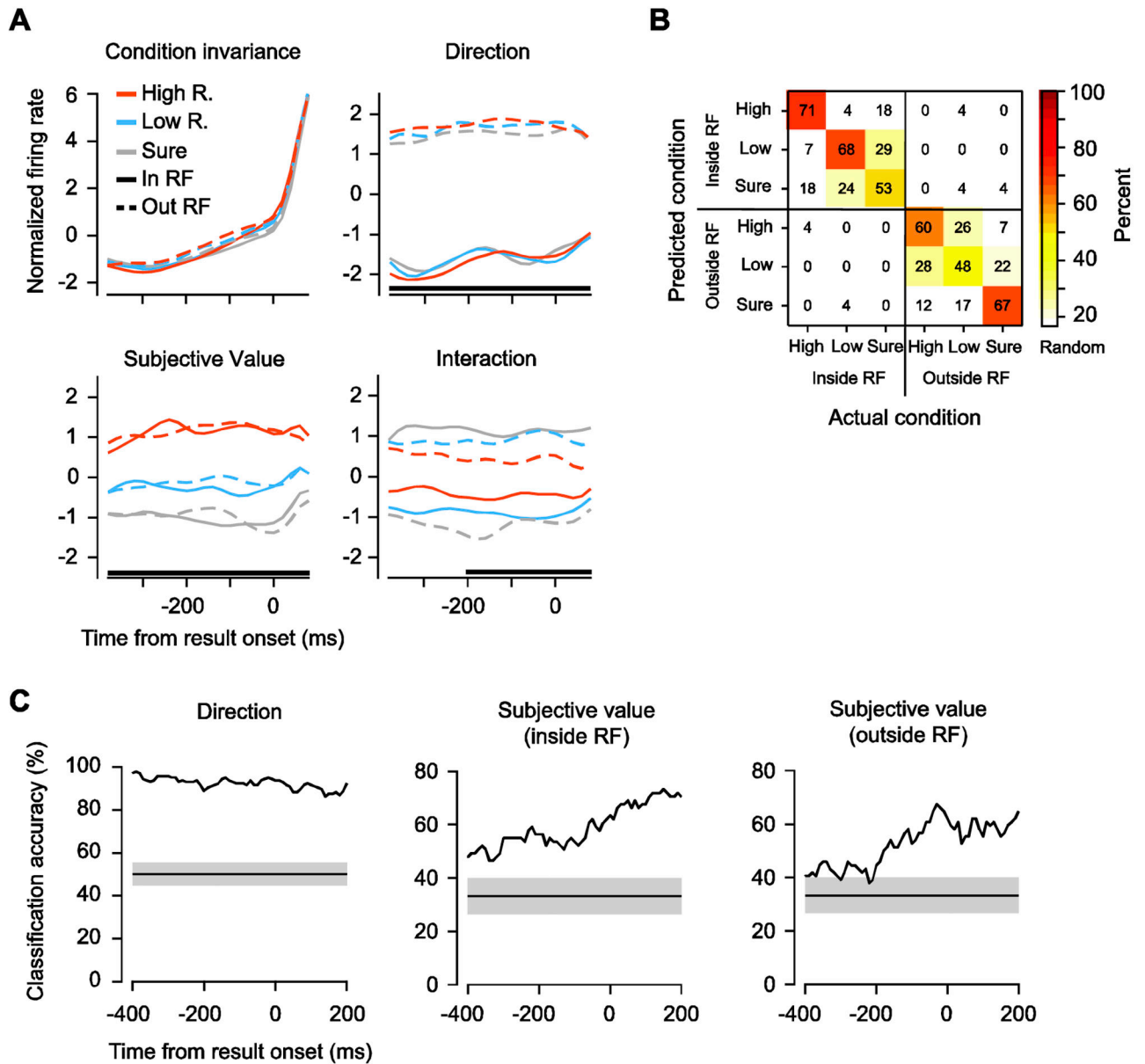


Fig. 5. Dynamics of FEF population activity signal both direction and subjective value of prior targets.

(A) Demixed principal components of the gamble task during the post-movement period. Time course of the projection of single components aligned on result onset. Top left: the condition-invariant component; top right: the target direction component; bottom left: reward component; bottom right: the interaction component between direction and reward. Black lines at bottom denote the time intervals during which the respective task parameter can be reliably extracted from population activity using linear classification. For each feature, the first component that captures the most variance is shown. (B) Performance of a linear classifier in distinguishing the 6 task conditions during the post-movement period. Numbers indicate the percentage of trials of each condition (x-axis) predicted by the classifier (y-axis). The number on the diagonal axis indicates the hit rates. (C) Temporal

profiles of classification accuracy in distinguishing prior target direction (left) and subjective value of prior targets inside (middle) and outside of (right) the RF. Shaded gray regions show the distribution of classification accuracies expected by chance as estimated by 1000 iterations of shuffling procedure.

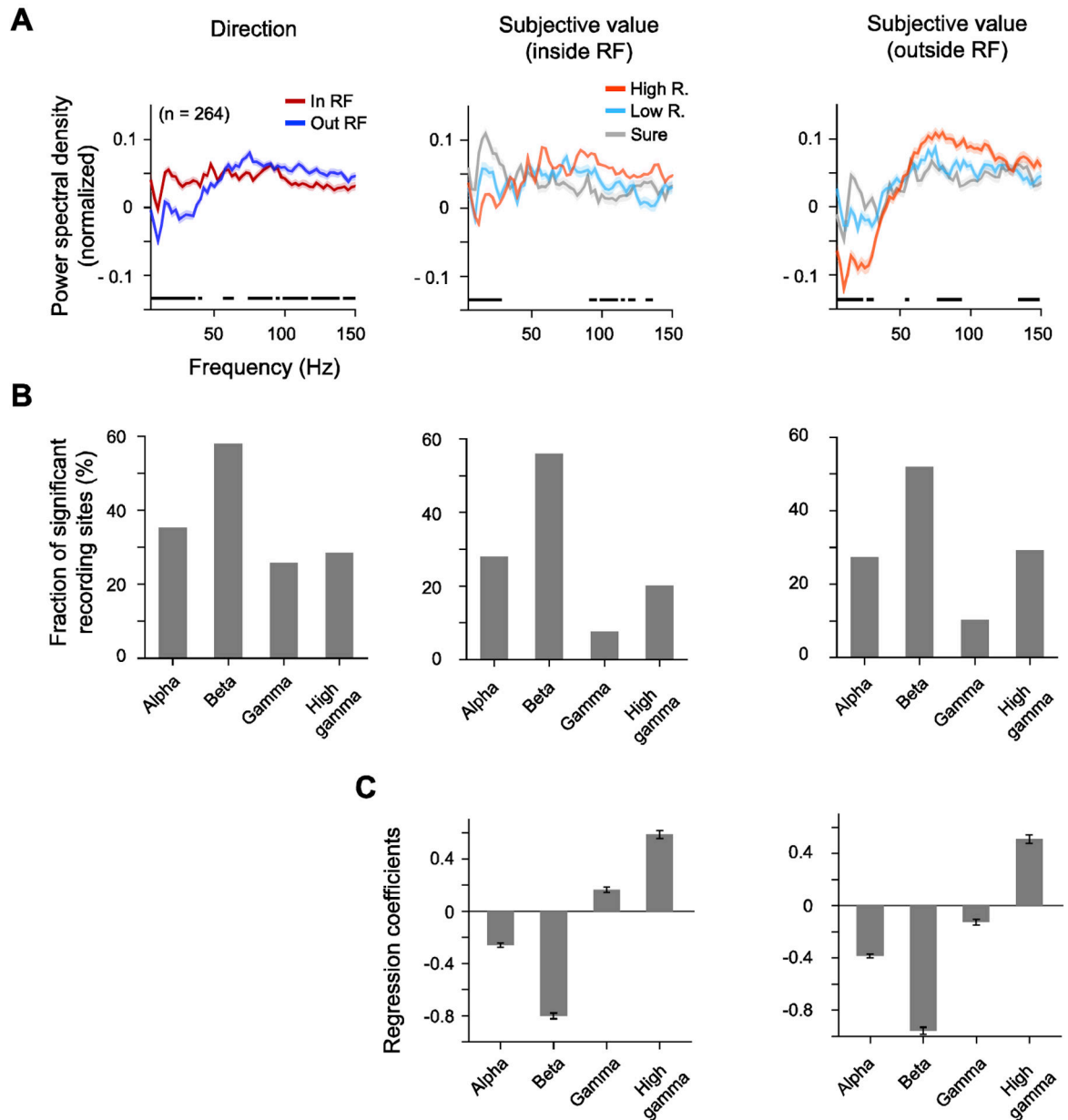


Fig. 6. Representation of prior movement direction and subjective value in FEF LFPs during the post-movement period.

(A) Normalized LFP power spectra shown for different movement directions (left), and subjective values of targets presented inside (middle) and outside (right) of the RF. The shaded area denotes \pm SEM and black lines at bottom denote significance. (B) The fraction of recording sites significantly modulated by prior movement direction (left), the subjective value of the targets presented inside (middle) and outside of the RF (right). (C) Regression coefficients for the subjective value of targets presented inside (left) and outside (right) of the RF. Error bars indicate standard error.