Title: *Frontal and parietal neurons encode reward prediction errors in multiple reference frames*

Authors: Nicholas C. Foley[1,2], Michael Cohanpour[1,2], Mulugeta Semework[1,2], Sameer A. Sheth[3] and Jacqueline Gottlieb[1,2,4]

Author affiliation: [1]Department of Neuroscience, [2] Zuckerman Mind Brain Behavior Institute, Columbia University, [3] Department of Neurosurgery, Baylor College of Medicine, [4]The Kavli Institute for Brain Science, Columbia University,

Corresponding author:

Jacqueline Gottlieb, PhD
Department of Neuroscience
Columbia University
3227 Broadway,
New York, NY 10025
E-mail: jg2141@ columbia.edu

Number of figures: 8

Acknowledgements: The work was supported by a Memory and Cognitive Disorder Award from the McKnight Foundation to JG.

Author Contributions: NCF and JG designed the experiment. NCF, SAS, MS and JG implemented the experiment and collected the data. NCF and MC analyzed the data. JG and NCF wrote the paper.

Disclosure statement:

The authors declare that they have no conflict of interest.

Abstract

Anticipated and experienced rewards modulate cognitive control and attention, but the mechanisms of these modulations are poorly understood. We compared neuronal responses in monkey dorsolateral prefrontal cortex and parietal area 7A, two areas strongly implicated in visual attention and working memory, related to probabilistic rewards signaled by familiar visual cues. Neurons in both areas encoded the magnitude, probability and expected value (EV) of the reward signaled by the cue. Strikingly, neurons also encoded across-trial memories of recent rewards, which, although statistically irrelevant to a trial's expectations, correlated with the monkeys' behavioral sensitivity to reward history. Finally, upon outcome delivery, neurons combined responses to the experienced outcome with renewed sensitivity to EV and reward history, allowing for population-level decoding of reward prediction errors (RPEs) relative to the trial's EV and reward history. Frontal and parietal areas explicitly encode expected and experienced outcomes and provide information relevant to computing model-free and model-based RPEs.

Introduction

Recent theoretical models of computational rationality and rational inattention postulate that utility-relevant outcomes such as the punishments or rewards possible in a task modulate not only an individual's choices of action, but the cognitive resources such as memory and attention that the individual allocates to the task[1, 2]. Despite increasing support for this view, empirical studies of the links between rewards and cognition are in their infancy. A topic of longstanding interest is the nature of reward-related activity in networks of executive control and attention[3].

The expected value of control theory, a recent conceptual framework of executive function, postulates that cognitive control involves two processes of, respectively, monitoring and regulation[4]. A medial frontal area, the anterior cingulate cortex (ACC), is thought to monitor the value of a task and choose the level of control (cognitive effort) to allocate to the task. More dorsal frontal and parietal areas, in contrast, are postulated to play a regulatory role in implementing the working memory and attentional policies selected by the ACC.

Two dorsal areas implicated in cognitive regulation are the dorsolateral prefrontal cortex (dlPFC) and parietal area 7A[5], which are adjacent to, but functionally distinct from the frontal eye field (FEF) and the lateral intraparietal area (LIP) that are associated with eye movements and visual attention. dlPFC and 7A neurons are sensitive to visual salience and target selection but, unlike FEF and LIP cells, have weaker visual responses and pre-saccadic activity[6], and instead are more sensitive to non-spatial factors such as task context and rules[7, 8]. Anatomically, 7A and dlPFC lack a strong connectivity to the superior colliculus but are reciprocally connected and project to FEF and LIP[9, 10]. Thus, these areas are excellent candidates for providing executive regulation, including sensitivity to incentives, to oculomotor and visual structures.

However, reward-related activity in these areas is incompletely characterized, especially as it relates to expectancy and surprise. The reinforcement learning literature postulates that cognitive regulation is sensitive to reward expectancy, and specifically to reward prediction errors (RPE) – the extent to which an outcome violates prior expectations. *Signed* RPEs, indicating whether an outcome is better or worse than expected, and their *unsigned* counterparts (absolute value of a signed RPE indicating general expectancy violation) are proposed to signal the need to reduce uncertainty by regulating learning rates, attention and information demand[11-13]. Moreover, computational studies show that RPEs can be calculated using model-based algorithms, based on probability distributions relevant to a current task state, or using model-free algorithms that rely only on reward history independently of task states to produce a potentially suboptimal but more computationally frugal cognitive strategy[14].

Despite the compelling arguments in the theoretical literature, little is known about the encoding of reward expectations in areas directly involved in cognitive regulation. A well-characterized neural representation of signed RPEs is found in midbrain dopamine (DA) cells, which show excitatory and inhibitory responses to outcomes that are, respectively, better or worse than anticipated[15]. In frontal cortical areas, however, RPE-like signals have been reported in the ACC, but their specific features are under debate[16, 17]. Reward-related activity in the dlPFC has only been reported in complex decision and learning tasks in which expectations and RPEs are difficult to infer (e.g.,[18-21]). 7A neurons have not been reported to have reward-related activity.

To examine these questions, we simultaneously recorded the responses of dlPFC and 7A neurons while monkeys performed a simple instructed saccade task in which reward magnitudes and probabilities varied randomly from trial to trial and were signaled by visual cues. By measuring the

monkeys' anticipatory licking as a behavioral index of expectancy, we could analyze the neural representations of reward expectation/surprise independently of the decision strategies.

We show that dlPFC and 7A neurons convey both explicit and implicit population-level information relevant to computing model-based and model-free RPEs. Neurons in both areas encoded the reward magnitude and probability conveyed by the visual cues and integrated this information to code the trial's expected value (EV). Strikingly, the neurons also encoded the value of the previous trial outcome that was statistically irrelevant to the current trial EV. These neural responses correlated with the monkeys' behavioral sensitivity to reward history, suggesting that they may underlie history-dependent biases that have been documented in the behavioral economics literature, such as overconfidence and the hot-hand fallacy[22]. Finally, upon outcome delivery, neurons in both areas had robust responses to reward size, alongside renewed sensitivity to the trial's EV and reward history. Responses to expected and experienced outcomes were intricately and asymmetrically organized, allowing for population-level decoding of rewards that were surprising in two reference frames: relative to the relevant, cue-signaled EV and relative to the statistically irrelevant reward history. dlPFC and 7A neurons thus convey rich information relevant to computing model-free and model-based reward expectations and RPEs.

Results

Task and behavior Two monkeys performed a visually guided saccade task in which they formed expectations about probabilistic rewards based on trial-specific visual cues (**Fig. 1A**). On each trial after achieving central fixation, the monkeys viewed a visual cue that signaled the magnitude and probability of the reward available on the trial (**Fig. 1A, cue**). The cue was presented for 300 ms at a peripheral location to the right or left of fixation and was followed, after a 600 ms memory period, by the presentation of the target for a subsequent saccade. After making the required saccade and completing a 350 ms post-saccadic fixation, the monkeys received the outcome – reward receipt or omission - according to the schedule predicted by the cue. By requiring monkeys to make instructed saccades, we could examine responses to expected rewards independently of decision strategies. Importantly, because the locations of the cue and saccade target were independently randomized, the cue only provided information about reward contingencies and not about the subsequent action.

To examine reward-related responses, the monkeys were familiarized with 20 distinct visual cues that signaled 10 combinations of reward probability and magnitude with 7 distinct levels of expected value (EV; the product of reward probability and magnitude; **Fig. 1B**). The reward contingencies included conditions that varied in reward probability but not magnitude or in reward magnitude but not probability (horizontal and vertical axes in **Fig. 1B**). The cues were colored checkerboard patterns that were controlled for discriminability, with the cue-reward associations counterbalanced across monkeys (**Fig. S1A**). Two distinct cues were assigned to each magnitude/probability combination to control for visual selectivity and all 20 cues were presented in random order throughout each daily session (**Fig. S1A**).

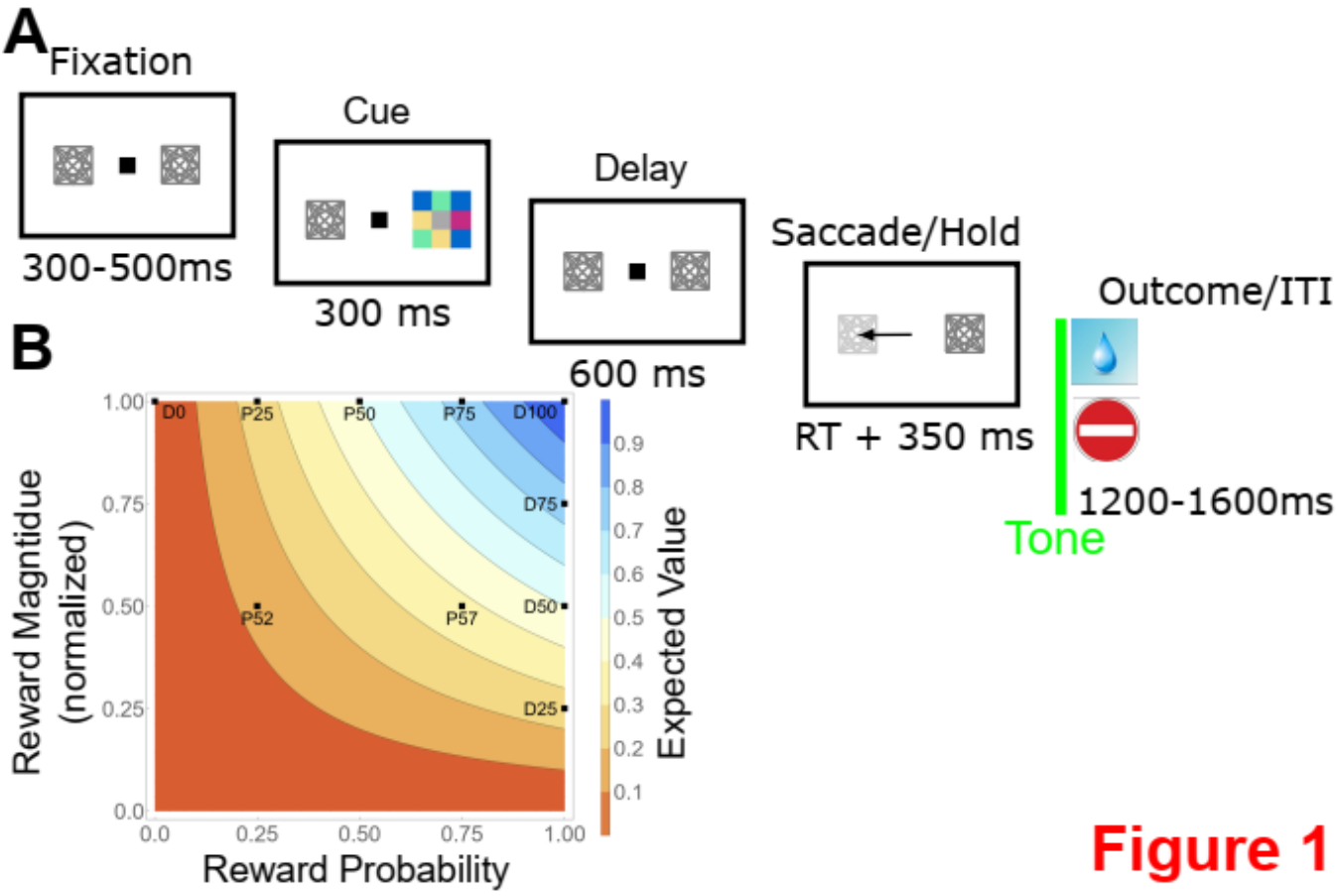

Figure 1. Task (A) A trial began with the presentation of a fixation point (small black square) flanked by two placeholders (gray squares). After the monkeys maintained fixation for 300-500 ms ("Fixation"), they were shown a reward cue for 300 ms ("Cue", colored checkerboard) at a randomly selected placeholder location, followed by a delay period ("Delay") and presentation of the saccade target ("Saccade & Hold", bright square). After maintaining fixation on the target for 350 ms, the monkeys heard a 200 ms tone signaling the onset of the outcome period, and received the outcome (reward omission, or a drop of water of the size that had been predicted by the cue). Outcome delivery was followed by an intertrial interval (ITI) during which eye position was unconstrained. (**B**) The entire set of reward contingencies that the monkeys experienced comprised 10 unique combinations of reward probability and magnitude (black points). "P" cues indicated rewards delivered probabilistically with 0.25, 0.5 or 0.75 probability and maximal size (400 ms, corresponding to 1.0 on the normalized scale; P25, P50 and P75, respectively) or with 0.25 and 0.75 probability and half size (P52 and P57, respectively). "D" cues indicated rewards delivered deterministically with either 0 probability (D0, a sure reward omission, which was coded as 0 probability at 1.0 magnitude in the analyses) or with 100% probability at different magnitudes (D25, D50, D100). Areas of constant EV are shown in color; the 10 cue contingencies indicated 7 distinct levels of EV.

We measured the monkeys' anticipatory licking as an index of their reward expectancy. After viewing the cue, licking scaled monotonically as a function of the cue-predicted EV, an effect that grew throughout the delay period (**Fig. 2A,** green trace). Regression analysis (*Methods, Eq. 3*) showed that the effect of EV was highly significant in each monkey (**Fig. 2A,** green marginal histograms; EV coefficient in the pre-tone epoch, mean ± standard error (SE): monkey 1: 0.21±0.01, $p < 10^{-10}$ relative to 0; monkey 2: 0.05±0.015, p = 0.0006). The response to EV had a quasi-categorical pattern, distinguishing most clearly between the two highest levels and 5 lower levels of EV (**Fig. 2B**, right),

suggesting that the monkeys tended to summarize the large set of contingencies as representing “high” or “low” EV.

Because reward contingencies were trial-wise randomized and explicitly cued, the history of recent rewards was statistically irrelevant to a trial’s expectations. Nevertheless, the monkeys’ licking was strongly sensitive to the reward on the previous trial (**Fig. 2A,** purple trace, Eq. 2). During the fixation period preceding cue presentation, licking scaled monotonically with previous reward size (PR; **Fig. 2B**, left) and the PR coefficient was significant in 79% of individual sessions and on average in each monkey; (**Fig**. **2A,** left purple histograms; monkey 1: 0.21±0.006, $p < 10^{-10}$; monkey 2: 0.04±0.01, p = 0.0001). The effect of PR remained significant in the epoch immediately preceding the outcome, when it coexisted with a significant effect of EV (**Fig**. **2A,** right purple histograms; monkey 1: 0.027±0.007, p = 0.0005; monkey 2: 0.054±0.01, $p = 10^{-6}$). PR sensitivity was driven mostly by the immediately preceding reward with no significant influence of earlier trials (Eq. 4; **Fig. 2B**, left, inset; beta PR for trial -1, 0.75±0.02 for mk 1, $p < 10^{-10}$; 0.65±0.04 for mk2, $p < 10^{-5}$; trials -2 to -5, all p > 0.2).

A potential explanation for the persistent influence of PR is that the monkeys used reward history to continuously update their value estimates for probabilistic cues. To examine this hypothesis, we examined the effects of cue-specific reward history, by comparing licking as a function of the last outcome that had been experienced for a specific cue, or pairs of cues signaling the same probability. The regression coefficients for these cue-specific prior rewards were not significantly different from 0, showing that the monkeys had stable estimates of probabilities (**Fig. S1B**). Therefore, the marked effect of PR seems to have occurred by default, independently of its relevance for learning or the current trial EV.

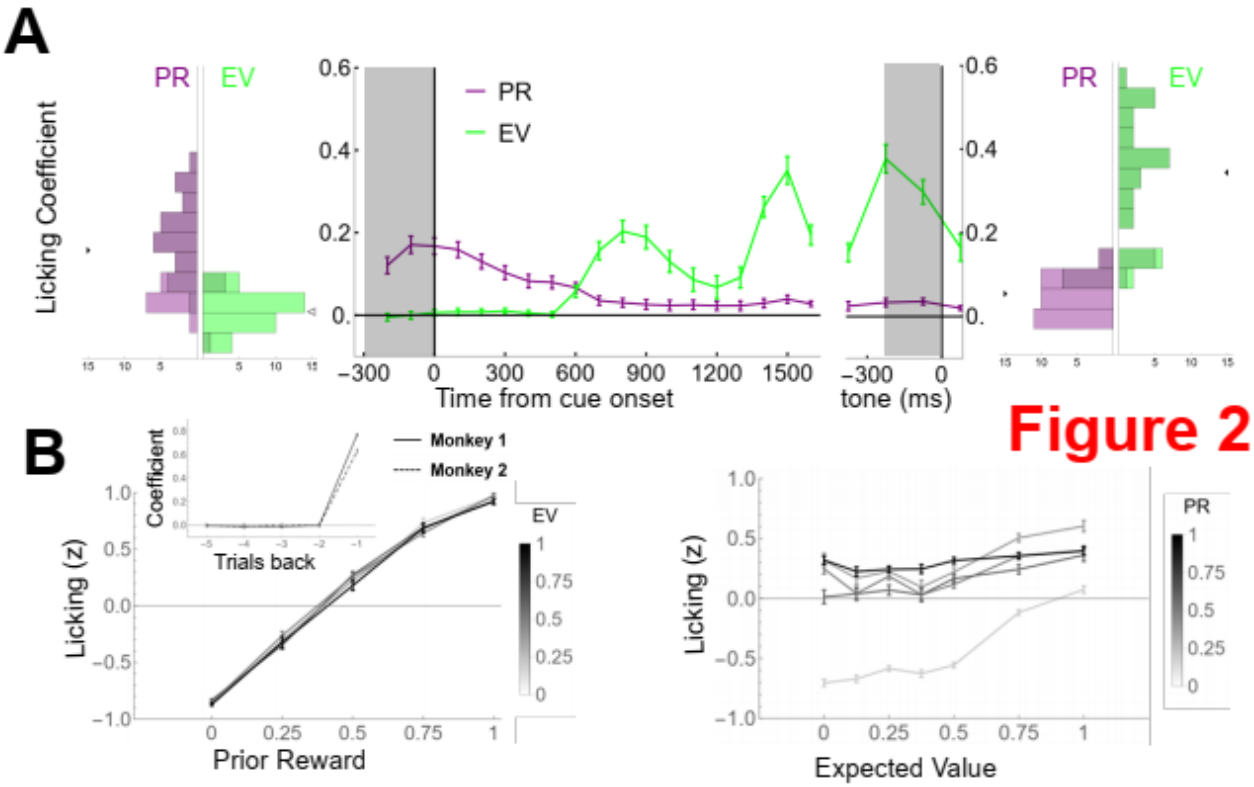

FIGURE 2. BEHAVIOR (A) LICKING IS SENSITIVE TO EV AND PR THROUGHOUT THE TRIAL. THE TRACES IN THE MIDDLE PANEL SHOW THE COEFFICIENTS ESTIMATING THE EFFECTS OF PR (PURPLE) AND EV (GREEN) FROM A REGRESSION

ANALYSIS (*METHODS*, EQS. 2 AND 3) COMPUTED IN CONSECUTIVE 100 MS TIME BINS ALIGNED ON CUE AND TONE ONSET. THE MARGINAL HISTOGRAMS SHOW THE DISTRIBUTIONS OF PR AND EV COEFFICIENTS ACROSS SESSIONS, IN THE 200 MS WINDOWS PRECEDING CUE ONSET (LEFT) AND TONE ONSET (RIGHT). THE TRIANGLES SHOW THE MEANS OF THE DISTRIBUTIONS, WITH FILLED COLORS INDICATING $p < 0.05$ RELATIVE TO 0. **(B) AVERAGE EFFECTS OF PR AND EV.** THE PANELS SHOW THE AVERAGE LR (MEAN AND SEM ACROSS SESSIONS) IN THE TIME WINDOWS HIGHLIGHTED IN A. FOR THE PRE-CUE EPOCH (LEFT), LR ARE SHOWN AS A FUNCTION OF PR AND SEPARATED ACCORDING TO EV (GRAYSCALE). FOR THE PRE-OUTCOME EPOCH (RIGHT), LR ARE PLOTTED AS A FUNCTION OF EV AND SEPARATED ACCORDING TO PR (GRAYSCALE). THE INSET SHOWS THE COEFFICIENTS OF PR (*METHODS*, EQ. 2) FOR 5 PRECEDING TRIALS (MEAN AND SEM ACROSS SESSIONS FOR EACH MONKEY).

Frontal and parietal neurons encode expected and experienced outcomes

To examine the neuronal encoding of reward expectations, we recorded spiking activity using "Utah" electrode arrays implanted in the pre-arcuate portion of the dlPFC and the posterior portion of area 7A (**Fig. S2**). We describe the responses of 2,034 neurons, of which 1,298 were in the dlPFC (917 in monkey 1) and 736 in area 7A (381 in monkey 1). We first describe the neurons' responses to reward anticipation during the pre-cue and delay periods, and to the experienced outcome during the intertrial interval (ITI), and end by describing their integration of signals of expected and experienced outcomes during the ITI. For all sections, detailed statistical analyses and comparisons between areas are presented in **Tables 1** and **2**.

Encoding of EV and PR during reward anticipation

In the pre-outcome epochs, dlPFC and 7A neurons conveyed information about PR and EV. To isolate the effects of these factors we evaluated each cell's firing rates with regression analysis that included EV and PR as covariates, along with cue location as a nuisance regressor to control for visuo-spatial selectivity (*Methods*, eq. 5).

During the delay period, 18% of cells in each area showed sensitivity to EV (**Fig. 3A; Table 1**). The sensitive cells showed both positive and negative scaling (increases or decreases of firing with increase in EV) with a small areal asymmetry whereby positive scaling was slightly more common in the dlPFC and negative scaling slightly more prevalent in 7A (**Table 1**). In both areas, EV sensitivity arose at median latencies longer than 300 ms after cue presentation and was sustained throughout the delay period (**Fig. 3A**, bottom, **Table 1**) indicating that it was not a mere visual response to the cues. Although the sensitive cells showed stronger modulations in the dlPFC (**Fig. 3A**, bottom, **Table 1**) the effect latencies did not significantly differ between the two areas (**Fig. 3A**, bottom, **Table 1**).

Selectivity to PR was as uncorrelated with sensitivity to EV (**Table 2**; $r = 0.06$ in dlPFC and $r = -0.05$ in 7A) and found in more than a third of the cells in each area (48% in the dlPFC and 35% in 7A; **Fig. 3B** top; **Table 1**). In both areas, PR-sensitive cells showed predominantly positive scaling (**Fig. 3B** top; **Table 1**) and sustained sensitivity throughout the fixation, cue and delay periods (**Fig. 3B**, bottom). However, PR sensitivity arose with latencies of more than 100 ms after fixation onset (**Fig. 3B**, bottom; **Table 1**), showing that it was not a mere persistence of the activity from the previous trial (a point to which we return below). PR-sensitive cells were more prevalent and showed stronger modulations in the dlPFC and, importantly, also had shorter latencies in dlPFC relative to 7A (**Fig. 3B, Table 1**). Thus, while dlPFC and 7A provided simultaneous information about the trial's EV, PR sensitivity was considerably stronger and arose earlier in the dlPFC.

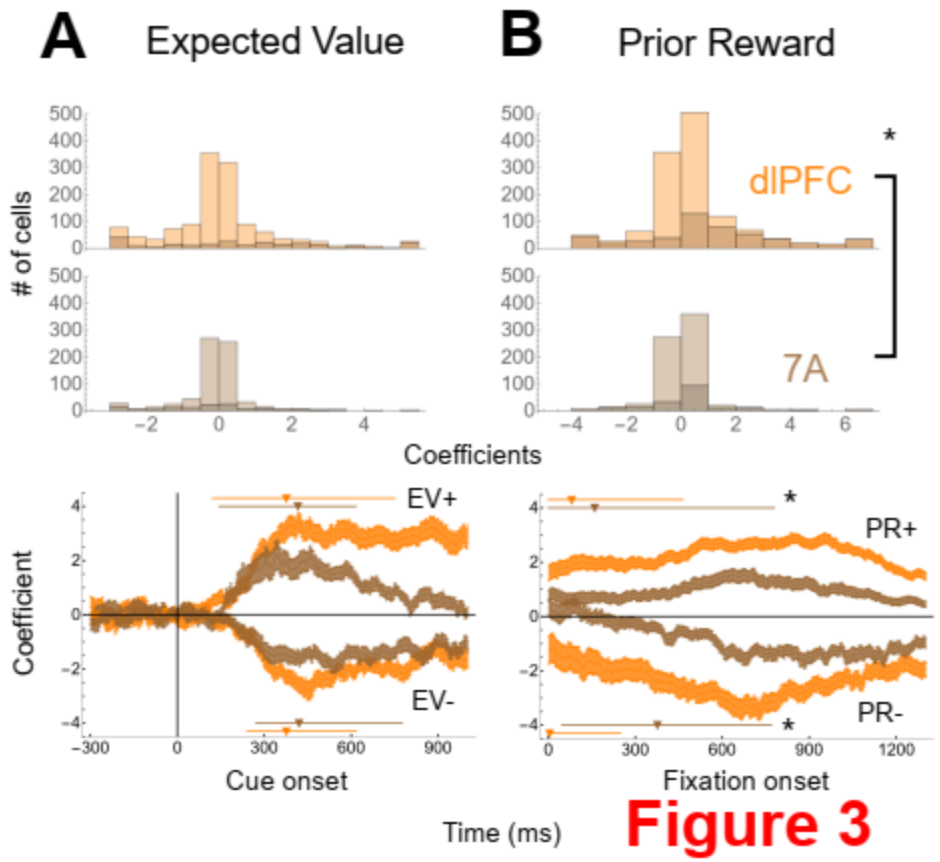

FIGURE 3. DLPFC AND 7A NEURONS ARE SENSITIVE TO EV AND PR (A) EV SELECTIVITY. THE TOP PANEL SHOWS THE DISTRIBUTION OF EV COEFFICIENTS (METHODS, EQ. 4) MEASURED IN THE DELAY PERIOD IN THE DLPFC (ORANGE) AND 7A (BROWN). DARKER SHADING INDICATES CELLS WITH SIGNIFICANT SELECTIVITY AS DEFINED IN *METHODS*. THE BRACKET AND STAR INDICATES $P < 0.05$ FOR A COMPARISON OF DLPFC AND 7A. THE BOTTOM PANEL SHOWS THE TIME COURSE OF EV SELECTIVITY CALCULATED BY APPLYING EQ. 4 IN A SLIDING WINDOW WITH 50 MS BINS AND 2 MS STEPS ALIGNED ON CUE ONSET. THE TRACES SHOW THE MEAN AND SEM COEFFICIENT FOR THE SUBSETS OF SELECTIVE NEURONS WITH POSITIVE AND NEGATIVE SCALING (EV+ AND EV-) IN DLPFC (ORANGE) AND 7A (BROWN). TRIANGLES AND LINES SHOW THE MEDIAN LATENCY AND INTERQUARTILE-INTERVAL FOR THE CORRESPONDING GROUP, AND STARS INDICATE $P < 0.05$ FOR THE LATENCIES IN DLPFC AND 7A. **(B) PR SELECTIVITY:** TOP PANEL SHOWS THE DISTRIBUTION OF PR COEFFICIENTS AND THE BOTTOM PANEL, THE TIME COURSE OF SELECTIVITY IN PR+ AND PR- CELLS IN THE SAME FORMAT AS IN **A**.

The neural responses to PR and EV followed similar patterns as the monkeys' licking response. EV-sensitive cells had a quasi-categorical firing rate pattern that resembled the monkeys' licking response, and was driven primarily by responses to the two highest levels of EV (whether this was an increase or decrease in firing for cells with, respectively, positive and negative modulations; **Fig. 4A**; cf **Fig. 2B**, right). To quantitatively evaluate the correspondence with the monkeys' expectations we conducted a representational similarity analysis, by computing the pairwise discriminability among 35 trial categories defined by 7 levels of EV and 5 levels of PR. We found strong positive correlations between the representations of these categories in the licking response and firing rates of dlPFC and 7A cells (**Fig. 4B;** monkey 1: PFC $r = 0.7$; 7A $r = 0.64$; monkey 2: PFC $r = 0.5$; 7A $r = 0.24$; all $p < 10^{-18}$), showing that the neural representation of PR and EV mirrored the monkeys' reward expectancies.

While EV is a key economic quantity, it is unknown how it is computed based on more primary variables of reward magnitude and probability. To examine this question, we separately estimated the neurons' sensitivity to each factor using trials with probabilistic and deterministic cues (**Fig. 1B**; (*Methods*, eq. 6,7). Despite being evaluated on different trials, neural sensitivity to reward probability and magnitude were highly correlated (**Fig. 4C, Table 2**) and each coefficient was correlated with EV sensitivity (**Table 2**). These correlations are in sharp contrast with the independence between EV and PR sensitivity and other task-related regressors (**Table 2**, see also below) suggesting that the neurons integrate information about probability and magnitude for computing EV.

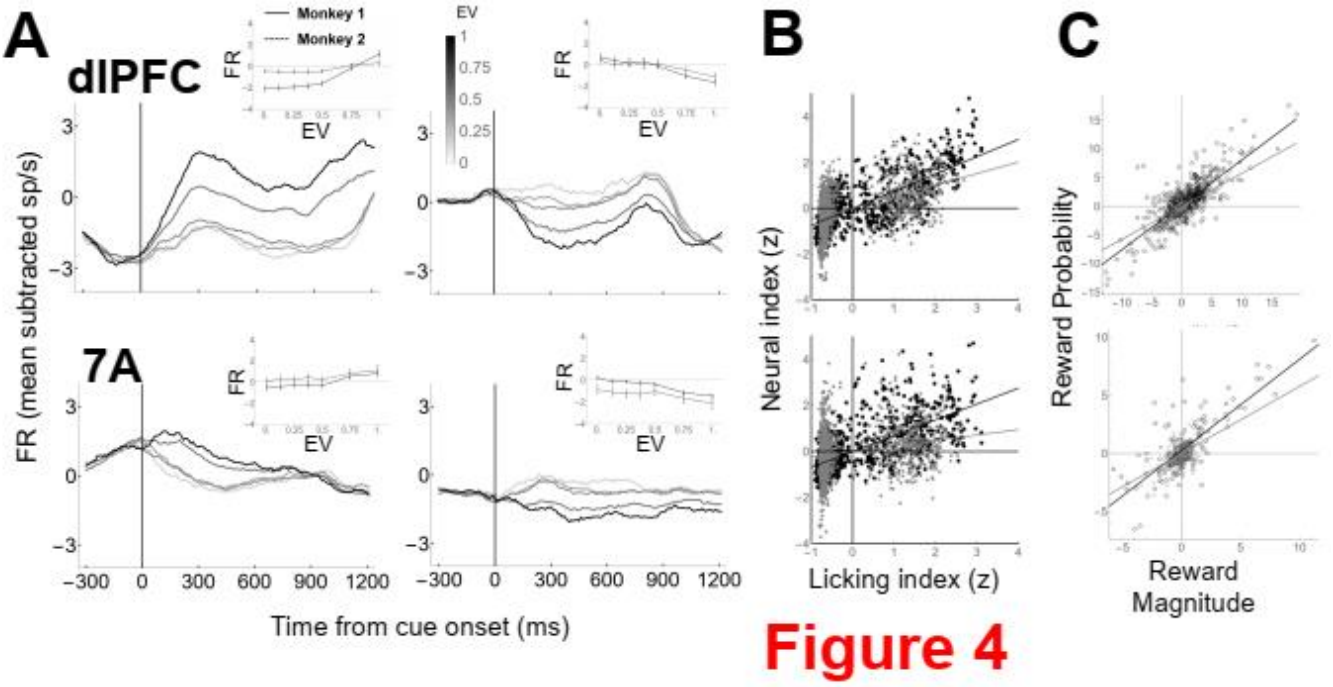

FIGURE 4. PROPERTIES OF THE EV SELECTIVITY (A) PERI-STIMULUS TIME HISTOGRAMS (PSTHS) OF CUE-ALIGNED FIRING RATES FOR EV+ AND EV- CELLS IN EACH AREA SHOW A QUASI-CATEGORICAL PROFILE, WITH THE STRONGEST MODULATIONS FOR THE TWO HIGHEST EV (0.75 AND 1.0). THE HISTOGRAMS ARE CONSTRUCTED BY CONVOLVING RAW FIRING RATES WITH A 100MS BOXCAR MOVING AVERAGE FOR 5 LEVELS OF EV (0, 0.25, 0.5, 0.75, 1). THE INSETS SHOW FIRING RATES (MEAN AND SEM DURING THE DELAY PERIOD) AS A FUNCTION OF EV **(B)** REPRESENTATIONAL SIMILARITY ANALYSIS PLOTTING THE DISCRIMINABILITY (AD INDEX) BETWEEN ALL PAIRS OF REWARD CONTINGENCIES IN THE LR (X AXIS) VERSUS FIRING RATES (Y AXIS). EACH POINT IS A PAIR OF CONTINGENCIES AND THE LINES SHOW THE BEST FIT REGRESSION (BLACK, MONKEY 1, GRAY, MONKEY 2). **(C)** CORRELATIONS BETWEEN REWARD MAGNITUDE COEFFICIENTS ON TRIALS WITH DETERMINISTIC CUES (**FIG. 1A**) AND PROBABILITY COEFFICIENTS ON TRIALS WITH PROBABILISTIC CUES (**FIG. 1A**). EACH POINT IS ONE NEURON AND THE LINES SHOW THE BEST FIT REGRESSION (BLACK, MONKEY 1, GRAY, MONKEY 2).

Encoding of the outcome

Large populations of neurons in both areas showed their strongest responses upon outcome delivery. Outcome related activity consisted of sensitivity to the magnitude of the current reward (CR), which was significant, and sustained throughout the ITI, in 63% of dlPFC cells and 47% of 7A cells (*Methods*, eq. 8; **Fig. 5A, Table 1**). CR sensitivity showed predominantly negative scaling, with more than 60% of the sensitive cells in each area responding most strongly to omission of reward (CR = 0; **Fig. 5A, Table 1**). Although very robust in both areas, CR sensitivity was stronger in dlPFC, as indicated by the higher

prevalence of significant selectivity (63% vs 47%, $p < 10^{-12}$; **Fig. 5A, Table 1**), and, in the sensitive cells, larger coefficients and shorter latencies in the dlPFC relative to 7A (**Fig. 5A, Table 1**)

The fact that CR encoding was sustained throughout the ITI raises the possibility that it carried over into the following trials and explained the neurons' sensitivity to PR. As noted above, the fact that PR sensitivity was often not present at the start of a trial argues against this possibility (**Fig. 4B**). As further evidence that CR and PR selectivity are distinct, the coefficients indexing these effects were uncorrelated (**Table 2**). As further verification we carried out cross-epoch decoding analyses testing whether classifiers trained to decode CR can provide effective read-outs of PR and vice versa (**Fig. 5B**). In both dlPFC and 7A, classifiers trained to decode CR during the last portion of the ITI were significantly less accurate at decoding the same quantity during the fixation period of the following trial (here denoted "PR") than they were during an equivalently distant time window in the early part of the ITI (**Fig. 5B, left**). Similarly, classifiers trained to decode PR during the fixation epoch were significantly less accurate in decoding the same quantity during the preceding ITI (here denoted CR) relative to an equivalently distant time period in the following trial (**Fig. 5B,** right). Therefore, PR sensitivity was not a mere persistence of a CR response but actively emerged in a distinct population of cells at the start of a trial.

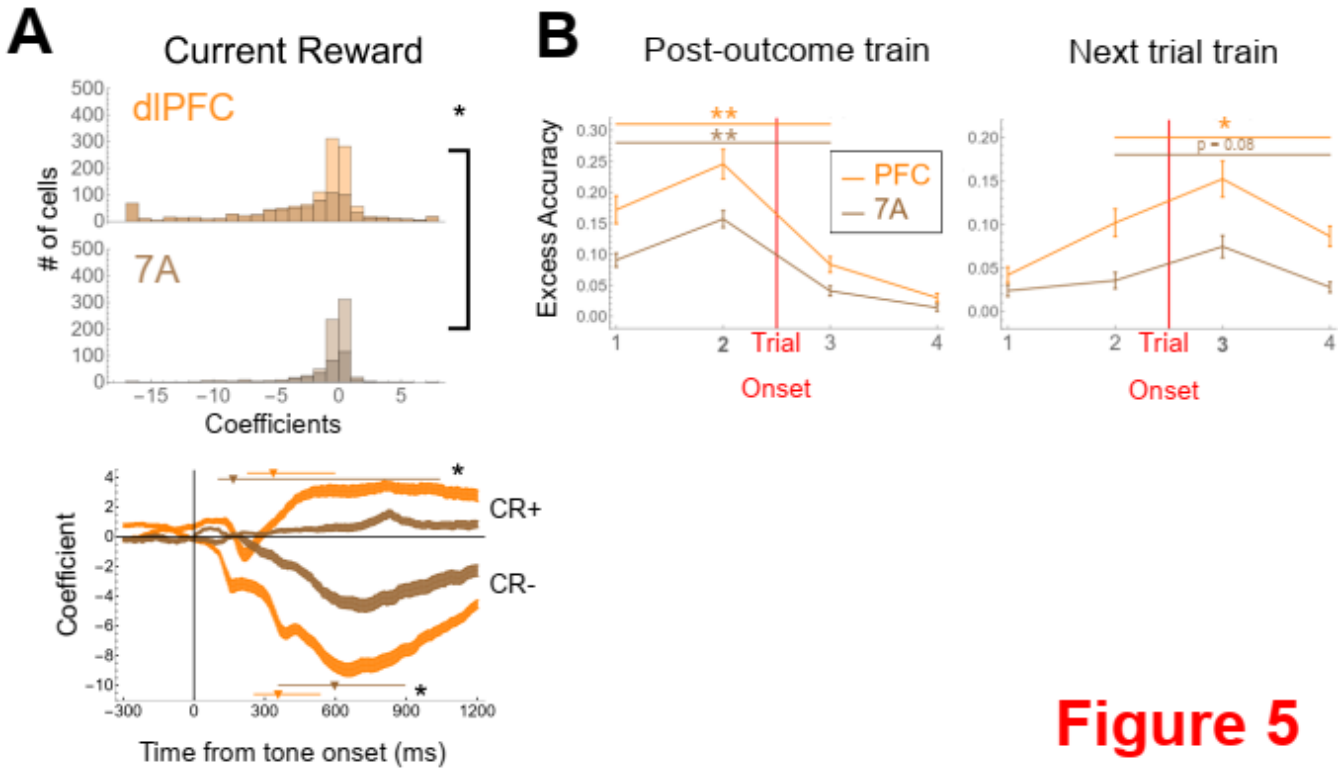

Figure 5

FIGURE 5. DLPFC AND 7A NEURONS ENCODE THE OBTAINED REWARD SIZE (A) THE DISTRIBUTION OF CR SELECTIVITY (TOP) AND THE TIMECOURSE OF SELECTIVITY IN THE SENSITIVE CELLS, ALIGNED ON TONE ONSET. OTHER CONVENTIONS AS IN FIG. 3. **(B) PR IS NOT A MERE PERSISTENCE OF THE CR RESPONSE.** ACROSS-EPOCH CLASSIFICATION ANALYSIS BETWEEN THE ITI AND FIXATION ONSET ON THE FOLLOWING TRIAL. FIRING RATES WERE DIVIDED IN 4 CONSECUTIVE 300 MS EPOCHS CENTERED ON TRIAL ONSET (FIXATION POINT ONSET). LOGISTIC CLASSIFIERS WERE TRAINED TO DECODE CR (OR PR) FROM THE ACTIVITY OF SIMULTANEOUSLY RECORDED CELLS IN EACH AREA USING THE WINDOW IMMEDIATELY PRECEDING TRIAL ONSET (LEFT PANEL, WINDOW 2) OR THE WINDOW IMMEDIATELY FOLLOWING THE TRIAL ONSET (RIGHT PANEL, WINDOW 3). THE TRACES SHOW CLASSIFICATION ACCURACY ABOVE THE SHUFFLE CONTROL (MEAN AND SEM ACROSS SESSIONS) FOR DLPFC (ORANGE) AND 7A (BROWN). FOR WINDOWS THAT WERE EQUALLY DISTANT FROM THE

TRAINING INTERVAL, CLASSIFICATION WAS SIGNIFICANTLY MORE ACCURATE WHEN TESTED WITHIN THE SAME TRIAL RELATIVE TO ACROSS TRIALS IN BOTH AREAS (* $P < 0.05$; ** $P < 0.001$).

Integration of experienced and expected outcomes implicitly signals RPE

In addition to encoding CR during the ITI epoch, many neurons carried information about EV and PR. We estimated sensitivity to EV and PR during the ITI (EV_ITI and PR_ITI) by entering EV and PR as regressors alongside with CR while controlling for cue and target location (*Methods*, eq. 8).

A straightforward possibility is that the post-outcome responses to EV and PR were mere continuations of the neuronal sensitivity to these variables preceding the outcome. Contrary to this hypothesis, coefficients to EV and PR were only weakly correlated with those for EV_ITI and PR_ITI (**Table 2**). Moreover, cross-epoch decoding showed that both variables were distinctively represented before versus after the outcome. In both dlPFC and 7A, classifiers trained to decode PR or EV immediately before outcome delivery were significantly less accurate at decoding the corresponding variable during the ITI, relative to an equally distant time window during the trial (**Fig. 6A, left**). Similarly, classifiers trained to decode PR or EV immediately after outcome delivery were significantly less accurate in decoding the variables before outcome delivery relative to an equally distant time period after the outcome (**Fig. 6A, right**). Thus, dlPFC and 7A cells showed new representations of EV and PR information during the ITI, which were distinct from those used during the trial.

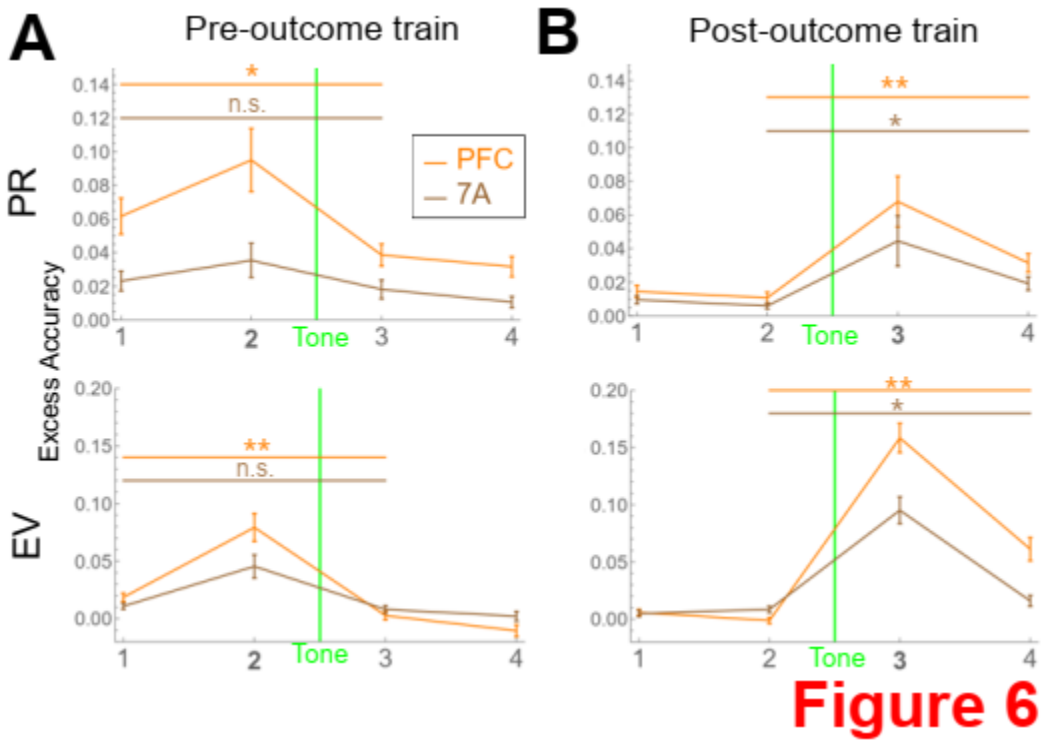

Figure 6. EV and PR encoding during the ITI is not a mere persistence of trial responses. Logistic classification performed as in Fig. 5B, but trained to decode PR (top) and EV (bottom) in 600 ms epochs immediately preceding **(A)** or following **(B)** the tone onset. All conventions as in **Fig. 5B.**

Although the coefficients to PR_ITI, EV_ITI were not correlated with the coefficients for CR (**Table 2**), sensitivity to PR_ITI and EV_ITI differed markedly depending on the trial's outcome, and this dependence was opposite for the two quantities. Moreover, the *polarities* of the PR_ITI and EV_ITI responses were skewed in opposite directions for the two quantities. Coefficients for PR_ITI were predominantly *positive*, indicating that most neurons had enhanced ITI firing for trials following a larger prior reward (**Fig. 7A, B; Table 1**). In addition, PR_ITI sensitivity was much stronger after reward *omission* rather than receipt, for the entire population (**Fig. 7A**) and for the subset of sensitive cells (**Fig. 7B**). These trends were highly robust in both areas; in the dlPFC, PR_ITI coefficients were 4.79±0.29 after reward omission versus 1.56±0.16 after reward receipt (all neurons, $p < 10^{-14}$); in 7A, the corresponding values were 2.56±0.45 vs 0.43±0.2 (all neurons, $p < 10^{-11}$).

The polarity and outcome dependence were opposite for EV_ITI, although EV_ITI modulations were overall weaker. The coefficients for EV_ITI were predominantly *negative*, indicating that most neurons had enhanced firing for trials with lower EV (**Fig. 7C, D**; **Table 1**). Moreover, EV_ITI sensitivity, was stronger (more negative) after reward *receipt* relative to omission in the subset of sensitive cells (**Fig. 7D**) and across the population (**Fig. 7C**). The dependence on outcome was only a trend in area 7A (EV_ITI coefficients of -0.26±0.12 after reward receipt vs -0.23±0.37 after omission, all neurons, $p = 0.74$) but was highly robust in the dlPFC (EV_ITI coefficients of -1.35±0.13 versus 0.67±0.69, all neurons, $p < 10^{-19}$).

Together, the asymmetries produced by the outcome and encoding polarity resulted in an emphasis on outcomes that were surprising relative to, respectively, PR and EV. Firing rates were strongest for reward *omissions* that were surprising relative to a rich reward history (**Fig. 7A, B).** In addition, firing rates were enhanced for *rewards* that were surprising relative to a low cue-signaled EV (**Fig. 7C,D**).

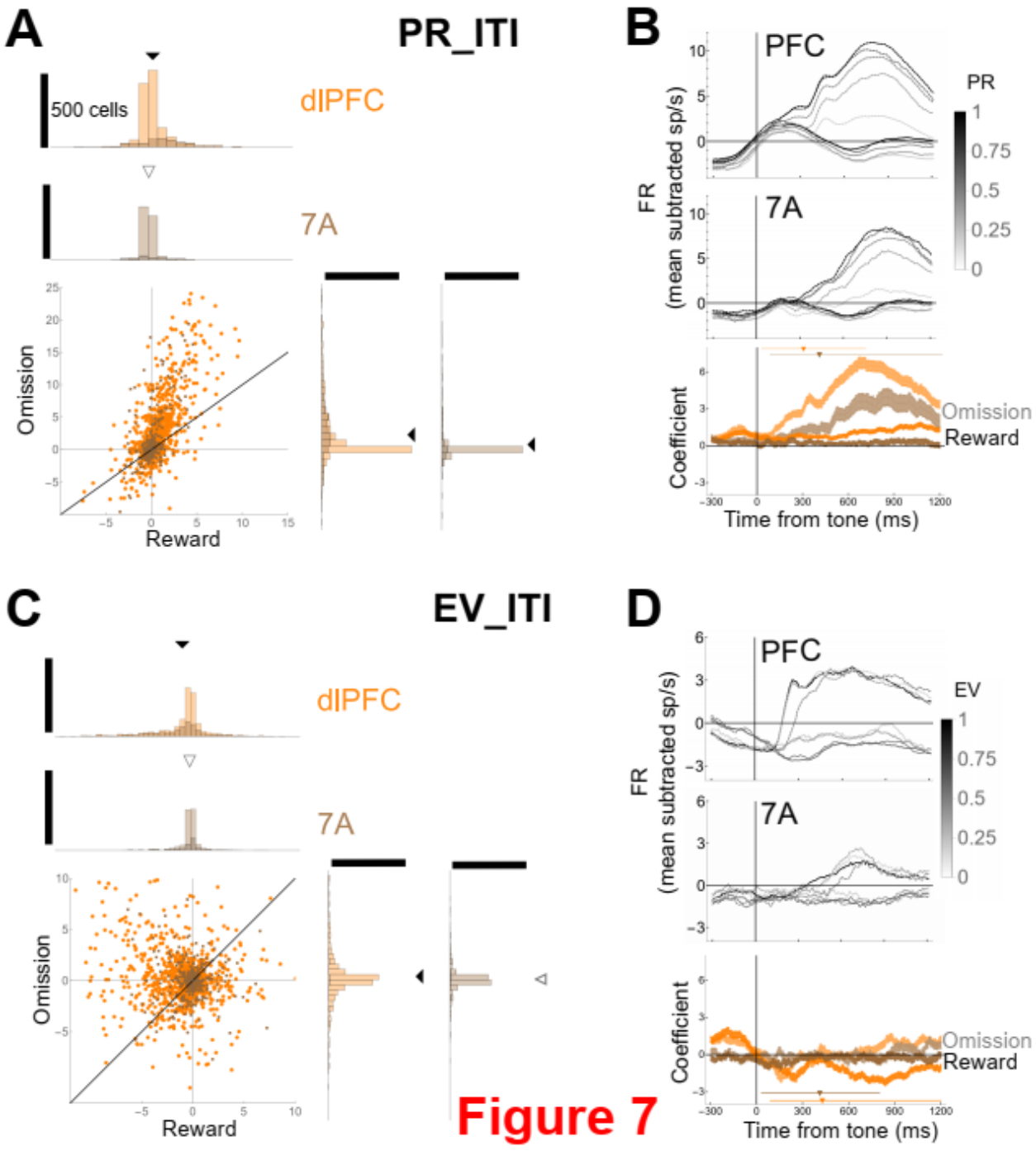

FIGURE 7. EV AND PR ENCODING DURING THE ITI DIFFER IN POLARITY AND INTERACTIONS WITH THE OUTCOME (A) ACROSS THE POPULATION, THE SENSITIVITY TO PR_ITI IS STRONGER AFTER REWARD OMISSION AND PREDOMINANTLY POSITIVE COMPARISON OF REGRESSION COEFFICIENTS FOR PR_ITI (*METHODS*, EQ. 8) ON REWARDED AND UNREWARDED TRIALS, FOR ALL THE RECORDED CELLS. EACH POINT IS ONE CELL (ORANGE: DLPFC; BROWN: 7A). IN THE MARGINAL HISTOGRAMS, DARKER SHADING SHOWS SIGNIFICANT CELLS. THE TRIANGLES SHOW MEANS WITH FILLED SYMBOLS INDICATING $P < 0.05$ RELATIVE TO 0. **(B) IN THE SENSITIVE CELLS, THE EFFECT OF PR_ITI IS STRONGER AFTER REWARD OMISSION AND PREDOMINANTLY POSITIVE** THE TOP TWO PANELS SHOW AVERAGE PSTHS FOR ALL THE CELLS WITH PR_ITI SENSITIVITY, ALIGNED ON TONE ONSET, FOR REWARDED VERSUS UNREWARDED TRIALS AND FOR EACH LEVEL OF PR (GRAYSCALE). THE TRACES SEPARATE BY THE TYPE OF OUTCOME, WITH HIGHER FIRING RATES AFTER REWARD OMISSION RELATIVE TO REWARD RECEIPT, DUE TO THE NEURONS' NEGATIVE SCALING WITH CR. FOR EACH OUTCOME, THE NEURONS

HAVE THE STRONGEST RESPONSES FOR THE HIGHEST PR, AND THIS EFFECT IS STRONGER AFTER OMISSION OF REWARD. THE BOTTOM PANEL SHOWS THE PR_ITI COEFFICIENTS FROM EQ. 8 COMPUTED IN A SLIDING WINDOW OF 50 MS WIDTH AND 2MS STEPS. THE TRACES SHOW THE MEAN AND SEM ACROSS THE SENSITIVE CELLS IN DLPFC (ORANGE) AND 7A (BROWN), SEPARATELY FOR TRIALS ENDING IN REWARD (DARKER SHADE) AND REWARD OMISSION (LIGHTER SHADING). **(C) ACROSS THE POPULATION, THE SENSITIVITY TO EV_ITI IS STRONGER ON REWARDED TRIALS AND PREDOMINANTLY NEGATIVE.** SAME CONVENTIONS AS IN **A**. **(D) IN THE SENSITIVE CELLS, THE EFFECT OF EV_ITI IS STRONGER ON REWARDED TRIALS AND PREDOMINANTLY NEGATIVE.** SAME CONVENTIONS AS IN **B**.

To determine whether these asymmetries reflected *bona fide* encoding of RPEs, we plotted firing rates as a function of RPEs defined in two references frames. We defined PR_RPE as the difference between the magnitude of the current and prior reward, capturing the surprisingness of an outcome relative to reward history. We defined EV_RPE as the difference between the reward magnitude and the trial's EV, capturing the surprisingness of an outcome relative to the cue-predicted EV. We then computed for each cell the absolute firing rate modulation relative to its average activity to capture how much the cell modulates as a function of PR_RPE and EV_RPE.

This analysis showed that individual cells did not have a strong linear encoding of PR_RPE or EV_RPE. For PR_RPE, the population showed an approximate skewed U-shape function (**Fig. 8A**), modulating more strongly at extreme relative to intermediate values of PR_RPE, and more strongly at negative relative to positive values PR_RPE. This non-linear effect, however, was not entirely consistent for rewarded and unrewarded trials (solid vs dashed traces in **Fig. 8A**), was weak for area 7A, and was found only in a minority of individual cells (**Fig. S3;** dlPFC: cluster 3, 24% of the cells; 7A: cluster 5, 14% of the cells). A consistent relation with EV_RPE was even more elusive. The population response showed a weak U-shaped pattern with a trend for stronger modulation for extreme values of EV_RPE (**Fig. 8B**), but this profile was very weak in area 7A and was only found in a small subset of cells in the dlPFC (**Fig. S3**, cluster 6, 11% of cells).

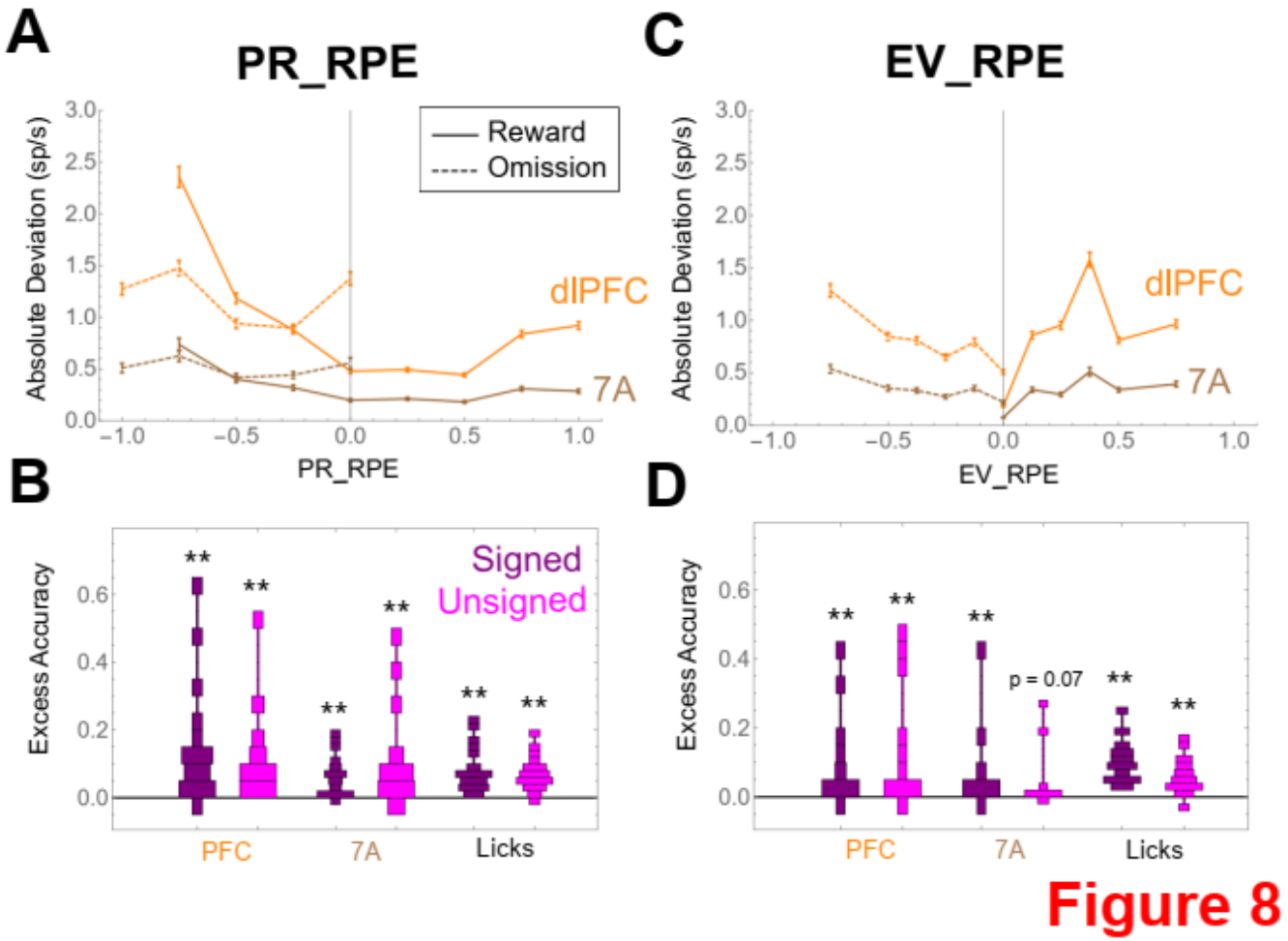

Figure 8

FIGURE 8. PR_RPE AND EV_RPE ARE NOT LINEARLY ENCODED IN INDIVIDUAL CELLS BUT CAN BE DECODED ACROSS THE POPULATION (A) NON-LINEAR RESPONSE TO PR_RPE THE ABSOLUTE DEVIATION FROM THE AVERAGE FIRING RATES AS A FUNCTION OF PR_RPE, AFTER REWARD RECEIPT (SOLID) AND OMISSION (DASHED). THE POINTS SHOW THE MEAN AND SEM FOR ALL THE CELLS IN DLPFC (ORANGE) AND 7A (BROWN). **(B) DECODING OF PR_RPE** DISTRIBUTIONS OF DECODING ACCURACY (AFTER SUBTRACTION OF ACCURACY IN SHUFFLED TRIALS) ACROSS THE RECORDING SESSIONS, FOR DLPFC, 7A, AND LR, AND FOR SIGNED AND UNSIGNED PR_RPE. (* $P < 0.05$; ** $P < 0.001$) **(C) NON-LINEAR RESPONSES TO EV_RPE** SAME CONVENTIONS AS IN **A**. **(D) DECODING OF EV_RPE** SAME CONVENTIONS AS IN **B**.

Therefore, the combination of asymmetric signals about the experienced and expected outcome did not amount to a formal encoding of RPE in individual cells. Nevertheless, this combination allowed decoding of PR_RPE and EV_RPE from the population response. Simple logistic decoders read out both signed and unsigned PR_RPE and EV_RPE with above-chance accuracy from the population response in both areas (**Fig. 8C, D**). Decoding accuracy was equivalent for signed and unsigned RPE, but was higher in dlPFC relative to 7A and was higher for PR_RPE relative to EV_RPE (3-way ANOVA across sessions, $p = 0.9$ for signed/unsigned; $p = 0.005$ for area; $p = 0.0003$ for PR_RPE vs EV_RPE). Thus, RPE was conveyed more reliably by the dlPFC relative to the 7A population and, remarkably, was conveyed more reliably if referenced to the statistically irrelevant reward history than to the task-relevant cue. Finally, PR_RPE and EV_RPE could be decoded from the monkeys' licking responses in both their signed and unsigned forms, showing that these quantities were behaviorally significant.

Discussion

Spearheaded by the early investigations of Mountcastle and his colleagues[23], recent studies of the inferior parietal lobe focused heavily on LIP, a small area on the lateral bank of the intraparietal sulcus that is associated with spatial orienting through eye movements and visual attention. Fewer studies, in contrast, targeted area 7A that occupies a much larger swath of cortex on the adjacent gyral surface. These studies showed that 7A neurons selectively respond to salient targets of visual search similar to dlPFC cells[24] and transmit information about task context and rules that at least partially reflect top-down signals from the dlPFC[7, 8]. Together with the strong reciprocal connections between 7A and the dlPFC[9], these findings suggest that this area is an important node in the fronto-parietal network mediating executive control and attention.

We extend these results by showing that 7A neurons also show sensitivity to expected and experienced rewards like that found in the dlPFC. Using a task that allowed us to focus on reward expectations independently of learning or decision strategies, we show that neurons in both areas carry signals encoding expected and experienced outcomes and combine these signals to allow decoding of RPEs in two reference frames - relative to the cues and relative to statistically irrelevant reward history. We discuss each finding in turn.

Expected value. In the task that we used, monkeys viewed familiar, extensively trained visual cues that signaled the reward distribution that was in effect on each trial. Stimuli with familiar reward associations gain salience and the ability to automatically bias attention[25]. We previously showed that, in area LIP, such stimuli evoke enhanced short-latency visual responses independently of the monkeys' reward expectations, suggesting that one mechanism of reward-based salience involves visual plasticity[26]. Such plasticity, however, did not account for the EV encoding in 7A and dlPFC. Instead, neurons in these areas had sustained encoding of EV that emerged at long latencies of nearly 300 ms and had a quasi-categorical response pattern that corresponded to the monkeys' expectations, suggesting that they signaled true predictive activity rather than salience or visual plasticity.

The EV-sensitive cells were equally likely to show positive or negative scaling, consistent with previous findings in the ACC and OFC in tasks involving reward-based decisions[17]. Moreover, we show that dlPFC and 7A cells had *bona fide* sensitivity to reward magnitude and probability, similar with previous findings during decision making in the ventromedial prefrontal cortex[27]. Thus, the properties of EV coding may be shared across frontal and parietal areas independent of active decision strategies.

Experienced reward. Upon receipt of an outcome, a large fraction of 7A and dlPFC cells had robust responses encoding reward size, with most cells showing negative scaling – enhanced firing to smaller relative to larger rewards. Strong reward-related responses with negative scaling are widespread throughout the PFC (e.g., [28-30]). Our finding that outcome information had shorter latencies in the dlPFC relative to 7A suggests that these responses are conveyed in a top-down fashion from the frontal to the parietal lobe, consistent with the proposed evolutionary role of the PFC in detecting (and eventually minimizing) foraging errors[31].

Reward memory. A perhaps more striking result was that, in addition to their sensitivity to actual rewards and task-relevant cues, the neurons conveyed memories of recent rewards that were statistically unrelated to a trial's EV. PR-sensitive cells in both areas encoded the size of the prior reward in ways that were not explained by mere persistence of activity and reflected the monkeys' behavioral sensitivity to reward history.

The ability to integrate information on longer time scales is a consistent feature of frontal and parietal areas, and these areas have been reported to convey across-trial information about context, rules, the presence of conflict as well as rewards (reviewed in [21]). In most previous studies, however, the extended memories were adaptive for performing a task, correlating with improved rule-based performance or adaptation to conflict[21], decision making[32, 33] and learning[19, 20, 34]. An additional study reported that dlPFC neurons respond to *irrelevant* reward history in a complex rule-based task, but found no manifestations of these responses in the monkeys' strategy[35]. Our findings extend the literature by showing that encoding of reward history is correlated with a strong behavioral bias that distorted the monkeys' expectations relative to the optimal inference based on the cue-signaled EV.

Our results thus provide a likely neural correlate for the "hot-hand" fallacy, a well-known economic bias whereby decision makers (humans and monkeys) choose as if they expected positive or negative outcomes to happen in streaks independently of their true autocorrelations[22]. Our findings suggest that biases due to irrelevant recent rewards may rely primarily on the frontal cortex, since PR sensitivity emerged earlier and was stronger in the dlPFC relative to 7A and may be conveyed from anterior to posterior areas (see also[17]). Thus, the specific contributions of frontal and parietal areas to effects of irrelevant reward-history in different task contexts (along with effects of irrelevant recent stimuli or actions[36]) are important topics for future investigations.

Implicit encoding of RPEs in different reference frames. A final striking aspect of our results is that, in addition to influencing reward expectation, sensitivity to PR was reinstated during the ITI, leading to an integration of information about consecutive outcomes. The neurons' simultaneous sensitivities to the current and previous outcome, although carried by independent populations of cells, produced enhanced firing for reward omissions that were surprising relative to large prior rewards. Interestingly, individual neurons did not fulfill the criteria of encoding RPEs, since they did not scale monotonically with PR_RPE or show anti-correlated sensitivity during expectation and outcome delivery (as reported in the ACC[17]). Nevertheless, reliable inferences about signed and unsigned PR_RPE could be obtained from the population. An analogous dynamic was found for EV information, which emerged in the ITI and resulted in higher responses for rewards that were surprising relative to the trial's EV, allowing decoding of EV_RPE despite a lack of explicit encoding of this quantity by individual cells. The ITI signals of PR and EV were carried by distinct populations of cells relative to those encoding these quantities during the trial, suggesting that they reflected active computation rather than mere persistence of the trial's activity. Consistent with this, PR_RPE and EV_RPE both influenced the monkeys' licking responses.

Prediction errors provide rich information about a task context and, when defined in different reference frames, can serve different behavioral goals[37]. RPEs calculated relative to task-relevant cues may serve to update state-specific reward expectations or monitor the validity of the informative cues[38]. RPEs defined relative to reward history, in contrast, may serve to monitor the overall reward rate for longer-range foraging decisions – e.g., whether to persist with a task or forage for alternative situations[31].

From a computational perspective, the EV-referenced RPEs we describe were based on the true transition probabilities given a task state and are thus similar to model-based RPEs. In contrast, the PR-referenced RPEs were based on state-independent reward history, consistent with model-free mechanisms. A conclusive mapping of our results on model-based and model-free algorithms will require further studies that use computational modeling or specialized tasks such as reinforcer devaluation[14, 39]. Nevertheless, our findings that these quantities are simultaneously represented in frontal and parietal cells is consistent with the fact that model-free and model-based mechanisms jointly

influence behavior[40] and, during observational learning, are both encoded in the human intraparietal sulcus[41]. Thus, our results provide a concrete neural mechanism that may underlie the computation of model-based and model-free RPEs, which can be evaluated against recently proposed alternative algorithms[14].

Methods

General methods. Data were collected from two adult male rhesus monkeys (Macaca mulatta; 9-12kg) using standard behavioral and neurophysiological techniques as described previously[42]. All methods were approved by the Animal Care and Use Committees of Columbia University and New York State Psychiatric Institute as complying with the guidelines within the Public Health Service Guide for the Care and Use of Laboratory Animals. Visual stimuli were presented on a MS3400V XGA high definition monitor (CTX International, INC., City of Industry, CA; 62.5by 46.5 cm viewing area). Eye position was recorded using an eye tracking system (Arrington Research, Scottsdale, AZ). Licking was recorded at 1 kHz using an in-house device that transmitted a laser beam between the tip of the juice tube and the monkey's snout and generated a 5V pulse upon detecting interruptions of the beam when the monkey extended his tongue to obtain water.

Task. A trial started with the presentation of two square placeholders (1˚ width) located along the horizontal meridian at 8º eccentricity to the right and left of a central fixation point (white square, 0.2˚ diameter). After the monkey looked maintained gaze on the fixation point for 300-500 ms (fixation window, 1.5-2˚ square) a randomly selected placeholders was replaced for 300 ms by a reward cue – a checkerboard pattern indicating the trial's reward contingencies (see **Fig. S1A** for detailed description of the visual appearance of the cues). After a 600ms delay period, the fixation point disappeared simultaneously with an increase in luminance of one of the placeholders (the target), whose location was randomized independently from that of the cue. If the monkey made a saccade to the target with a reaction time (RT) of 100 ms – 700 ms and maintained fixation within a 2.0-3.5˚ window for 350 ms, he received a reward with the magnitude and probability that had been indicated by the cue. An auditory tone (200 ms, 500 Hz) signaled the end of the post-saccadic hold period on all trials, providing a temporal marker for the onset of the outcome/ITI period whether a reward was received or omitted. Rewards, when delivered, were linearly scaled between 0.28 to 1.12mL. The ITI – from tone onset to the onset of the fixation point on the following trial lasted for 1200-1600 ms. Error trials (resulting from fixation breaks, premature, late or wrong-direction saccades) were immediately repeated until correctly completed, precluding the monkeys from aborting trials in which they anticipated lower rewards. Monkeys were extensively familiarized with the task and all the cues before recordings began.

Neural recordings. After completing behavioral training, each monkey was implanted with two 48-electrode Utah arrays (electrode length 1.5 mm) arranged in rectangular grids (1 mm spacing; monkey 1, 7x7 mm, monkey 2, 5x10 mm) and positioned in the pre-arcuate portion of the dlPFC and the posterior portion of area 7A (**Figure S2**). Data were recorded using the Cereplex System (Blackrock, Salt Lake City, Utah) over 22 sessions spanning 4 months after array implantation in monkey 1, and 12 sessions spanning 2 months after implantation in monkey 2.

Data analysis. Error trials were discarded and not considered further (13.7% in monkey 1, 14.3% in monkey 2). All statistical analyses were preceded by tests of normality and symmetry ($p < 0.05$). For univariate comparisons we used the Wilcoxon-signed-rank test if the symmetry criterion was met, and the Mann-Whitney U-test otherwise. Correlation coefficients were computed using the Spearman Rank test.

Behavior. Eye position was digitized at 220 Hz, and saccade RT was defined using velocity and acceleration criteria[43]. While RT showed some effects of reward contingencies and spatial congruence between cue and target locations, these effects were not consistent and are not reported here.

Trial-by-trial licking rates (LR) were defined as the proportion of time spent licking in a time window of interest. To estimate the effects of PR we focused the analysis on pairs of consecutive correct trials. We used a 3-step hierarchical analysis (eq. 1-3) to separately estimate the influence of

PR and EV on LR. In the first step we partitioned out the effect of the prior trial's outcome *type* (reward receipt of omission):

$$LR = \beta_0 + \beta_1 * PRNR + \varepsilon \quad (1)$$

where RNR is an indicator equal to 1 if the prior trial was reward and 0 otherwise. In the second step used the residuals from eq. 1 to estimate the effect of PR:

$$LR_{PRNR} = \beta_0 + \beta_1 * PR + \varepsilon \quad (2)$$

where LR_{PRNR} are the residuals from eq. 1, and PR is the magnitude of the prior reward (0, 0.25, 0.5, 0.75, 1). Thus, the PR coefficients we report estimate the monkeys' sensitivity to the size of the prior reward above and beyond the mere presence of a reward. Finally, in the third step we used the residuals from eq. 2 to estimate the effect of EV:

$$LR_{PR} = \beta_0 + \beta_1 * EV + \varepsilon \quad (3)$$

Where LR_{PR} are the residuals from equation 2, and EV is the expected value of the current cue (0, 0.125, 0.25, 0.375, 0.5, 0.75, 1).

To assess the temporal window over which reward history exerted effects, licking was regressed against the value of the prior reward (0, 0.25, 0.5, 0.75, 1) for the 5 previous trials, omitting error trials:

$$LR = \beta_0 + \beta_1 * PR_{n-1} + \beta_2 * PR_{n-2} + \beta_3 * PR_{n-3} + \beta_4 * PR_{n-4} + \beta_5 * PR_{n-5} + \varepsilon \quad (4)$$

Neural responses. Raw spikes were sorted offline using WaveSorter[44] and analyzed with MatLab (MathWorks, Natick, MA) and Mathematica (Champaign, IL). Only neurons with waveforms clearly separated from noise were included in the analysis. All neural analyses were computed on unsmoothed firing rates (FR) that had been normalized within each cell by subtracting the FR averaged across the entire epoch from fixation onset until the end of the ITI.

We used regression analyses to measure the sensitivity to EV, PR and CR. All regressors ranged between 0 and 1, and took values of [0, 0.125, 0.25, 0.375, 0.5, 0.75, 1] for EV, and [0, 0.25, 0.5, 0.75, 1] for PR and CR. Coefficients are reported in units of sp s^{-1}, and sensitive cells are defined as those showing a coefficient with p-value < 0.05.

We included cue location (CL) and target location (TL) as nuisance regressors (coded as 0 or 1 for the hemifield that was, respectively, ipsilateral or contralateral to the recording site), ensuring that we estimate sensitivity to reward variables independently of spatial coding or reward x space interactions.

To estimate the effects of PR and EV (**Fig. 3** and **Table 1**) we fit FR using the equation:

$$FR = \beta_0 + \beta_1 * PR + \beta_2 * EV + \beta_3 * CL + \beta_4 * (PR \cdot CL) + \beta_5 * (EV \cdot CL) + \varepsilon \quad (5)$$

We defined a cell as being PR-sensitive if it showed a significant β_1 coefficient in the interval 0 – 1,000 ms after fixation point onset, and EV-sensitive if it showed a significant β_2 coefficient in the delay period (300 - 900 ms after cue onset).

To estimate the effects of reward probability (P) we fit firing rates on trials with probabilistic cues using the equation:

$$FR = \beta_0 + \beta_1 * PR + \beta_2 * P + \beta_3 * CL + \beta_4 * (PR \cdot CL) + \beta_5 + \varepsilon \quad (6)$$

To estimate the effects of reward magnitude (RM) we fit firing rates on trials using deterministic cues using the equation:

$$FR = \beta_0 + \beta_1 * PR + \beta_2 * RM + \beta_3 * CL + \beta_4 * (PR \cdot CL) + \beta_5 + \varepsilon \quad (7)$$

We defined a cell as being sensitive to probability or magnitude if it had a significant coefficient in the delay period (same interval as that used to measure sensitivity to EV).

To estimate the effects of CR, EV_ITI and PR_ITI we fit FR using the equation:

$$FR = \beta_0 + \beta_1 * PR + \beta_2 * EV + \beta_3 * CR + \beta_4 * CL + \beta_5 * TL + \beta_6 * (CR \cdot CL) + \beta_7 * (CR \cdot TL) + \varepsilon \quad (8)$$

To estimate sensitivity to CR we applied this equation to all trials and defined a cell as being CR-sensitive if it showed a significant β_3 coefficient in the interval 200 – 1,200 ms after tone onset.

To measure the sensitivity to EV_ITI and PR_ITI while accounting for the asymmetry of these modulations, we re-applied eq. 8 separately to trials ending in reward and reward omission. (For the latter trials, the CR term was dropped as it was always equal to 0). We defined a cell as being "sensitive" based on the trials with the strongest modulations (i.e., if it had a significant β_2 coefficient for EV_ITI on rewarded trials, or a significant β_1 coefficient for PR_ITI on unrewarded trials).

To estimate effect latencies, we focused on the subset of cells that were sensitive for each factor and analyzed them with reduced models that included only an intercept term and the regressor of interest and was applied in a 50 ms window stepped by 2 ms after the corresponding trigger point (fixation point onset for PR, cue onset for EV and tone onset for CR, EV_ITI and PR_ITI). We identified the earliest pair of consecutive bins showing $p < 0.01$ for the respective coefficient and defined the latency as the start of the first of these bins.

Because linear models of PR_RPE and EV_RPE that included all the necessary covariates (i.e. EV, PR and CR) produced inconsistent results, we examined the encoding of these quantities using cluster analysis and population decoding as described in the text and below.

Representational similarity analysis. To examine the representation of the reward contingencies in the licking response, we calculated the LR on each trial in the 900 ms interval starting at cue onset, pooled the trials across all sessions within each monkey, and partitioned the pooled dataset into 35 bins defined by distinct combinations of 7 levels of EV and 5 levels of PR. We then computed the Anderson-Darling statistic (AD) as a measure of distance between the LR distributions in each possible pair of conditions (1,225 pairs; including those with identical contingencies) We repeated this procedure for each recorded cell using FR in the 900 ms interval starting at cue onset and pooling trials across all neurons that contributed at least 2 trials within a bin. We then calculated the correlation coefficient, across the 1,225 pairs, between the AD distances in LR and FR, for each area and each monkey.

Classification Analyses used the logistic classifier in the Classify[] function of Mathematica with 80/20 cross validation and 100 random replications for each variable (EV, PR, CR, EV_RPE, PR_RPE). We measured accuracy as the fraction of test trials that were assigned to the correct category. We

obtained the baseline level of accuracy given the label distribution for each replication by computing 10 additional classifications on the same trials but with shuffled labels. We defined the *excess accuracy* as the difference between the average classification accuracy on the real and shuffled data. Excess accuracy was estimated for each session and statistics were conducted across sessions. For neuronal responses, the input to the classification was the trial-by-trial FR of all the neurons that were simultaneously recorded in each area in that session. For licking classification analyses, the predictor was the trial-by trial LR over the time window of [-300 300] from cue onset for PR_RPE, and [0 900] from cue onset for EV_RPE. The target classes were EV_RPE or PR_RPE (signed and unsigned separately) trained using all trials.

	Fraction significant			Regression coefficients			Latencies		
	dlPFC	7A	dlPFC vs 7A	dlPFC	7A	dlPFC vs 7A	dlPFC	7A	dlPFC vs 7A
Pre-Reward									
EV (expected value)									
EV+	0.11 (142)	0.08 (60)	0.02	2.44±0.19*	1.39±0.22*	0.00001	371.2±18.3	362.8±27.8	0.62
EV-	0.07 (95)	0.1 (72)	0.02	-2.33±0.17*	-1.5±0.16*	0.001	376.9±21.1	415.2±23.4	0.22
All	**0.18 (237)**	**0.18 (132)**	**0.42**	**0.52±0.2***	**-0.19±0.18**	**0.004**	**373.4±13.8**	**400.7±17.9**	**0.18**
PR (prior reward)									
PR+	0.39 (378)	0.25 (189)	0.048	2.58±0.16*	0.89±0.11*	10^{-21}	154.4±12.3	208.9±20.8	0.013
PR-	0.19 (250)	0.09 (70)	10^{-9}	-2.63±0.16*	-1.04±0.12*	10^{-8}	104.8±12.4	274.1±42.1	10^{-8}
All	**0.48 (628)**	**0.35 (259)**	10^{-9}	**0.51±0.15***	**0.37±0.1***	**0.63**	**134.8±8.9**	**227.7±19.2**	10^{-10}
P (probability)									
P+	0.1 (125)	0.09 (70)	0.46	3.2±0.28*	1.4±0.18*	10^{-7}	380.1±19.1	409.6±25.2	0.25
P-	0.07 (92)	0.08 (59)	0.22	-2.5±0.18*	-1.8±0.18*	0.007	344.3±24.2	429.5±29.7	0.02
All	**0.16 (217)**	**0.17 (129)**	**0.32**	**0.78±0.26**	**-0.09±0.19**	**0.054**	**364.6±15.1**	**419.7±19.4**	**0.02**
M (magnitude)									
M+	0.09 (119)	0.08 (61)	0.25	3.19±0.28*	1.4±0.18*	10^{-6}	380.8±22.5	448.4±36.5	0.13
M-	0.05 (68)	0.07 (57)	0.01	-2.2±0.18*	-1.7±0.18*	0.35	445.0±31.6	430.5±28.4	0.95
All	**0.14 (187)**	**0.15 (118)**	**0.16**	**1.21±0.27***	**-0.1±0.19**	**0.001**	**407.5±18.7**	**438.6±22.6**	**0.22**
Post-reward/ITI									
CR (Current reward)									
CR+	0.16 (204)	0.18 (136)	0.054	2.35±0.21*	0.7±0.14*	10^{-13}	246.8±14.4	274.6±24.5	0.07
CR-	0.47 (614	0.28 (211)	10^{-17}	-6.91±0.31*	-3.27±0.14*	10^{-14}	230.7±6.6	346.6±14.5	10^{-15}
All	**0.63 (818)**	**0.47 (347)**	10^{-12}	**-4.6±0.27***	**-1.71±0.28***	10^{-12}	**234.7±6.1**	**319.4±13.1**	10^{-6}
PR_ITI (Prior reward)									
R trials									
PR_ITI+	0.18 (234)	0.06 (46)	10^{-14}	2.70±0.13*	1.86±0.2*	0.17	284.3±19.2	352.0±49.5	0.15
PR_ITI-	0.06 (76)	0.06 (44)	0.45	-1.91±0.19*	-1.05±0.14*	0.003	200.3±29.6	289.7±62.1	0.012
All	**0.24 (310)**	**0.12 (90)**	10^{-10}	**1.56±0.16***	**0.43±0.20**	10^{-5}	**226.1±16.5**	**369.5±38.9**	**0.012**
NR trials									
PR_ITI+	0.33 (428)	0.14 (103)	10^{-21}	6.19±0.29*	5.0±0.54*	0.005	232.3±12.6	271.9±31.2	0.14
PR_ITI-	0.05 (75)	0.08 (57)	0.04	-3.24±0.27*	-1.89±0.29*	10^{-5}	186.65±30.9	368.1±45.1	10^{-5}
All	**0.38 (503)**	**0.21 (160)**	10^{-15}	**4.79±0.29***	**2.56±0.45***	10^{-8}	**225.4±11.7**	**305.8±26.0**	**0.0005**
EV_ITI (Expected value)									
R trials									
EV_ITI+	0.15 (202)	0.15 (113)	0.45	1.59±0.13*	0.68±0.13*	10^{-9}	324.82±26.4	348.7±42.0	0.46
EV_ITI-	0.33 (431)	0.13 (94)	10^{-24}	-2.73±0.13*	-1.41±-0.16*	10^{-6}	321.9±17.9	418.3±41.8	0.016
All	**0.49 (633)**	**0.38 (207)**	10^{-10}	**-1.35±0.13***	**-0.26±0.12**	10^{-8}	**322.8±14.8**	**382.6±29.7**	**0.017**
NR trials									
EV_ITI+	0.02 (30)	0.02 (19)	0.35	3.84±0.67*	1.0±0.29*	0.002	274.0±51.1	516.9±121.9	0.078
EV_ITI-	0.02 (23)	0.02 (15)	0.33	-3.45±0.68*	-1.79±-0.52*	0.17	268.0±66.7	381.3±114.7	0.3
All	**0.04 (53)**	**0.04 (34)**	**0.28**	**0.67±0.69***	**-0.23±0.37**	**0.33**	**271.5±40.2**	**449.1.6±82.9**	**0.059**

Table 1. **Neural modulations and area comparisons.** The major vertical sections show, from left to right, the proportion of cells with significant effects (fraction (number)), the regression coefficient across the significant cells (mean±sem) and the latencies for significant cells (mean±sem). Each section gives the data for dlPFC and 7A, and the p-value for a statistical comparison across the two areas (left column, z-test of proportions; center and right column, two-sample tests as indicated in the text). The major horizontal sections show the signals discussed in the text, for positive encoding cells (+), negative encoding cells (-), and all the significant cells. EV, P and M effects are based on analyses in the 300 – 900 ms after cue onset; PR selectivity is based on 0 – 1,000 ms after fixation onset; and all the post-reward (ITI) effects were measured between 200 and 1,200 ms after tone onset. The analyses are based on the full data set (1,298 neurons in the dlPFC and 736 in 7A).

dlPFC							7A					
	EV	P	M	CR	PR_ITI	EV_ITI	EV	P	M	CR	PR_ITI	EV_ITI
PR	0.06*	0.06*	0.04	-0.066*	0.1**	0.03	-0.05	-0.08*	0.02	0.007	0.25***	0.05
EV		**0.5*** **	**0.69*** **	-0.078*	0.08*	-0.01		**0.47*** **	**0.68*** **	-0.06	-0.03	0.04
P			**0.85*** **	0.02	0.11*	-0.02			**0.45*** **	-0.01	-0.007	-0.06
M				0.037	0.08*	-0.0003				0.046	-0.02	0.05
CR					-0.04	- 0.01					-0.09*	0.11*
PR_ITI						**-0.39*** **						-0.13*

Table 2. **Correlations between regression variables.** Each entry shows the Spearman correlation between the regression coefficients of a pair of variables (measured as in Table 1) across all the neurons in each area. White/gray shading indicates signals estimated during, respectively, the pre-reward and post-reward epochs. * p < 0.05; **p < 0.01; ***p < 0.0001**

References

1. Gershman, S.J., Horvitz, E.J. & Tenenbaum, J.B. Computational rationality: A converging paradigm for intelligence in brains, minds and machines. *Science* **349**, 273-278 (2015).
2. Caplin, A. & Dean, M. Revealed Preference, Rational Inattention and Costly Information Acquisition *American Economic Review* **105**, 2183-2203 (2015).
3. Wallis, J.D. & Kennerley, S.W. Heterogeneous reward signals in prefrontal cortex. *Current opinion in neurobiology* **20**, 191-198 (2010).
4. Shenhav, A., Botvinick, M. & Cohen, J. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217-240 (2013).
5. Katsuki, F. & Constantinidis, C. Unique and shared roles of the posterior parietal and dorsolateral prefrontal cortex in cognitive functions. *Front Integr Neurosci* **6**, 17 (2012).
6. Barash, S., Bracewell, R.M., Fogassi, L., Gnadt, J.W. & Andersen, R.A. Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. *Journal of Neurophysiology* **66**, 1095-1108 (1991).
7. Stoet, G. & Snyder, L.H. Single neurons in posterior parietal cortex of monkeys encode cognitive set. *Neuron* **42**, 1003-1012 (2004).
8. Crowe, D.A. *et al.* Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat Neurosci* **16**, 1484-1491 (2013).
9. Cavada, C. & Goldman-Rakic, P.S. Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *J. Comp. Neurol.* **287**, 422-445 (1989).
10. Saleem, K.S., Miller, B. & Price, J.L. Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. *J Comp Neurol.* **522**, 1641-1690 (2014).
11. Fan, J. An information theory account of cognitive control. *Front Hum Neurosci.* **8** (2014).
12. Pearce, J.M. & Mackintosh, N.J. *Two theories of attention: a review and a possible integration*. (Oxford University Press, New York; 2010).
13. Iigaya, K., Story, G.W., Kurth-Nelson, Z., Dolan, R.J. & Dayan, P. The modulation of savouring by prediction error and its effects on choice. *eLife* **April 21**, pii: e13747 (2016).
14. Russek, E.M., Momennejad, I., Botvinick, M.M., Gershman, S.J. & Daw, N.D. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput Biol* **13**, e1005768 (2017).
15. Schultz, W. Neuronal Reward and Decision Signals: From Theories to Data. *Physiol Rev.* **95**, 853-951 (2015).
16. Hayden, B.Y., Heilbronner, S.R., Pearson, J.M. & Platt, M.L. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *Journal of Neuroscience* **31**, 4178-4187 (2011).
17. Kennerley, S.W., Behrens, T.E. & Wallis, J.D. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci* **14**, 1581-1589 (2011).
18. Kennerley, S.W. & Wallis, J.D. Reward-dependent modulation of working memory in lateral prefrontal cortex. *J Neurosci* **29**, 3259-3270 (2009).
19. Asaad, W.F., Lauro, P.N., Perge, J.A. & Eskandar, E.N. Prefrontal Neurons Encode a Solution to the Credit-Assignment Problem. *J Neurosci.* **37**, 6995-7007 (2017).
20. Asaad, W.F. & Eskandar, E.N. Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. *J Neurosci.* **31**, 17772-17787 (2011).
21. Mansouri, F.A., Egner, T. & Buckley, M.J. Monitoring Demands for Executive Control: Shared Functions between Human and Nonhuman Primates. *Trends Neurosci* **40**, 15-27 (2017).

22. Blanchard, T.C., Wilke, A. & Hayden, B.Y. Hot-hand bias in rhesus monkeys. *J Exp Psychol Anim Learn Cogn* **40**, 280-286 (2014).
23. Mountcastle, V.B., Lynch, J.C., Georgopoulos, A., Sakata, H. & Acuna, C. Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *Journal of Neurophysiology* **38**, 871-908 (1975).
24. Katsuki, F. & Constantinidis, C. Early involvement of prefrontal cortex in visual bottom-up attention. *Nat Neurosci* **15**, 1160-1166 (2012).
25. Anderson, B. The attention habit: how reward learning shapes attentional selection. *Ann N Y Acad Sci.* **1369**, 24-39 (2016).
26. Peck, C.J., Jangraw, D. C., Suzuki, M., Efem, R. & Gottlieb, J. Reward modulates attention independently of action value in posterior parietal cortex. *J Neurosci* **29**, 11182-11191 (2009).
27. Strait, C.E., Blanchard, T.C. & Hayden, B.Y. Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron* **82**, 1357-1366 (2014).
28. Kennerley, S.W. & Wallis, J.D. Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *Eur J Neurosci* **29**, 2061-2073 (2009).
29. Sajad, A., Godlove, D.C. & Schall, J.D. Cortical microcircuitry of performance monitoring. *Nat Neurosci* **22**, 265-274 (2019).
30. Hayden, B.Y., Nair, A.C., McCoy, A.N. & Platt, M.L. Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron* **60**, 19-25 (2008).
31. Genovesio, A., Wise, S.P. & Passingham, R.E. Prefrontal-parietal function: from foraging to foresight. *Trends Cogn Sci* **18**, 72-81 (2014).
32. Seo, H., Barraclough, D.J. & Lee, D. Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. *J Neurosci* **29**, 7278-7289 (2009).
33. Seo, H., Barraclough, D.J. & Lee, D. Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cereb Cortex* **17 Suppl 1**, i110-117 (2007).
34. Histed, M.H., Pasupathy, A. & Miller, E.K. Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* **63**, 244-253 (2009).
35. Genovesio, A., Tsujimoto, S., Navarra, G., Falcone, R. & Wise, S.P. Autonomous encoding of irrelevant goals and outcomes by prefrontal cortex neurons. *J Neurosci.* **34**, 1970-1978 (2014).
36. Akrami, A., Kopec, C.D., Diamond, M.E. & Brody, C.D. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature* **554**, 368-372 (2018).
37. Oemisch, M. *et al.* Feature-specific prediction errors and surprise across macaque fronto-striatal circuits. *Nat Commun* **10**, 176 (2019).
38. Kuhns, A.B., Dombert, P.L., Mengotti, P., Fink, G.R. & Vossel, S. Spatial Attention, Motor Intention, and Bayesian Cue Predictability in the Human Brain. *Journal of Neuroscience* **37**, 5334-5344 (2017).
39. Doll, B.B., Simon, D.A. & Daw, N.D. The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol* (2012).
40. Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P. & Dolan, R.J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204-1215 (2011).
41. Dunne, S., D'Souza, A. & O'Doherty, J.P. The involvement of model-based but not model-free learning signals during observational reward learning in the absence of choice. *J Neurophysiol.* **115**, 3195-3203 (2016).
42. Oristaglio, J., Schneider, D.M., Balan, P.F. & Gottlieb, J. Integration of visuospatial and effector information during symbolically cued limb movements in monkey lateral intraparietal area. *J Neurosci* **26**, 8310-8319 (2006).
43. Nystrom, M. & Holmqvist, K. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behav Res Methods* **42**, 188-204 (2010).
44. Phillips, M.H. in Society for Neuroscience Vol. #508.12 (San Diego, CA.; 2012).

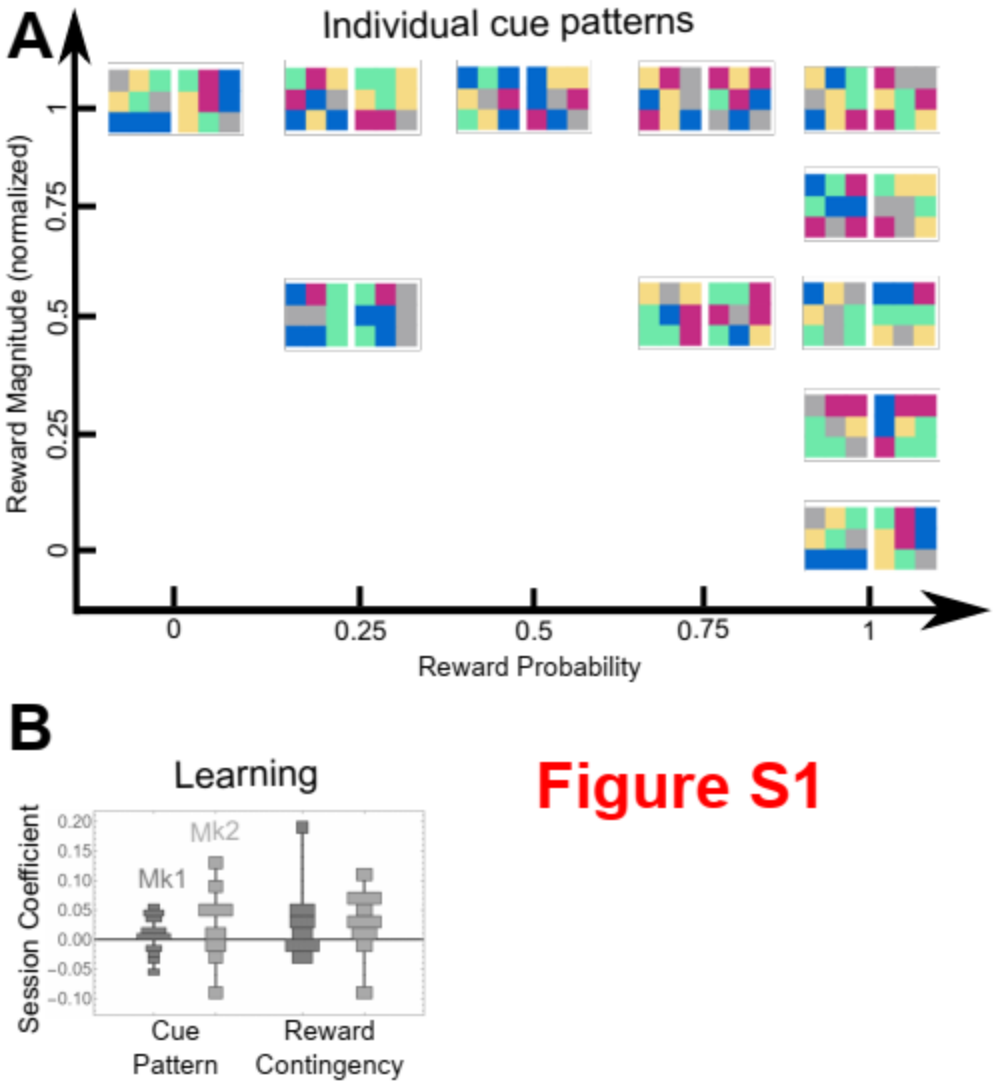

Figure S1. **Cues and learning.** (**A**) **The 20 cues and their assignment to combinations of reward magnitude and probability.** Each cue was a 3x3 colored checkerboard ("Mondrian"), with each tile taking one of 5 possible colors that were defined in DKL space and isoluminant within 2cd. The distance between a pair of cues was the number of tiles that would need to be replaced for the cues to become identical. Individual cues were generated by randomly assigning colors to tiles, with the constraint that all distances are at least 4 tiles. The cues shown here were used for both monkeys. The reward mapping is shown for monkey 1 and was randomized for monkey 2 (not shown). (**B**) **Absence of cue-specific learning.** For each probabilistic cue (P52, P25, P57, P50, P75 in **Fig. 1B**) we computed the LR modulation based on PR (*Methods*, Eq. 1-3) restricting the trials to previous presentations of the same cue pattern or the same reward contingency. The histograms show the distribution of PR coefficients across sessions for each monkey. Had the monkeys updated the cue values the coefficients should be positive, indicating more LR after a large, versus a small, prior reward for that cue or contingencies. However, no distribution was different from 0 for the individual cues (monkey 1, p = 0.15; monkey 2, p=0.16) or reward contingency (monkey 1, p = 0.1; monkey 2, p = 0.06), showing that the monkeys were using pre-learned distributions for these highly familiar cues.

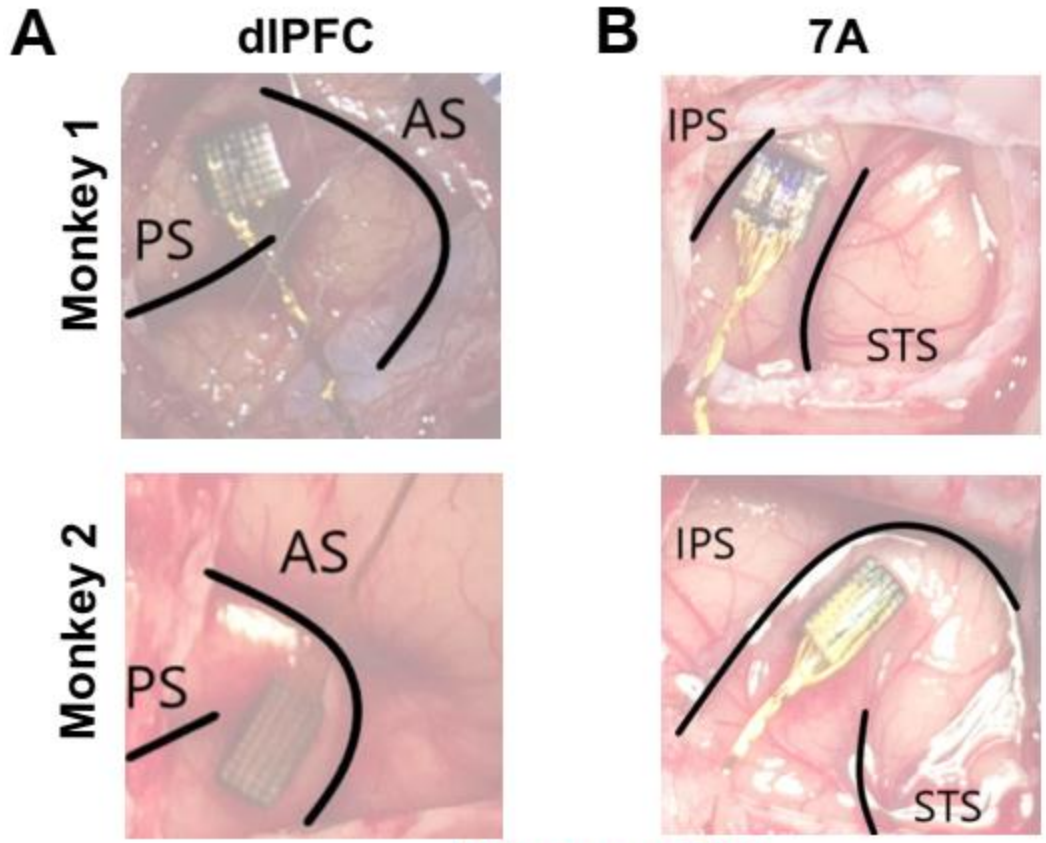

Figure S2

Figure S2. **Recording sites.** Intraoperative photographs showing array placements. (**A**) The dlPFC arrays were implanted between the arcuate sulcus (AS) and the principal sulcus (PS), slightly more dorsal in monkey 1 relative to monkey 2 because of vascular anatomy. (A) The 7A arrays were implanted between the intraparietal sulcus (IPS) and superior temporal sulcus (STS), in the posterior portion of this area that has been targeted in recent studies.

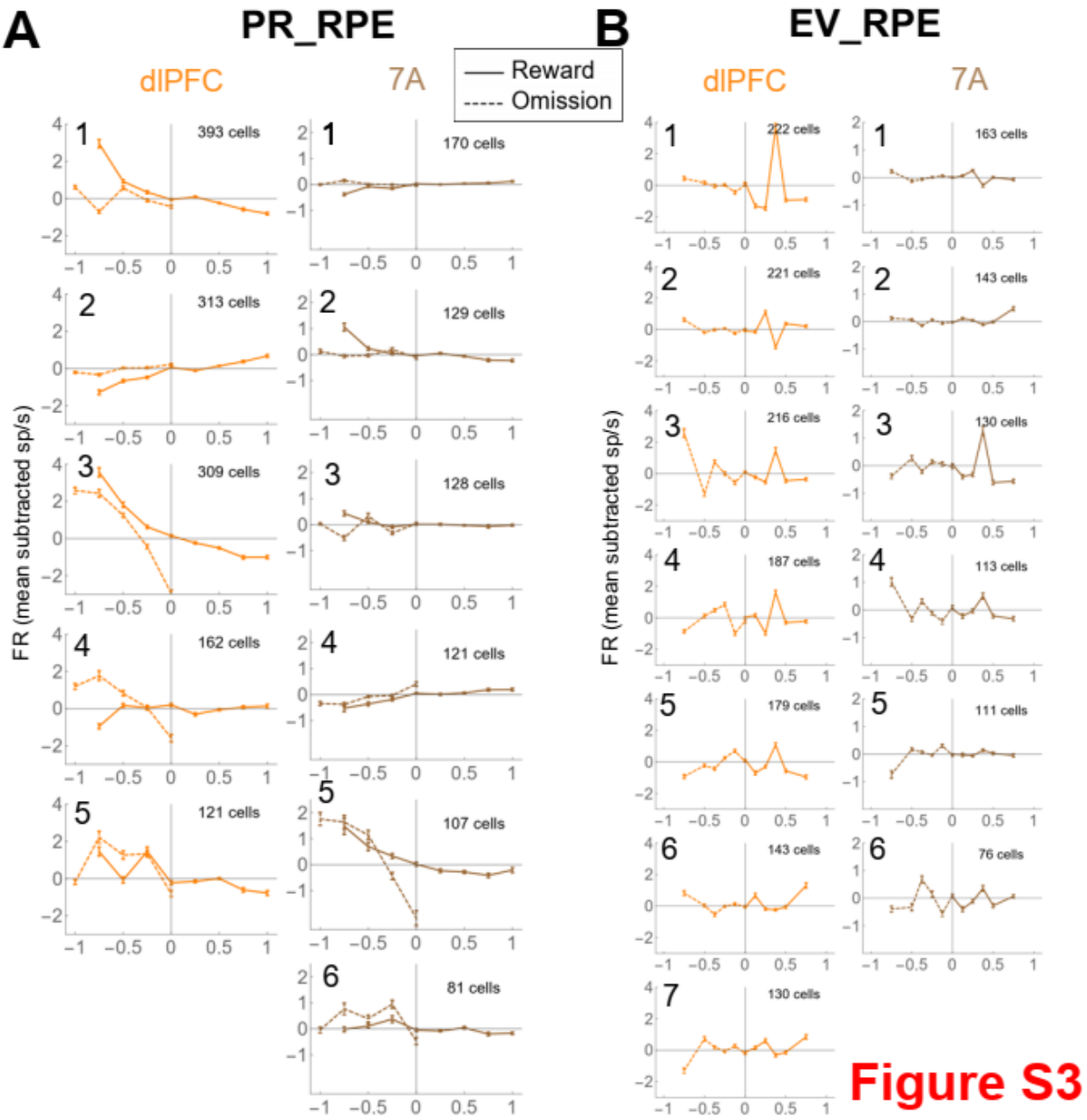

Figure S3. **Diversity of individual neuron responses to PR_RPE (A) and EV_RPE (B)** Vectors of mean deviations were constructed for each cell (as described for Fig. 8A,B, but using the signed rather than absolute values), and then analyzed with k-means clustering with correlation-distance and k chosen for each area based on scree plots. Clusters are ordered according to size in each area. All other conventions are as in **Fig. 8A,B.**