Behavioral/Systems/Cognitive

Encoding of Both Positive and Negative Reward Prediction Errors by Neurons of the Primate Lateral Prefrontal Cortex and Caudate Nucleus

Wael F. Asaad and Emad N. Eskandar

Rhodan Center for Nervous System Repair, Department of Neurosurgery, Massachusetts General Hospital, and Harvard Medical School, Boston, Massachusetts 02114

Learning can be motivated by unanticipated success or unexpected failure. The former encourages us to repeat an action or activity, whereas the latter leads us to find an alternative strategy. Understanding the neural representation of these unexpected events is therefore critical to elucidate learning-related circuits. We examined the activity of neurons in the lateral prefrontal cortex (PFC) and caudate nucleus of monkeys as they performed a trial-and-error learning task. Unexpected outcomes were widely represented in both structures, and neurons driven by unexpectedly negative outcomes were as frequent as those activated by unexpectedly positive outcomes. Moreover, both positive and negative reward prediction errors (RPEs) were represented primarily by increases in firing rate, unlike the manner in which dopamine neurons have been observed to reflect these values. Interestingly, positive RPEs tended to appear with shorter latency than negative RPEs, perhaps reflecting the mechanism of their generation. Last, in the PFC but not the caudate, trial-by-trial variations in outcome-related activity were linked to the animals' subsequent behavioral decisions. More broadly, the robustness of RPE signaling by these neurons suggests that actor-critic models of reinforcement learning in which the PFC and particularly the caudate are considered primarily to be "actors" rather than "critics," should be reconsidered to include a prominent evaluative role for these structures.

Introduction

Unexpected events drive learning (Rescorla and Wagner, 1972; Pearce and Hall, 1980). Reinforcement learning models offer useful accounts of behavioral plasticity that is generated by mismatches between predicted and unpredicted events (Sutton and Barto, 1998). The difference between the actual outcome of a situation or action and the expected outcome is the reward prediction error (RPE). A positive RPE indicates the outcome was better than expected while a negative RPE indicates it was worse than expected; the RPE is zero when events transpire according to expectations. Midbrain dopamine neurons exhibit activity that conforms in many aspects to this predicted RPE signal (Hollerman and Schultz, 1998; Waelti et al., 2001): many display phasic activations consistent with positive RPEs and transient inhibitions correlated with negative RPEs. The modulation of dopamine in downstream cortical and basal ganglia targets is thought to promote neural plasticity and behavioral change (Reynolds et al., 2001; Canales et al., 2002; Tsai et al., 2009).

There is no consensus regarding the extent and manner in which positive and negative RPEs are represented by neurons in these downstream areas targeted by dopamine. Studies directly investigating the activity of single neurons in the lateral PFC, which is thought to be critical for associative learning, have described the influence of choice and reward history (Barraclough et al., 2004; Ichihara-Takeda and Funahashi, 2006; Seo et al., 2007; Histed et al., 2009). However, some of these have failed to reveal evidence of RPEs in the lateral PFC (Matsumoto et al., 2007; Kennerley and Wallis, 2009), whereas others have found activity related to unexpected outcomes, but did not explore differences between positive and negative RPEs (Seo et al., 2007).

Meanwhile, in the basal ganglia, despite the variety of learning-related phenomena observed here (Schultz et al., 1993; Tremblay et al., 1998; Brasted and Wise, 2004; Schmitzer-Torbert and Redish, 2004; Barnes et al., 2005; Pasupathy and Miller, 2005; Williams and Eskandar, 2006; Kimchi et al., 2009), the encoding of RPEs by striatal neurons (and in particular, by the medium spiny neurons that constitute the significant majority of striatal neurons) has been reported as rare, if present at all (Schultz et al., 1998; Kim et al., 2009; Oyama et al., 2010).

Therefore, we sought to examine the manner and extent to which neurons in the lateral PFC and anterior caudate nucleus represent positive and negative RPEs. We hoped to begin to elucidate their potential mechanistic substrates by examining the pattern of their corepresentation by individual neurons, and by

Received July 25, 2011; revised Sept. 9, 2011; accepted Sept. 29, 2011.

Author contributions: W.F.A. and E.N.E. designed research; W.F.A. performed research; W.F.A. analyzed data; W.F.A. and E.N.E. wrote the paper.

This work was supported by grants from the National Eye Institute (1R01EY017658), National Institute on Drug Abuse (1R01NS063249), National Science Foundation (IOB 0645886), NIDA Conte Award 1P50MH086400-03, and the Howard Hughes Medical Institute. W.F.A. was also supported by a Tosteson Fellowship. We thank Anne-Marie Amacher, Ming Cheng, John Gale, Kenway Louie, Matt Mian, Mayank Mehta, Shaun Patel, Sameer Sheth, and Ziv Williams for helpful discussions or technical assistance.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Wael F. Asaad, Neurosurgery, APC 6th Floor, Rhode Island Hospital, 593 Eddy Street, Providence, RI, 02903. E-mail: wfasaad@alum.mit.edu.

W.F. Asaad's present address: Brown Institute for Brain Science and the Department of Neurosurgery, Alpert Medical School, Brown University and Rhode Island Hospital, Providence, RI 02903.

DOI:10.1523/JNEUROSCI.3793-11.2011

 0270-6474/11/3117772-16$15.00/0

investigating their relative timing, both within and across structures. We trained two monkeys to perform a learning task with multiple reversals, and with a mnemonic component of the sort known to activate these regions (Fuster and Alexander, 1971; Kojima and Goldman-Rakic, 1982; Rainer et al., 1998). The task was designed to employ simple and stable learning rules, but also to promote the occurrence of unexpected positive and negative outcomes. We then recorded the activity of individual lateral PFC and caudate neurons while the animals performed this task.

Materials and Methods

Task. Two monkeys ("M1" and "M2") performed a temporally delayed, on-line learning task in which they had to determine by trial-and-error which of four picture cues or spatial locations was currently rewarded within a particular block. Object-learning and spatial-learning blocks were interleaved in a pseudorandom fashion such that each of the four cues or locations was designated correct before any were repeated, and so that the number and type of transitions between tasks were balanced. Trials requiring object learning were identical in all sensory and motor respects to those requiring spatial learning. No explicit cue identified the type of learning required, or when a block switch occurred. The monkeys needed to learn both the type of feature that was rewarded (object or location) and the particular exemplar (specific object or feature from among the four possibilities in each class) simply by trial-and-error.

Each trial began with the acquisition and maintenance of central fixation for 1000 ms. Four cues were then presented simultaneously in the corners of a visual display (lasting 500 ms), followed by a short delay (1000 ms), followed by a "go" signal (extinguishing of the central fixation dot) instructing them to choose one of those four locations by executing a saccade to it. If the chosen location had contained the rewarded cue or was the currently rewarded direction, then a generic, positive visual reinforcer was presented (a green circle for 500 ms), and then a juice reward was delivered. If the chosen location was incorrect, a generic negative reinforcer was presented (a red "x" for 500 ms), which was not followed by reward delivery. Trials were separated by a 2000 ms intertrial-interval.

A learning criterion of four consecutive correct choices was used because the random probability of that event with four response options was less than one percent ($0.25^4 = 0.0039$). Once the association was learned, and the monkeys' behavior had been allowed to plateau for an additional number of correct trials (40), a new block began and the identity of the rewarded cue was changed without any external signal to the animal. At this point, each animal was required to abandon the previously learned association and learn a new one. At least 16 blocks were completed in any recording session.

Each day, four new cues were selected pseudo-randomly from a pool of 50 naturalistic images, familiar to each animal, each subtending 2 degrees of visual angle. The use of familiar stimuli allowed us to discount the influence of perceptual learning on neural activity, and the use of a sizeable pool of familiar images ensured that the same set of four objects was never used more than once, so that the cue array was unique in each session. During cue presentation, each picture was presented at a distance of 7 degrees of visual angle from fixation, at 45, 135, 225, or 315 degrees from the horizontal. The particular configuration of the four cues among the four positions was chosen pseudo-randomly on each trial (i.e., each of the 24 possible configurations was selected randomly without replacement, until exhausted). Therefore, in object learning blocks, the rewarded location associated with the correct cue varied randomly from trial to trial and could not be used to predict reward. In spatial learning blocks, these same cue arrays were irrelevant to determine the rewarded direction.

The behavioral task was executed using custom psychophysical software ("MonkeyLogic") running on a specially configured Windows XP-based PC (Microsoft) in MATLAB (MathWorks). The adequate temporal precision of this software has been confirmed and reported (Asaad and Eskandar, 2008), and further temporal refinements had since been made (see www.monkeylogic.org). Eye position was monitored at 120 Hz using an infrared optical eye-tracking system (iScan).

Subjects. Two male rhesus monkeys (*macaca mulatta*) weighing 6.5 and 4.5 kg were trained to perform the behavioral tasks using apple juice as reward. They were water restricted so that they received their daily allotment of fluid (40 ml/Kg) only during task performance. They were each implanted with a head bolt restraint to allow reliable optical tracking of eye movements, and to permit stable single-unit neuronal isolation. Once initial training was complete, they were implanted with recording chambers positioned over the left lateral prefrontal cortex and the head of the caudate nucleus using standard stereotactic coordinates (Paxinos et al., 2000). Both animals were always handled strictly in accordance with NIH policies and those of the Massachusetts General Hospital animal care and use committee.

Neurophysiology. In each session, multiple single neurons were recorded simultaneously, extracellularly, from 6–16 individual electrodes driven by a custom-built multielectrode microdrive system. We relied on structural T1-weighted MRI scans with fiducial markers representing potential electrode trajectories to position electrodes over the lateral PFC, centered on the principal sulcus and angled slightly posteriorly to allow deeper recordings from the anterior caudate nucleus. The total lateral PFC recording area in each animal was $\sim$2.2 cm^2 across the two-dimensional surface tangent to the lateral cortical surface of the PFC. Trajectories to the anterior caudate nucleus were limited to a smaller area within a posterior subregion, encompassing $\sim$1.0 cm^2. In each animal, a clearly bimodal distribution of recording depths was obtained with no overlap, corresponding to locations in the lateral PFC superficially and caudate nucleus more deeply, as expected based upon trajectories planned according to the MRI scans.

Electrodes were driven until well isolated single units were encountered, without regard to any behavioral correlation of activity. After isolation, at least 2–3 h elapsed before behavior and data acquisition commenced, to allow the signals to stabilize. On each channel, all waveforms surpassing a threshold voltage were sampled at 40 kHz and stored for off-line spike sorting (Plexon). Spikes were classified as belonging to individual neurons based upon an analysis of waveform features, including maximum and minimum voltage, voltage range, and waveform principal components. Spike clusters that were not separable from low-amplitude multiunit activity, or from other single-unit clusters, were left unsorted and were not included in any analysis. A low fraction of short-latency events on an interspike-interval (ISI) histogram ($< 0.5\%$ <2 ms) was necessary but not sufficient for the grouping of spikes into a single-unit. Taking into account all sorted spikes, $<0.1\%$ of those assigned to single units had ISIs less than or equal to 1.5 ms.

Data and statistical analysis. Behavioral and neurophysiological data were coregistered using digital codes sent from the behavioral computer to the neurophysiology computer to time-stamp the occurrence of key events. These data were then analyzed using MATLAB (MathWorks).

For this report, we focused on neuronal activity beginning with the onset of the feedback cue, after each behavioral choice. We included correct and incorrect trials, so long as those incorrect trials were technically successful in every other respect but for the correctness of the chosen target (e.g., trials with failure to fixate, break fixations, or early responses were excluded from analysis). We assigned each of these trials to one of four categories, depending upon the outcome in the current trial and the outcome in the immediately preceding trial: Correct trials that followed a preceding correct trial were termed expectedly positive (EP). Correct trials that followed a preceding incorrect trial were termed unexpectedly positive (UP). Incorrect trials that followed a preceding incorrect trial were expectedly negative (EN), and incorrect trials that followed a preceding correct trial were unexpectedly negative (UN). This method provided a straightforward, minimal-model method of evaluating the influence of expectation and outcome on neuronal activity without including additional parameters (e.g., the number of trials to include as part of the history, and the weighting given to each). Supporting this approach, previous work has demonstrated that while the activity of some lateral PFC neurons reflect outcomes at least two or three trials back, the immediately preceding trial was indeed the most potent influence (Seo et al., 2007), and in any case the exact integration time is likely to vary with the stability of any particular task (Sugrue et al., 2004; Kennerley et al., 2006). We also present behavioral data in Results to demonstrate the stationarity of the animals' strategies, and their lack of reliance on more extended trial histories.

To determine reliable outcome-selectivity for each neuron, two sliding receiver operating characteristics (ROCs) were calculated across the trial (200 ms width moving in 50 ms steps), and the time between feedback-onset and feedback-onset +1000 ms was considered. The first ROC reflected the difference in neuronal activity between unexpected and expected positive outcomes (UP and EP), while the second reflected the difference between unexpected and expected negative outcomes (UN and EN). In neither ROC was the comparison confounded by immediate differences in visual stimulation (the feedback cues were identical within each pairing), or the presence versus absence of reward. The only factor differentiating these trials was the outcome history as defined by the single preceding trial. Sensory and motor features (cue arrangement, saccade direction, and rewarded cue or location) were balanced within each response category, and so did not differentially influence one over another.

ROC areas-under-the-curve >0.5 (chance) indicated neuronal preferences for the unexpected over the expected condition within each pair (i.e., such neurons were UP- or UN-responsive), whereas areas <0.5 indicated the opposite (neurons were EP- or EN-responsive). To assess the significance of these ROC areas, bootstrap tests were performed. For each pair of conditions comprising an ROC value (UP-EP or UN-EN), the trial-by-trial spike rates were shuffled and reassigned among the pair of conditions 1000 times, followed by a calculation of the ROC area. This created a distribution of randomized ROC areas from which the p value of the actual ROC area could be obtained. This was repeated for each ROC value in each 200 ms bin for each neuron. An α level of 0.05 was used with a correction applied for multiple comparisons (Benjamini and Hochberg, 1995); neurons with significant outcome-related responses were first identified using a starting threshold of: α/(number of bins), or $0.05/20 = 0.0025$. Similarly, to get an estimate of the total number of outcome-responsive neurons across each population (i.e., within each area in each animal), the lowest p value from each neuron was tested against a starting threshold of: α/(total number of neurons in that population). For subsequent analyses, when multiple bins were found to be significant for a particular neuron within a specific category of outcome-related activity (UP, UN, EP, or EN), the average spike rate across these bins was used.

To assess the contribution of a preceding trial n-k on neuronal activity in trial n, we used an ROC analysis on spike rates sorted by preceding outcomes. Specifically, this analysis calculated the area under the ROC comparing the distributions of spike rates in trial n when the outcome on trial n-k was correct versus incorrect. The independent contribution of trial n-k was found by performing separate ROCs for each subset of trials in which the outcomes on trials n-k+1 to n were held constant, and then taking the mean. For example, the ROC at trial n-3 represents the ability to discriminate the firing rates at trial n using the outcome at trial n-3 while holding constant the sequence of outcomes at trials n, n-1, and n-2. For each neuron, spike rates in significant time bins during the feedback period were used, as found above.

We also used a linear regression to assess the contribution of the correctness of preceding trials on outcome-related neuronal activity. We considered each of the four response types individually, and included correct versus incorrect (1 or 0), chosen object (1–4), or chosen location (1–4), in trials n-5 to n-1, as potential factors contributing to the observed spike rate on trial n. Only time bins with significant outcome-related activity—as determined by the initial ROC analysis—were used. The population means of the resulting regression coefficients were then compared with zero using two-tailed t tests, with a correction applied for multiple comparisons, as above.

To determine whether neurons were responding to the outcome of the previous trial rather than that of the current trial, we calculated for each neuron an index based upon the mean level of activity for each of the four outcome types (UP, EP, UN, or EN) using the same time bins as for the ROC analysis, above. The index equaled the following:

$$\ln\left(\frac{|\mathrm{UP} - \mathrm{EP}| + |\mathrm{UN} - \mathrm{EN}|}{|\mathrm{UN} - \mathrm{EP}| + |\mathrm{UP} - \mathrm{EN}|}\right) \quad (1)$$

The numerator is derived from the absolute differences in neuronal activity between pairs of trials that share an outcome on the current trial, but are preceded by trials of differing outcomes (UP and EP trials are both correct trials, but the former is preceded by an incorrect trial and the latter by a correct trial; likewise UN and EN trials are both incorrect, but differ in the outcome of the preceding trial). Meanwhile, the denominator is derived from the absolute differences in activity between trials that differ in the current outcome, but share a prior trial's outcome (e.g., UN and EP trials are both preceded by correct trials, but have different outcomes on the current trial). Thus, when the outcome of the previous trial was the major determinant of a neuron's activity, the numerator would be larger and so the natural log of this ratio would be greater than zero. In contrast, a larger denominator reflects differences that are driven more by the outcome of the current trial than by the previous one; in that case, the log would be less than zero. The mean value of this index across time bins was calculated for each neuron in either the pre-feedback period (specifically, the fixation, cue, and delay periods) or in the post-feedback period (feedback onset to feedback offset + 500 ms). Two-tailed t tests were then used to assess whether the means of each distribution differed from zero. All neurons were used in this analysis, regardless of significance on the ROC.

The number of neurons responding to more than one category of outcome was represented using an overlap fraction. This was simply the number of neurons responding to both of a pair of categories divided by the sum of neurons responding to either of the pair, alone or in combination with the other. It varied from zero to one, corresponding to strict absence of combined responses to exclusive presence of combined responses. Because this measure varies according to the individual frequencies of each category, we also calculated a distribution of expected overlap values for each of the six possible pairs of categories, for each animal and for each recorded area (PFC or caudate). This was done by reassigning the observed number of significant responses for each category randomly across neurons, and recalculating the overlap fraction 10,000 times. In other words, to create the random distributions, the frequency of each response was maintained but the particular neuronal assignments were shuffled. The resulting two-tailed p value was then derived by comparing the actual overlap fraction against the associated random distribution.

Because certain categories of outcome represent opposite polarities of the same analysis (UP-EP or UN-EN), it was inherently less likely that a neuron would represent both types of outcome from a given pair because time spent signaling one category (say UP) leaves less time available for a neuron to represent the opposite type of outcome (here, EP). Indeed, we found that in most cases neurons were less likely to show both UP and EP activity, or to show both UN and EN activity (see Fig. 8, Table 3). These findings were therefore expected and considered artifacts. Instead, we focused on the potential association or dissociation of UP and UN responses.

The latencies to appearance of UP or UN activity were based on ROCs calculated in 50 ms time bins stepped 5 ms across the 500 ms feedback period (feedback-onset to feedback-offset). The significance of each ROC area-under-the-curve was determined with a bootstrap analysis as above. The first significant time bin (if any) with UP or UN activity was found for each neuron, and the cumulative latency curves of these ascending-sorted values were then calculated. These curves were scaled to account for the differing numbers of neurons across groups that would otherwise distort the results due to sampling bias by normalizing with respect to the relative sizes of the populations being compared (along the x-axis) and by the total area of each distribution (on the y-axis). Cumulative distributions are shown because they are particularly sensitive to the higher moments of each distribution of latencies (such as skew), which could reveal differences in the timing of information arrival. Significant differences in the underlying (non-normalized) distributions of first-significant-bin latencies that gave rise to these cumulative latency curves were assessed with two-sample, two-tailed Kolmogorov–Smirnov tests.

To examine the association of neuronal activity with behavior, we performed a form of reverse correlation of the latter with the former. We hypothesized that trial-by-trial variations of neuronal activity might be correlated with different levels of influence on the behavioral output (here, the animal's choice). Thus, if UP- and UN-related neuronal activity were linked to behavior, a relatively higher level of neuronal activity

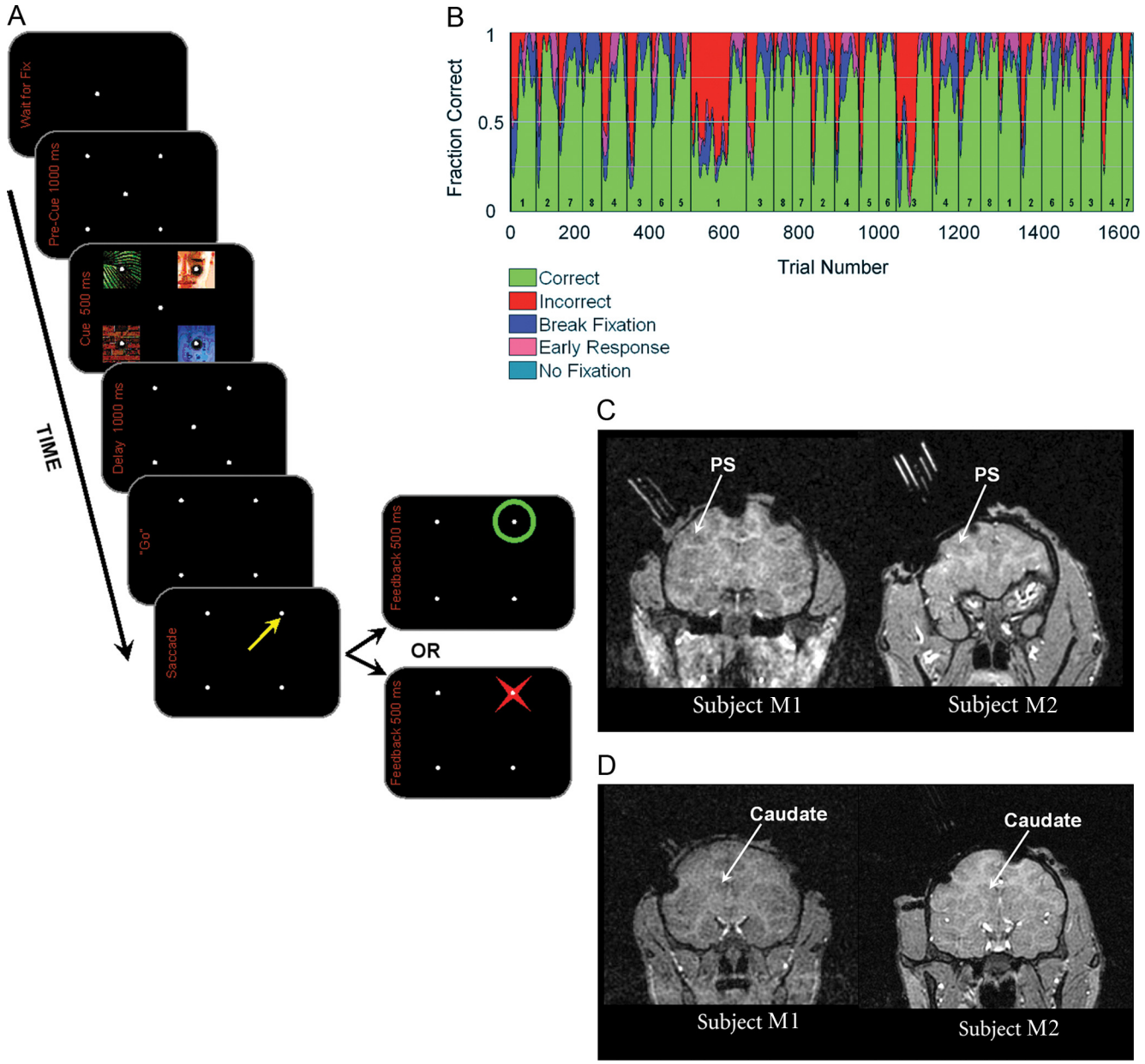

Figure 1. Behavioral task and recording areas. ***A***, A schematic representation of the behavioral task (see Materials and Methods). ***B***, A typical recording session is shown with trial number along the *x*-axis and behavior (calculated using a 10-trial moving average) along the *y*-axis. Correct trials are represented by the green area, and incorrect choices are shown in red. The other colors represent procedural errors; neuronal activity from these latter trial types, reflecting technical errors rather than decisional ones, was not used in any analysis. The vertical lines represent reversals between blocks, and the numbers at the bottom of each block represents the particular cue (1–4) or direction (5–8) that was designated correct during that block. ***C***, T1-weighted MRIs with fiducal markers placed within the recording chambers to demonstrate a subset of trajectories into the left lateral prefrontal cortex of each monkey, M1 and M2. PS, Principal sulcus. ***D***, Fiducial markers within the posterior portion of the same recording chambers demonstrate sample electrode trajectories into the anterior caudate nucleus.

on a UP trial (by UP-responsive neurons) should more likely be associated with selection of the same behavioral response on the subsequent trial, whereas for those neurons responding to UN outcomes, a higher level of activity on UN trials should more likely be associated with selection of a new behavioral response on the subsequent trial. Therefore, we used the binary factor (new vs same choice on the subsequent trial) to sort neuronal activity corresponding to either UP or UN trials. Then, for each neuron, we subtracted the mean spike rate in those trials followed by "same" choice behavior from the mean spike rate in the remaining trials that were followed by "new" choice behavior. This difference provided a simple measure to assess how trial-by-trial variations in neuronal response within an outcome category (UP or UN) were linked to subsequent behavior. Only the first three trials from each block were used (though the results are insensitive to the exact number), because a spatially based choice strategy was evident most strongly in earlier trials within each block (later, subjects tended to adopt an object-based strategy if the correct feature appeared to be an object rather than a location); this allowed us to use selection of a new or same direction as the behavioral key.

Results

Behavior

Animals learned the spatial task (Fig. 1) more quickly than they learned the object task (Fig. 2*A*,*B*). This was due, at least in part, to a default reliance on a spatial strategy. In other words, when a correct outcome was encountered after a prior trial had been incorrect (i.e., after an unexpected positive, or "UP" trial), they

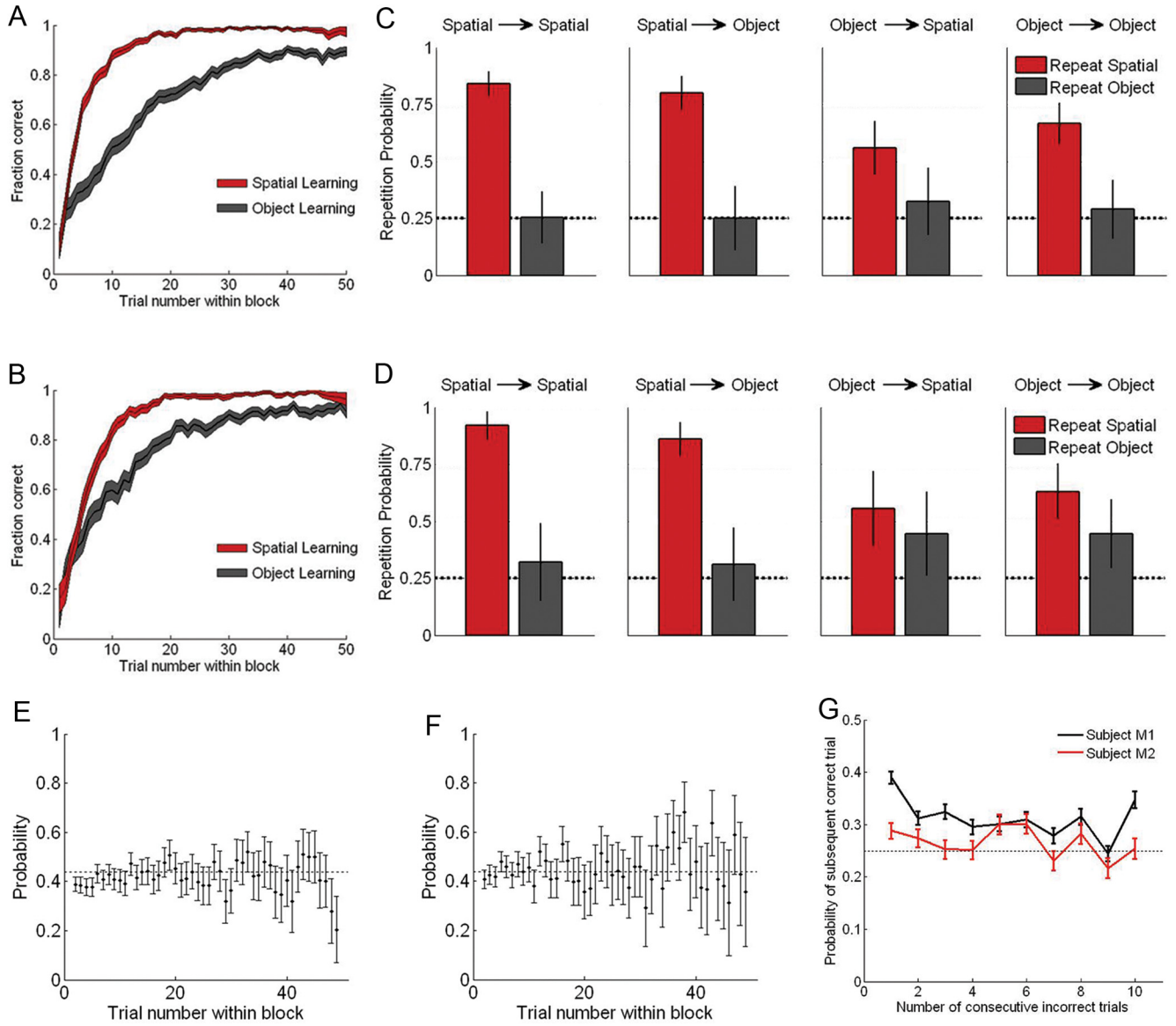

Figure 2. Behavioral performance and strategy. ***A***, ***B***, Percentage correct as a function of trial number within a block for subjects M1 and M2, is plotted separately as a function of learning rule ("object" or "spatial"). The shaded area bounds the mean ±SE. Animals learned more quickly in spatial learning blocks than in object-learning blocks. ***C***, ***D***, Behavioral strategy following a UP trial is shown. If subjects were using a spatial strategy, then the occurrence of an UP outcome should lead them to repeat the chosen direction, regardless of the object that had cued that direction (i.e., "repeat spatial"). If they were using an object-based strategy, they should instead "follow the object" that had cued the correct location just chosen, wherever it appears on the subsequent trial ("repeat object"). The initial strategy used by each animal (based on the first 3 occurrences of UP trials in each block) was influenced by the type of the preceding block, but was generally biased toward a spatial strategy. This dependency on the preceding block type weakened as the number of trials into a block increased, although it was still evident across the entire block (data not shown). The probability of selecting the same direction or object by chance was 0.25 (indicated by the dashed line). A three-way ANOVA with main factors of (1) previous block type (spatial vs object); (2) current block-type; and (3) strategy (i.e., repeat spatial vs repeat object) revealed a significant main effect of strategy (M1: $p < 0.0001$; M2: $p < 0.0001$), and a significant interaction between previous block type and strategy (M1: $p = 0.0057$; M2: $p = 0.0014$). No other main effects or interactions were significant in either subject. ***E***, ***F***, Behavioral strategy after UN trials is plotted as a function of trial number within a block. The chance probability of reselecting either the same object or direction was 0.4375 (indicated by the dashed line). For each subject, performance returned approximately to chance after any incorrect trial, regardless of its timing within a block. Over all trials, M1 was slightly less likely to repeat the same object or direction after an error ($p = 0.025$), whereas the pattern of choices made by M2 did not differ from chance ($p = 0.402$). ***G***, The likelihood of a correct choice after *n*-consecutive incorrect trials is plotted for each subject. Subject M1 was slightly more likely than M2 to make a correct response after an incorrect response (likely related to this subject's slightly greater tendency to avoid reselecting a previously chosen object or direction, as shown in ***E***). However, there was no large variation in performance as a function of the number of consecutive incorrect choices for either subject, suggesting they relied predominantly on a guessing strategy, rather than accumulate information from each consecutive incorrect response.

tended to reselect the chosen direction of that correct trial on the subsequent trial, rather than "follow the object" that had cued the chosen location. This bias was evident in the types of errors made by each animal within each block type: in object-learning blocks, 75.0% (subject M1) and 73.2% (subject M2) of errors could be attributed to a spatial strategy (i.e., reselecting the same location after a correct trial, resulting in an incorrect choice), while in spatial-learning blocks, 52.7% (M1) and 54.0% (M2) of errors could be attributed to an object-based strategy (reselecting the same object after a correct trial, resulting in an incorrect choice).

The degree of reliance on a spatial rather than object strategy was biased by the type of preceding block (Fig. 2*C*,*D*); when the preceding block employed a spatial rule, use of a spatial strategy was most pronounced. However, when the preceding block em-

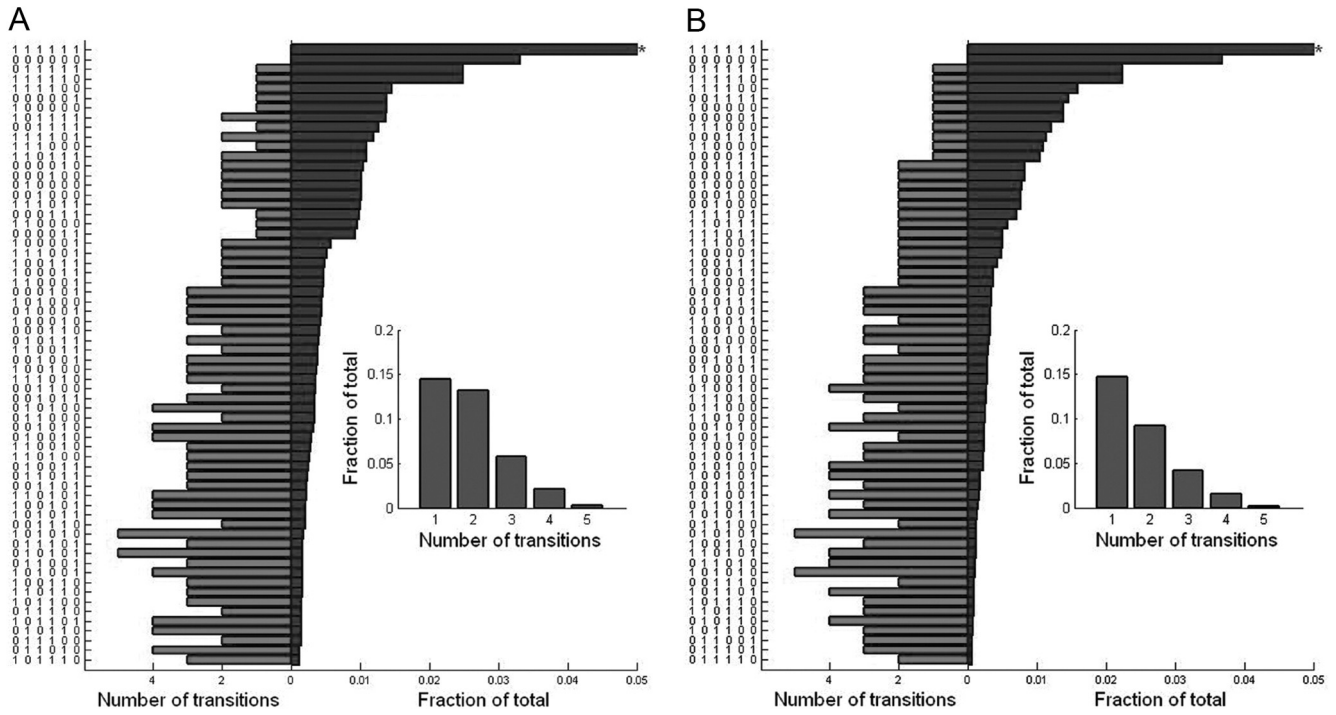

Figure 3. Incidence of specific behavioral patterns. ***A***, ***B***, The relative incidence of particular sequences of behavioral outcomes is plotted for M1 (***A***) and M2 (***B***). Each of the 64 possible 6-trial sequences of correct (labeled "1") and incorrect ("0") outcomes is arranged along the *y*-axis in order of frequency. The proportion of all 6-trial segments corresponding to that pattern is shown on the *x*-axis, to the right (in blue). The total fraction represented by the first bar (indicated by an asterisk) was 0.61 and 0.66 for M1 and M2, respectively. The red bars to the left reflect the number of transitions between correct and incorrect trials contained in each sequence. For example, the sequence "110110" has 3 transitions (trials 2–3, 3–4, and 5–6). The inset plots the frequency of these different patterns according to the numbers of transitions they contain. These plots provide a measure of behavioral stability, and also would reflect reliance on idiosyncratic strategies that depend upon particular sequences of outcomes. The simplest sequences are most common. Therefore, most unexpected responses occur in regimes of relatively stable behavior.

ployed an object rule, the bias toward a spatial strategy was reduced or abolished (but not reversed).

It was possible that the stage of learning (approximately, number of trials into a block) influenced behavioral strategy. For example, subjects might more readily explore alternative options after an incorrect trial early in a block, whereas they might persist in choosing that same incorrect option after learning had occurred, due to a recent history of successive positive reinforcement. Furthermore, they could perhaps expect that, after many trials, a reversal is imminent, and so again be more willing to explore after an incorrect trial. In actuality, examining the tendency to reselect the same object or direction after an incorrect trial, we found that performance essentially returned to chance whenever a negative outcome was encountered, regardless of the number of trials into a block (Fig. 2*E*,*F*). This rapid unlearning is very similar to the behavior observed in another learning task with multiple reversals (Fusi et al., 2007), and likely reflects behavioral adaptation to an unstable task environment.

If subjects' behavior were optimal, each successive incorrect choice would lead to the accumulation of information, constraining subsequent choices to the remaining, unexplored alternatives. Therefore, a correct outcome encountered after four consecutive incorrect choices should be less surprising than one encountered after two consecutive incorrect choices. However, examining the likelihood of a correct choice after n consecutive incorrect choices (Fig. 2*G*) revealed that there was relatively little variation above chance (especially compared against theoretically optimal behavior, in which there would be no sequence of trials or strategy that would require >7 incorrect choices to achieve 100% correct performance, given four locations and four objects, sampled independently). Therefore, the animals' experience did not provide a basis for differential expectation after particular sequences of incorrect trials; again, they appeared to rely in large part on random guessing, ignoring the potential information available from consecutive incorrect choices.

The relative stationarity of these performance measures allowed us to apply a simplified analysis framework to examine the neuronal data. Specifically, each trial was assigned to one of four outcome categories based solely on the outcome of the immediately preceding trial: A correct trial that followed an incorrect trial was termed a UP outcome, whereas a correct trial that followed another correct trial was termed an EP outcome; likewise, an incorrect trial that followed a correct trial was UN, and one that followed another incorrect trial was EN. Thus, when comparing UP to EP trials, the visual cues and delivery of reward were identical. Only the previous trial differed across these conditions. Similarly, a comparison of UN and EN trials was not confounded by immediate events, as the negative visual feedback and absence of reward were identical. Furthermore, the visual cues and response directions were balanced across these conditions, such that they did not differentially influence any of these four outcome categories.

In essence, we assumed that an animal's expectation for the current trial was set primarily by its experience on the immediately preceding trial. The advantage of this method is that it is a simple, almost model-free way of determining expectation. However, there may be times when a subject's performance is unstable, vacillating between correct and incorrect trials so frequently as to confound any sense of expectation. Alternatively, a complicated sequence of outcomes could reflect a rational strategy that is overlooked by considering only the immediately preceding trial: For example, if a subject is using a spatial strategy (i.e., "reselect a

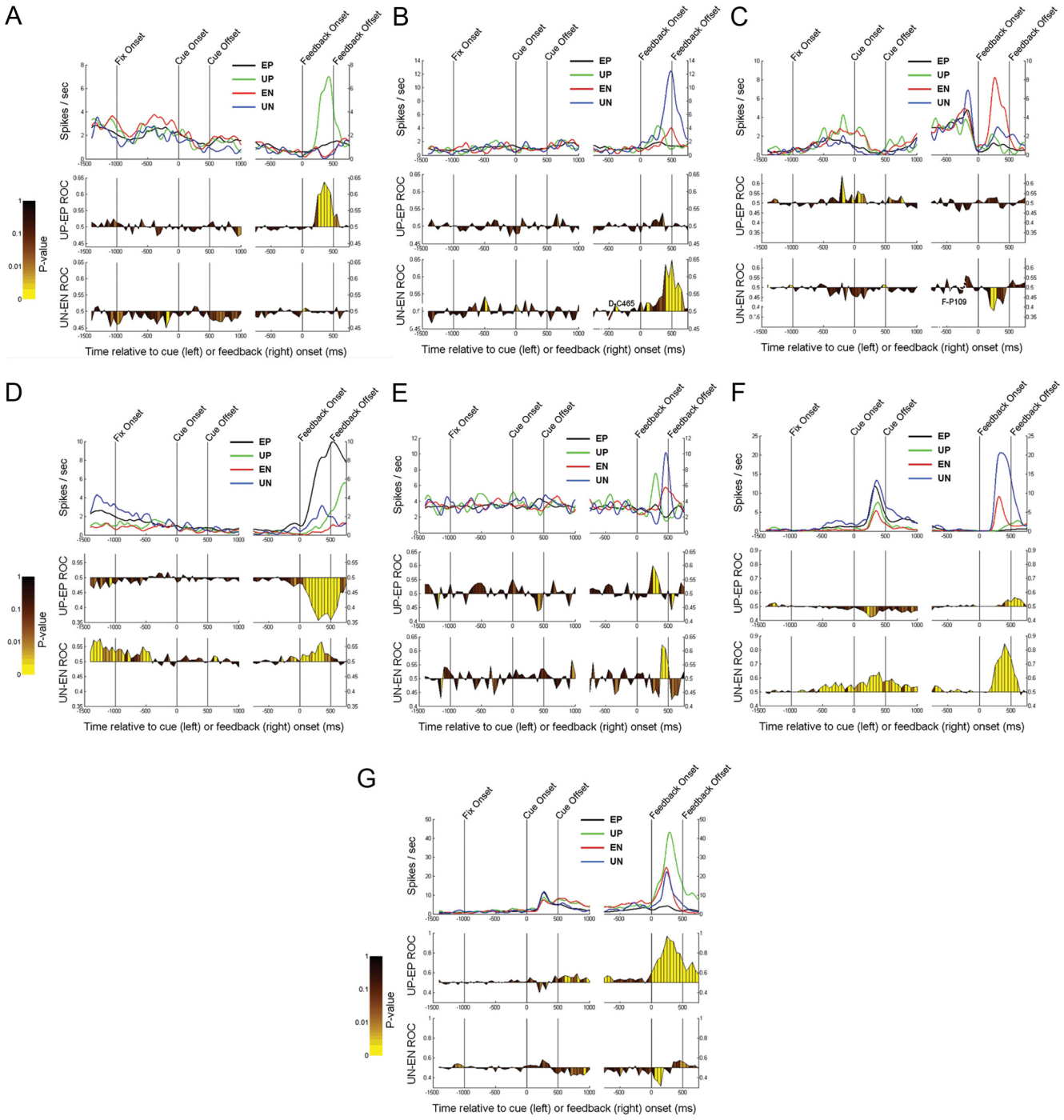

Figure 4. Single neuron examples. ***A***, Top, This caudate neuron demonstrated a feedback-related response to unexpectedly positive outcomes. A slight depression of activity was seen for both expected and unexpected negative outcomes. Activity was less modulated by expected positive outcomes. This neuron recapitulated the pattern of activity described by others for midbrain dopamine neurons. Neuronal activity was smoothed with a sliding Gaussian kernel with $\sigma = 50$ ms. The left-side portion of the figure is aligned to the cue onset (at 0 ms); fixation onset occurred at −1000 ms and cue offset occurred at 500 ms, both marked by vertical lines. The right-side portion of the figure is aligned to the onset of visual feedback (at 0 ms), while offset of visual feedback occurred at 500 ms, coincident with the onset of reward delivery in correct trials. Green, UP; black, EP; blue, UN; red, EN. Middle, A sliding ROC comparing UP and EP responses. Deviations >0.5 indicate the UP response was greater than the EP response; those <0.5 reflect greater activity for the EP than for the UP condition. Each bar is color-coded to represent the *p* value calculated from a bootstrap distribution created for each 200 ms bin, slid in 50 ms steps. Bottom, The sliding ROC comparing UN and EN responses. Deviations >0.5 indicate greater activity in UN trials whereas deviations <0.5 reflect greater activity in EN trials. Conventions in ***B***–***G*** are as for ***A***. ***B***, This caudate neuron showed greater activity, beginning during visual feedback, for UN outcomes. ***C***, A PFC neuron that showed greater activity for EN outcomes. ***D***, A PFC neuron that exhibited greater activity on trials with EP outcomes. This neuron also had a small selective response on UN trials. ***E***, This caudate neuron displayed brief phasic responses of approximately similar magnitude to both UP and UN outcomes. ***F***, This PFC neuron had a larger, earlier response to UN outcomes followed by a smaller response to UP outcomes. It also had greater activity earlier in the trial around the time of the cue and early delay period, when the previous trial was correct, as suggested by the slight separation between UN and EP trials (which are preceded by correct trials) from UP and EN trials (which are preceded by incorrect trials), similar to what has been reported previously by others (Seo et al., 2007; Histed et al., 2009). ***G***, A PFC neuron that displayed a small early response to EN outcomes, but a larger and more sustained response to UP outcomes. This neuron also demonstrated slightly increased firing rates during the delay period when the previous trial was incorrect.

correct location, regardless of object") in an object-learning block, it may by chance make a correct choice, followed subsequently by an incorrect choice when the correct object shifts to a new location, followed again by a correct choice when the subject infers that it had been the object rather than the location that led to the first correct outcome in this sequence. Here, one could justifiably argue that the last correct trial in this 0-1-0-1 sequence (where "0" is an incorrect trial and "1" is a correct trial) is completely expected, and therefore should not be assigned to the UP category.

Therefore, to assess the prevalence of such patterns and to examine the local variability of behavior, we tabulated the frequency of all combinations of correct and incorrect outcomes over six trial sequences. The resulting distribution of frequencies for each of the 2^6 possible sequences is shown for each animal in Figure 3. The simplest sequences, basically corresponding to the most straightforward and stable behavior, were the most common. Specifically, only ~10% of sequences contained >2 transitions. Furthermore, even if a higher-order strategy were in use, it was unlikely to be evident in all of these more complicated sequences; rather, it likely would have manifested as only some subset of that ~10%. These behavioral data therefore supported our approach of inferring expectation from the immediate history of outcomes, as it was much less likely that these category assignments were drawn from regimes of unstable behavior devoid of expectation, or that more complicated sequences, such as that hypothesized above (containing at least three transitions between different outcomes), contributed significantly to the animals' experience.

Of course, it is possible that in any isolated instance the prior trial did not make the inferred impression, or that a particular sequence of outcomes did contribute uniquely to expectation. However, these analyses demonstrate that subjects did not tend to rely on such higher-order behavioral strategies; neuronal results sorted according to the four simple outcome categories defined above are therefore unlikely to be significantly biased by such factors.

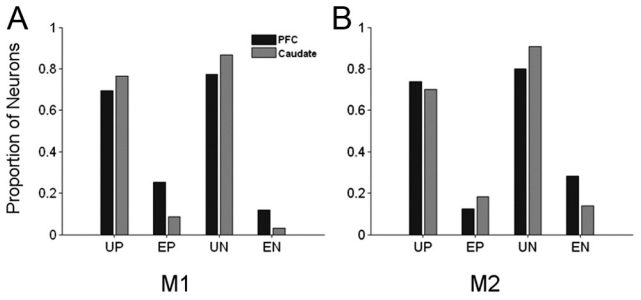

Figure 5. Prevalences and magnitudes of the four types of outcome-related responses. ***A***, This bar graph depicts the fraction of neurons within the lateral PFC (black) or caudate nucleus (gray) of subject M1 that exhibited each category of outcome-related response. A neuron was counted as responding to a particular outcome category if it had at least one time bin (after feedback onset) that was significant in the relevant ROC analysis (see Materials and Methods). Fractions sum to greater than one within each area because neurons could display more than one response type. Numerical values are included in Table 1. ***B***, Data from subject M2 displayed in the same manner as in **A**.

Table 1. Prevalences of neurons exhibiting significant responses to particular outcome types (related to Fig. 5)

	Subject M1 PFC ($n = 394$)	Subject M1 caudate ($n = 128$)	Subject M2 PFC ($n = 240$)	Subject M2 caudate ($n = 87$)
UP	274 (69.5%)	98 (76.6%)	177 (73.8%)	61 (70.1%)
UN	305 (77.4%)	111 (86.7%)	192 (80.0%)	79 (90.8%)
UP and UN	211 (53.6%)	85 (66.4%)	140 (58.3%)	56 (64.4%)
UP or UN	368 (93.4%)	124 (96.9%)	229 (95.4%)	84 (96.6%)
EP	100 (25.4%)	11 (8.6%)	30 (12.5%)	16 (18.4%)
EN	47 (11.9%)	4 (3.1%)	68 (28.3%)	12 (13.8%)
EP and EN	11 (2.8%)	1 (0.8%)	13 (5.4%)	6 (6.9%)
EP or EN	136 (34.5%)	14 (10.9%)	85 (35.4%)	22 (25.3%)

Numbers and associated percentages are shown separately for the PFC and caudate populations for subject M1 and M2. The percentages are based upon the total number of neurons in each anatomical area, in each animal. The "UP and UN" and "EP and EN" rows reflect the number of neurons with combined responses, and the associated fraction is with respect to the total number of neurons (i.e., this is not equal to the "overlap fraction" shown in Fig. 8 or Table 3). The "UP or UN" and "EP or EN" rows reflect the total number of neurons responding to at least some form of unexpected or expected outcome, respectively.

Neuronal population

We recorded 635 single neurons from the lateral PFC (M1 $n = 394$; M2 $n = 240$) and 233 neurons from the anterior caudate nucleus (M1 $n = 139$; M2 $n = 94$) of two monkeys while they performed this task (sample electrode trajectories are shown in Fig. 1*C*,*D*). In the caudate, we excluded the small number of cells that had firing characteristics typical of tonically active neurons (Aosaki et al., 1995) (11 in M1 and 7 in M2, representing 7.91% and 7.45% of all recorded caudate neurons in each animal). Our caudate population therefore consisted of neurons most likely to be striatal projection neurons, a.k.a. medium spiny neurons. We did not prescreen neurons in either area for task-related responses.

Neuronal representation of expected and unexpected positive and negative outcomes

For this report, we focused on neuronal activity beginning at the time of visual feedback onset, near the end of each trial. For each neuron in the PFC or caudate, we examined firing rates reflecting the immediate outcome in each trial (correct or incorrect) as a function of the prior trial's outcome.

A sliding ROC analysis was performed on the outcome-related activity of each neuron with a bootstrap control to determine significance (see Materials and Methods). Neurons were considered to code for a particular type of outcome if an ROC area comparing UP and EP trials or comparing UN and EN trials significantly differed from chance in any time bin during the 1000 ms beginning at feedback onset (the direction above or below chance indicated which of the pair more strongly activated the neuron). These particular pairings formed the basis of the ROC analysis because a comparison restricted to positive trial outcomes (UP or EP) was not confounded by immediate events, nor was a comparison of negative trial outcomes (UN or EN). Therefore, a significant ROC during the feedback period reflected a difference in expectation, as set by the previous trial's outcome, rather than a difference of visual cue or reward.

Consistent with the importance of unexpected events as motivators of learning, we found that many neurons in both the PFC and the caudate were preferentially activated by unexpected (UP or UN) outcomes (Fig. 4*A*,*B*). Fewer neurons were preferentially activated by expected (EP or EN) outcomes (Fig. 4*C*,*D*). In both the PFC and caudate populations, responses to unexpected outcomes were strikingly more common (Fig. 5; Table 1). In fact, >90% of neurons in each animal responded selectively to at least one of the two valences of unexpected outcome. Randomly allocating the total number of significant responses across the four outcome categories (UP, EP, UN, or EN) demonstrated that the probability the top two categories (UP and UN) would show the observed separation from the bottom two (EP and EN) simply by chance was very low ($p \ll 0.0001$ in each area in each subject).

Contribution of more distant outcomes to neuronal activity

To assess our method of relying on the outcome of only the immediately preceding trial as a determinant of neuronal activity, we analyzed the ability to discriminate spike rates as a function of the outcome of a given preceding trial in a sequence leading up to the current trial. The independent contribution of each preceding trial to the current trial's outcome was assessed with an ROC analysis that sorted trials according to the outcome on trial *n*-*k* (correct or incorrect), and compared the resulting distributions of spike rates in trial *n* using the area under the ROC, while holding the sequence of intervening trials constant (see Materials and Methods). In the context of our task, the immediately preceding trial (*n*-1) did indeed provide the most information about neuronal activity, with much less discriminability provided by trial *n*-2 and the other preceding trials (Fig. 6). This is consistent with the behavioral evidence, above, which demonstrated that the animals did not rely on higher-order, multiple-trial sequences of outcomes to guide behavior.

As a cross-validation of this result, a linear regression was performed to assess the contribution of the correctness of preceding trials to outcome-related neuronal activity, as well as the potential contribution of the chosen objects or spatial locations on those preceding trials (Fig. 7). We found that, using this method as well, the correctness of the immediately preceding trial was indeed the most potent factor driving outcome-related neuronal activity, with less-consistent and smaller contributions of earlier trials. The chosen objects or locations did not significantly contribute to outcome-related neuronal activity (regression coefficients were nearly uniformly <0.1, and none were significant).

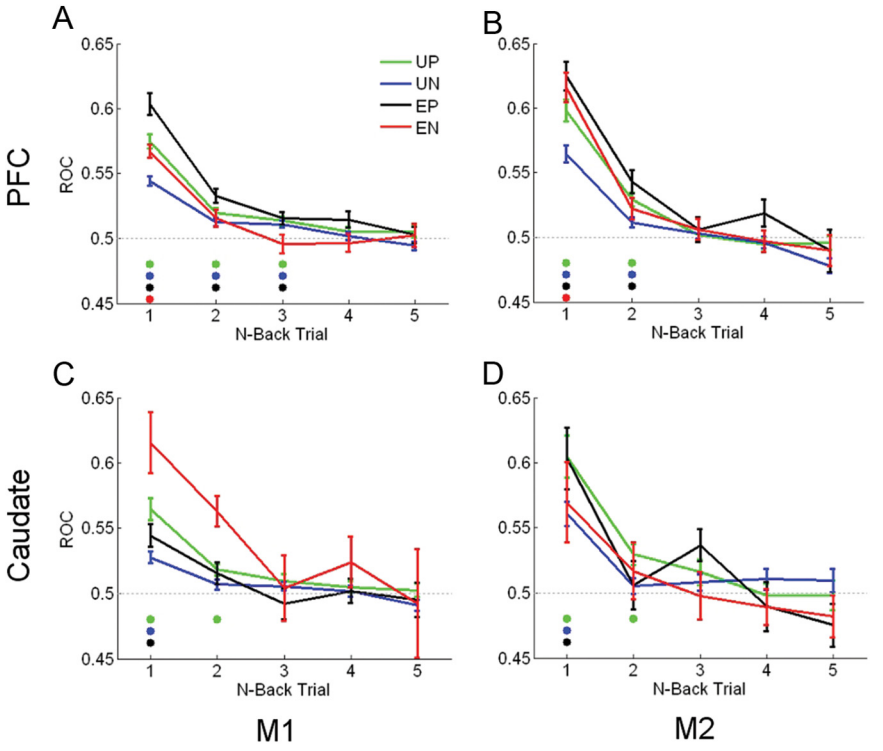

Figure 6. The independent contributions of preceding trials to spike rates. ***A***, The discriminability of spike rates by ROC in a given index trial, *n*, afforded by the outcome of a preceding trial (*n*-1 through *n*-5) is plotted (see Materials and Methods). ROCs for activity corresponding to each of the four categories of outcome are plotted separately. Here, the contributions of preceding trial outcomes to neuronal activity are shown for the PFC of subject M1. Most of the ability to discriminate spike rates was indeed found in the outcome of the immediately preceding trial. The colored dots identify which points were significantly above chance for the population. ***B*—*D***, The contribution of preceding trials to spiking activity in the PFC of M2 (***B***), the caudate of M1 (***C***), and the caudate of M2 (***D***) were similar. Note that it was possible that the neural responses to different outcome categories could have been differently influenced by preceding outcomes; however, after correcting for the number of pairwise comparisons, no such differences were found to be significant.

Lack of modulation of RPE-related activity by task context or chosen features

While neurons that exhibit RPE-related activity (i.e., UP- or UN-selective) could in principle also convey information about particular task features that led to the observed outcome, the linear regression described above did not find a significant influence of chosen object or location. To more directly test for the possible influence of these factors, we now explicitly considered only the most- or least-preferred objects or locations for each neuron, and also considered the task context (object learning or spatial learning).

We sorted spike rates for all UP- or UN-selective neurons (on UP or UN trials, respectively, and using only significantly modulated bins, as determined earlier) according to object versus spatial task, most- versus least-preferred object, or most- versus least-preferred direction. Two-tailed *t* tests were then applied to reveal significant differences, with *post hoc* correction for multiple comparisons (Benjamini and Hochberg, 1995). The results, shown in Table 2, demonstrate that RPE-related activity was not significantly modulated by any of these parameters, consistent with the notion that RPEs are generic signals reflecting only the overall quality of an outcome, regardless of specific task elements (Sutton and Barto, 1998).

Representation of the current versus previous outcome in the feedback period

Because a comparison of UP-versus-EP trials, or UN-versus-EN trials, is based upon a difference in the outcome of the previous trial, it was possible that feedback-period differences in a current trial simply reflected a difference in the prior trial's outcome, rather than being driven by the immediate outcome and its expectation. Indeed, many neurons (including those in Fig. 4*C*,*D*,*F*) did show significant ROCs earlier in a trial that must have reflected the prior trial's outcome, because the current trial's outcome was yet unknown. Therefore, to measure the degree to which activity was driven by the previous trial outcome rather than the current one, we calculated an index for each neuron that cast the differential contribution of the prior versus current trial as a ratio (see Materials and Methods). This index was positive when a neuron's activity was more closely related to the previous trial's outcome and negative when it was more a function of the current trial's outcome. As expected, the indices calculated before feedback onset tended to be greater than zero (M1 and M2: $p < 0.0001$ by two-tailed *t* test; Fig. 7), suggesting a larger contribution of the previous trial, while the indices calculated after feedback onset tended to be less than zero (M1 and M2: $p < 0.0001$), consistent with a greater contribution of the current trial's outcome. The ROC differences beginning at feedback onset, there-

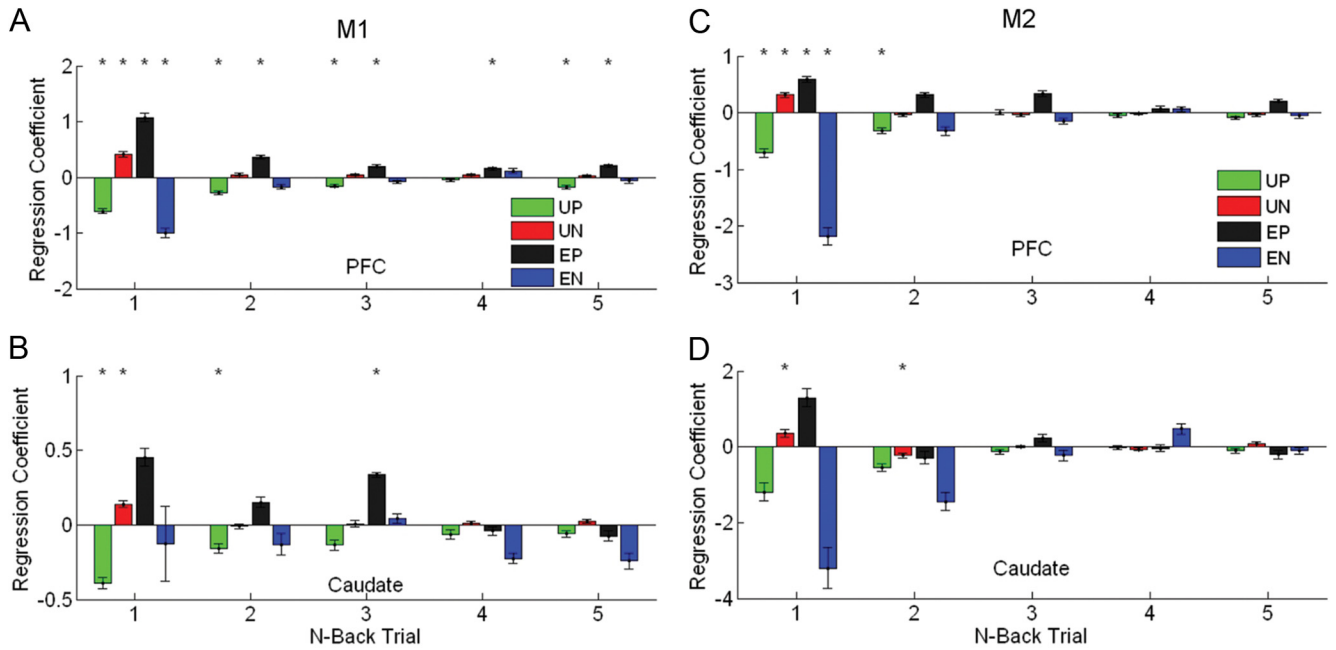

Figure 7. The contribution of behavioral factors on preceding trials to neuronal activity. The contribution of the correctness, the chosen object, and the chosen spatial location on preceding trials to outcome-related neuronal activity was assessed with multiple linear regression. In neither area and in neither animal were chosen object or location significant regressors on any preceding trial. Plotted here are the regression coefficients for each type of outcome-related response, as a function of the correctness of the *N*-back trial. ***A***, The data for PFC neurons in subject M1. Positive deflections indicate a positive contribution of a preceding correct trial at that lag, and negative deflections indicate a positive contribution of an incorrect trial at that lag. Therefore, UP and EN responses, which are by definition based upon a preceding incorrect trial, both deflect negatively; meanwhile UN and EP responses, which are identified based upon a preceding correct response, both deflect positively. In this animal, the UP and EN activity of PFC neurons tended to be significantly influenced by outcomes several trials into the past, but the largest contributor was the immediately preceding trial (*n*-1). ***B***, The data for caudate neurons in subject M1. ***C***, The data for PFC neurons in subject M2. The contribution of more distant trials was not replicated in this animal. ***D***, The data for caudate neurons in subject M2. In all cases, as for the independent ROC method (Fig. 6), the immediately preceding trial was indeed the most potent influence on outcome-related neuronal activity.

Table 2. RPE activity is not modulated by task, chosen object, or chosen location

	PFC		Caudate	
	UP cells	UN cells	UP cells	UN cells
Subject M1				
Object vs Spatial Task	**0 (0.0%)**	**0 (0.0%)**	**0 (0.0%)**	**0 (0.0%)**
	14 (5.1%)[a]	28 (9.2%)[b]	4 (4.1%)[c]	4 (3.6%)[d]
Best vs Worst Cue Object	**1 (0.4%)**	**0 (0.0%)**	**0 (0.0%)**	**0 (0.0%)**
	60 (21.9%)[a]	44 (14.4%)[b]	16 (16.3%)[c]	10 (9.0%)[d]
Best vs Worst Direction	**2 (0.7%)**	**0 (0.0%)**	**0 (0.0%)**	**0 (0.0%)**
	19 (6.9%)[a]	20 (6.6%)[b]	4 (4.1%)[c]	5 (4.5%)[d]
Subject M2				
Object vs Spatial Task	**1 (0.6%)**	**1 (0.5%)**	**0 (0.0%)**	**0 (0.0%)**
	7 (4.0%)[e]	8 (4.2%)[f]	1 (1.6%)[g]	2 (2.5%)[h]
Best vs Worst Cue Object	**0 (0.0%)**	**0 (0.0%)**	**4 (6.6%)**	**0 (0.0%)**
	29 (16.4%)[e]	29 (15.1%)[f]	15 (24.6%)[g]	8 (10.1%)[h]
Best vs Worst Direction	**2 (1.1%)**	**0 (0.0%)**	**0 (0.0%)**	**0 (0.0%)**
	22 (12.4%)[e]	19 (9.9%)[f]	4 (6.6%)[g]	2 (2.5%)[h]

The RPE activity of very few neurons was influenced by the task (object vs spatial learning), the chosen object, or the chosen location. Neurons with significant UP or UN activity were assessed for significant modulation of this activity across tasks or most- versus least-preferred ("best versus worst") features using two-tailed *t* test. The values in bold reflect the numbers and percentages of neurons showing significant differences after correction for multiple comparisons, using $\alpha = 0.05$. The numbers immediately beneath these, within each cell, are the uncorrected values, using the same alpha level. The results for subject M1 and for M2 are shown separately.

[a]($n = 274$).

[b]($n = 305$).

[c]($n = 98$).

[d]($n = 111$).

[e]($n = 177$).

[f]($n = 192$).

[g]($n = 61$).

[h]($n = 79$).

fore, were not simply the persistence of pre-feedback activity reflecting the previous trial's outcome, but rather reflected the current outcome as a function of the expectation set by the prior outcome.

Overlap of outcome representations in single neurons

Surprisingly, the outcome-related activity of individual neurons did not exclusively adhere to one particular category of response (Fig. 4*E–G*): Many individual neurons were observed to have separate epochs within a trial exhibiting different types of outcome-related activity. We wondered whether particular types of outcome-related activity were more or less likely to be found in conjunction with specific other types of activity within single neurons. Specifically, we were interested to know whether UP and UN responses were derived from a common substrate reflecting unexpected events, and were therefore more likely to be observed together within individual neurons. Therefore, we calculated the percentage overlap for each of the six possible pairings of the four different outcome categories (UP, UN, EP, or EN), where overlap was defined as the number of instances in which a given pair is found within single neurons divided by the sum of that number plus the total number of instances in which either member of the pair is found in isolation from the other. Thus, if each of the types within a pair never overlapped with the other within single neurons, the overlap fraction would be zero; if, on the other hand, they only ever were observed as paired within single neurons, the overlap fraction would be one. A bootstrap analysis was used to determine whether the observed incidence of pairings was more or less likely than expected by chance, given the underlying distributions (see Materials and Methods).

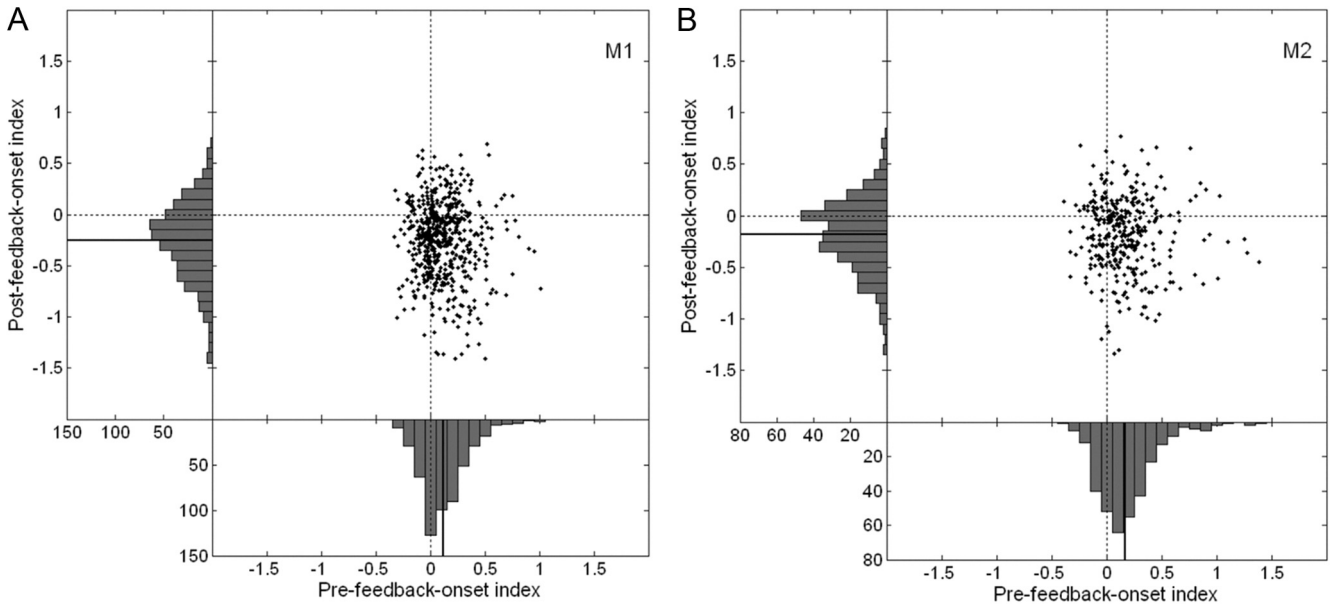

Figure 8. Relative contributions of the prior versus the current trial's outcome on neuronal activity. ***A***, A positive index reflected neuronal activity that was driven primarily by the prior trial's outcome, whereas a negative index reflected activity that was more related to the current trial's outcome (see Materials and Methods). Here, the data from subject M1 is shown. A scatter plot is shown depicting the mean pre-feedback-onset and post-feedback-onset indices for each neuron. Population histograms depicting these preindices and postindices are shown along the *x*- and *y*-axes, respectively, with the mean of each distribution depicted by a solid black line. The means of these distributions were 0.12 and −0.26, respectively, each differing significantly from zero ($p < 0.00001$ in each case, by two-tailed *t* test). ***B***, The data from subject M2 is plotted as in ***A***. The means of the pre-feedback-onset and post-feedback-onset index distributions were 0.17 and −0.18, respectively. These also significantly differed from zero ($p < 0.00001$ in each case). Splitting the data to consider neurons recorded from either the PFC alone or from the caudate alone (independently for each subject) did not change the pattern or significance of these results.

This analysis revealed that in neither the PFC nor the caudate of either animal was the frequency of UP-UN association within individual neurons higher or lower than predicted by their individual incidences across the population ($p > 0.2$ in all cases; Fig. 8, Table 3). This suggests that information about unexpectedly positive events and unexpectedly negative events was independently distributed across each neuronal population.

Relative timing of UP and UN activity

Because the time of appearance of UP or UN selectivity might provide insight into the underlying circuit mechanisms that generate these signals, we calculated the latencies to the appearance of significant UP or UN activity, defined as the earliest time bin during the feedback period with a significant ROC for each type of activity (see Materials and Methods). Next, we plotted the cumulative latencies to the appearance of UP and UN activity for each recorded area in each animal (Fig. 9). Significant differences between these latency curves were then assessed with a Kolmogorov–Smirnov (KS) test. We found that, in the PFC of one animal and the caudate of both animals, UP activity did appear with a significantly shorter latency than UN activity (KS test comparing distributions across neurons of latencies to first significant bins: M1 PFC: $p < 0.001$; M1 caudate: $p = 0.003$; M2 PFC: $p = 0.232$; M2 caudate: $p = 0.033$), consistent with the notion that extra processing may be required to detect or invert a negative signal relaying a negative prediction error. No difference in the latency of appearance of UP or UN activity was observed between the PFC and caudate ($p > 0.3$ in all four cases).

Trial-by-trial variations in single neuron activity and behavioral strategy

We next sought to determine whether the activity of the UP- and UN-selective neurons—representing non-zero reward prediction errors—was linked to subsequent behavioral decisions. Op-

Table 3. Numerical overlap fractions and associated *p* values (related to Figure 8)

	UP	EP	UN	EN
(A) Subject M1 PFC				
UP	**1**			
EP	**0.16 (0)***,#	**1**		
UN	**0.57 (0.336)**	**0.24 (0.611)**	**1**	
EN	**0.13 (0.901)**	**0.08 (0.306)**	**0.06 (0)***,#	**1**
(B) Subject M1 caudate				
UP	1			
EP	**0.07** (0.095)	**1**		
UN	**0.69** (0.401)	**0.08** (0.158)	**1**	
EN	**0.03** (0.234)	**0.07** (0.663)	**0.02** (0.013)*,#	**1**
(C) Subject M2 PFC				
UP	**1**			
EP	**0.05** (0)*,#	**1**		
UN	**0.61 (0.222)**	**0.13 (0.590)**	**1**	
EN	**0.26 (0.504)**	**0.15 (0.045)****	**0.22 (0.006)***,#	**1**
(D) Subject M2 caudate				
UP	**1**			
EP	**0.12 (0.013)***,#	**1**		
UN	**0.67 (0.557)**	**0.16 (0.039)***	**1**	
EN	**0.16 (0.757)**	**0.27 (0.006)****,#	**0.12 (0.072)**	**1**

(A) Subject M1 PFC: The overlap fractions across pairs of outcome-type representations in single neurons in the PFC of subject M1 are shown in bold. The bootstrapped *p* value for each is shown adjacent, in parenthesis.

(B) Subject M1 caudate: The overlap fractions for caudate neurons in subject M1, as shown for PFC neurons in (A).

(C) Subject M2 PFC: The overlap fractions for PFC neurons in subject M2.

<mc>(D) Subject M2 caudate: The overlap fractions for caudate neurons in subject M2. In subject M2 there was an unreplicated trend for expected responses of the same valence (EP and EN) to overlap more often than expected by chance. The decreased frequency of overlap for UP-EP and UN-EN pairings is likely an artifact of the analysis structure (see Materials and Methods).

*Overlap fractions that were significantly below expected ($\alpha = 0.05$), given the independent prevalence of each type of outcome representation. **Overlap fractions that were higher than expected. #*p* Values that remained significant after correction for multiple comparisons; these are also indicated by color (green for less frequent than expected by chance, red for more frequent than expected).

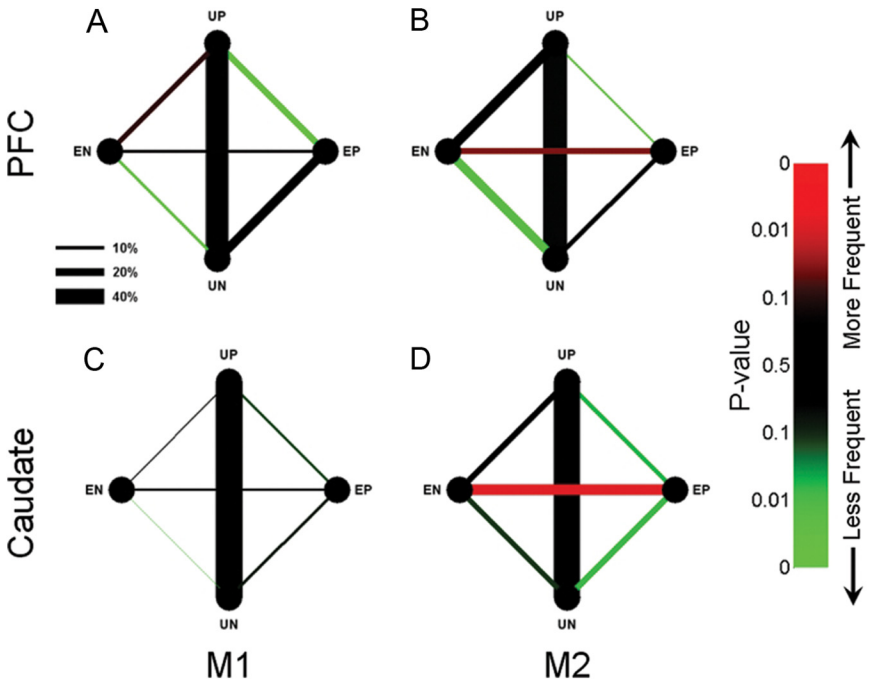

Figure 9. Overlap fractions between pairs of outcome types. The overlap fraction between each pair is shown by the thickness of the lines connecting them (as indicated by the thickness scale and numerical values in Table 3). The significance of these values, factoring in the baseline frequencies of each outcome type, is depicted by the color of the lines. ***A***, The overlap fractions for outcomes in the PFC of M1. ***B***, The overlap fractions for the outcomes in the PFC of M2. ***C***, The overlap fractions for the outcomes in the caudate of M1. ***D***, The overlap fractions for the outcomes in the caudate of M2. There was no evidence for an increased or decreased tendency for UP- or UN-selective responses to be found in the same neurons. Note that the decreased tendency for UP-EP and UN-EN pairings is a consequence of the analysis structure (see Materials and Methods).

timally, a UP outcome on a given trial should lead to reselection of the same choice on the following trial, whereas a UN outcome should lead to selection of a new choice on the next trial. Because directly comparing behavior against neuronal activity in response to UP versus UN outcomes would be confounded by differences in the feedback cue and reward delivery, a different within-outcome-category measure was needed. Therefore, we examined trial-by-trial variations of neuronal activity in response to a single type of outcome (UP or UN) in relation to subsequent behavioral choice. We hypothesized that these trial-by-trial fluctuations in outcome-related neuronal activity could be linked to the animals' subsequent behavioral strategy, such that higher UP or UN activity would be associated with more optimal response selection. For example, if an animal responded appropriately to a UP outcome in a given index trial by reselecting that correct response on the subsequent trial, outcome-related activity in the index trial should have been higher than in those instances when an UP outcome was followed less appropriately by selecting a different response. Likewise, higher UN-related activity should be associated with subsequent selection of a different response, whereas lower UN-related activity might be associated with suboptimal selection of the same incorrect response on the next trial.

Specifically, we assessed neuronal activity as a function of whether the next trial's chosen direction was the same or different as in the index (UP or UN) trial. We used the earliest trials in each block when a spatial strategy was most evident. For each UP- or UN-activated neuron, we subtracted the average activity associated with a subsequent "same" response from the average activity associated with a subsequent "new" (different) response. If the prediction holds, the population of UP activity differences should be shifted to the left (higher activity for "same" direction trials), whereas the population of UN response differences should be shifted to the right (higher activity for "new" direction trials). Indeed, over the respective neuronal populations (UP- and UN-selective neurons), we found this prediction to be true in the PFC, but not the caudate (UP PFC: $p = 0.024$; UN PFC: $p = 0.010$; UP and UN caudate: $p > 0.1$, by t test), although the absolute magnitude of this effect was quite small (Figs. 10, 11). While this difference may reflect an inherently weaker link between caudate RPEs and subsequent behavioral choice—similar to the lack of tight correlation observed between midbrain dopamine neuron RPEs and behavior (Bayer and Glimcher, 2005)—it may simply reflect a lack of sufficient statistical power in the smaller caudate dataset to replicate the effect seen in the PFC.

Performing this same analysis assuming an object-based, rather than spatially based, strategy revealed no significant link between trial-by-trial neuronal activity fluctuations and subsequent choice behavior (UP PFC: $p = 0.992$; UN PFC: $p = 0.459$; UP caudate: $p = 0.052$; UN caudate: $p = 0.368$). This is unsurprising, given the clear reliance on a spatial strategy early within each block (Fig. 2*C*,*D*).

An average of 53 trials in subject M1 and 44 trials in M2 was used for each neuron for this analysis. Examining the activity fluctuations of individual neurons for subsequent significant choice-predicting behavior revealed no significant correlation for any single cell, after correction for multiple comparisons. The subtle link between neuronal activity fluctuations and subsequent behavior was evident only across the entire population of neurons.

Discussion

We observed clear evidence of positive and negative RPEs in a large proportion of lateral PFC and caudate neurons, whereas many prior studies had not seen strong evidence of RPEs in these structures (Matsumoto et al., 2007; Kennerley and Wallis, 2009; Kim et al., 2009; Roesch et al., 2009; Oyama et al., 2010). The influence of outcomes, regardless of expectation, has been described in the lateral PFC and caudate (Histed et al., 2009). Other work has provided evidence of RPEs in the lateral PFC, but did not specifically consider differences between positive and negative RPEs (Seo et al., 2007). Meanwhile, neurons in the striatum have been described as encoding reward expectation, but specifically not the deviation of actual reward from expectation, that is, the RPE (Schultz et al., 1998). That some studies failed to see evidence of RPEs may be due at least in part to the types of behavioral paradigms used. Ours was a temporally delayed, on-line instrumental learning task with multiple reversals, requiring selective visual attention to process the peripherally presented cues; any or all of these components may have been essential to achieving strong neuronal activation. Indeed, some forms of short-term memory and reversal learning may depend critically upon the lateral PFC (Fuster and Alexander, 1971; Kubota and Niki, 1971; Kojima and Goldman-Rakic, 1982; Funahashi et al.,

1993; Dias et al., 1996; Asaad et al., 1998), and therefore tasks that engage these features might be expected to produce greater and more selective neuronal activation.

It has been a matter of some debate whether positive and negative prediction errors are processed by separate, parallel neural systems, or by the same circuit. Several human functional imaging studies have suggested that positive and negative outcomes might be processed in separate brain regions (Ramnani et al., 2004; Nieuwenhuis et al., 2005; Yacubian et al., 2006; Liu et al., 2007; Seymour et al., 2007). Others did find overlap, but observed opposite responses to positive versus negative RPEs (McClure et al., 2003; D'Ardenne et al., 2008). In addition, several imaging studies report evidence for RPEs in the ventral, but not dorsal striatum (O'Doherty et al., 2004; Abler et al., 2006; Hare et al., 2008), contributing to the notion that the dorsal striatum may be the "actor" whereas the ventral striatum is the "critic," such as in actor-critic models of reinforcement learning. In contrast to these reports, our single-unit data demonstrate that both types of signal do indeed coexist—even at the single-neuron-level—in both the lateral PFC and caudate nucleus of the nonhuman primate, and that a positive representation predominates for both positive and negative RPEs.

Recent work has demonstrated that the dorsal anterior cingulate cortex also exhibits positive signals for both positive and negative unexpected events (Hayden et al., 2011). Combined with our work, this suggests increased activation for both positive and negative RPEs may be a general phenomenon across the corticostriatal system. Those results were attributed to a "surprise" signal. Similarly, it may be tempting to construe our findings as reflecting a more general surprise signal related to the occurrence of any unexpected event. However, the data show that in the lateral PFC and caudate, at least, positive and negative unexpected events had independent, though sometimes overlapping, representations; a common substrate, such as surprise, should have caused these signals to coincide more often than expected by chance, but this was not observed. Outputs from these UP- and UN-driven neurons may converge onto downstream targets to provide a more generic surprise signal, but their independence in the lateral PFC and caudate would facilitate the contrasting influence errors of opposite valence should have on behavior. In other words, while the purpose of a combined surprise signal may be to increase arousal or attention generally,

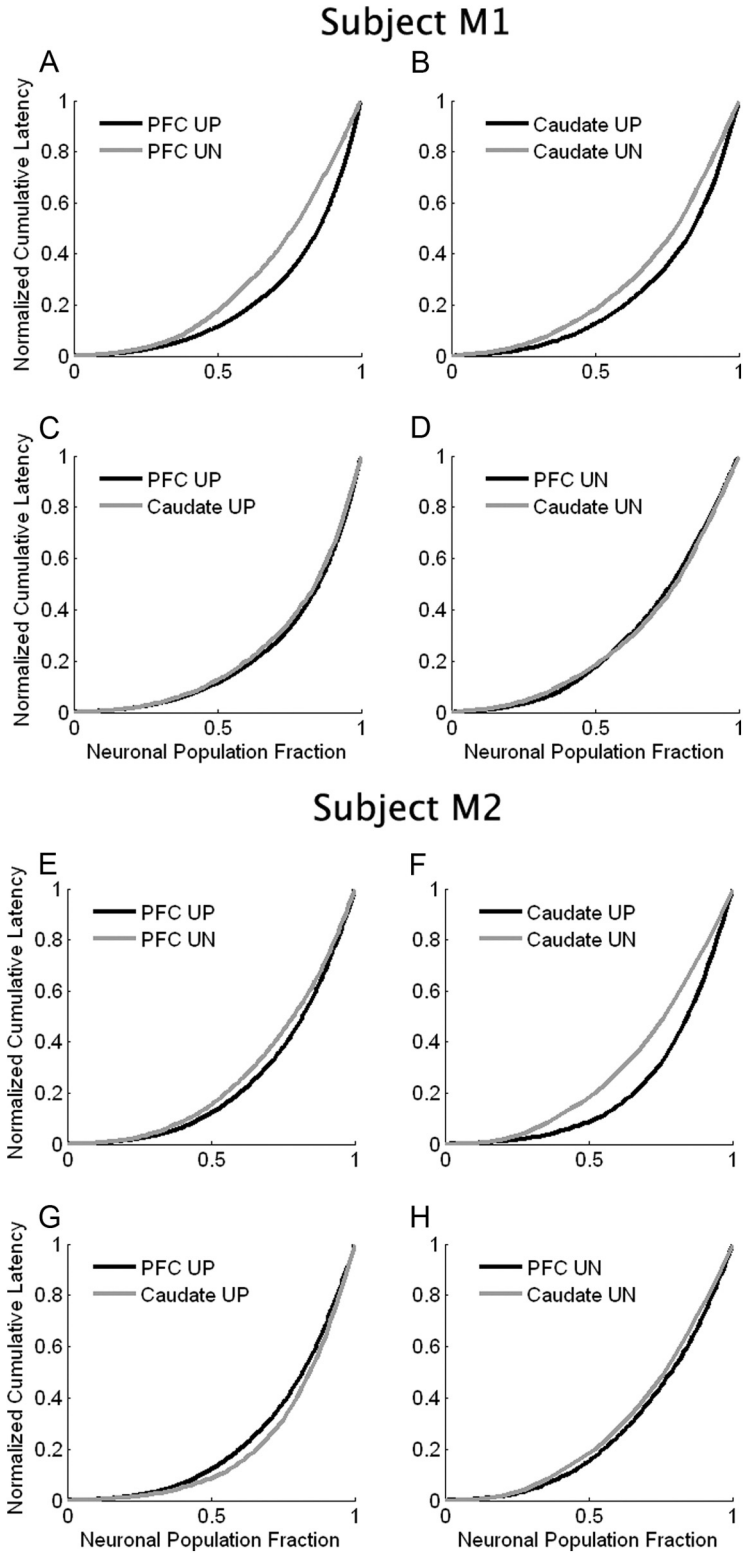

Figure 10. Latencies to the appearance of UP or UN activity. ***A***, The cumulative latencies to the first time bin during the feedback period at which significant UP or UN activity was detected are plotted for the PFC neuronal population in subject M1 (see Materials and Methods for details). Lower values indicate lower latencies. UP activity tended to arrive earlier than UN activity ($p < 0.001$). ***B***, The cumulative latencies for UP and UN activity in the caudate of M1 are plotted. Again, UP activity tended to arrive earlier than UN activity ($p = 0.003$). ***C***, No difference between the latencies of arrival of UP activity across the PFC and caudate were observed ($p = 0.700$). ***D***, No difference was observed between the latencies of arrival of UN activity across the PFC and caudate ($p = 0.638$). ***E–H***, The same data are shown for subject M2. A significant difference was observed between the latencies for UP and UN activity in the caudate (***F***: $p = 0.033$), but not in the PFC (***E***: $p = 0.232$). Again, there were no differences observed across areas (***G***: $p = 0.374$ for UP activity; ***H***: $p = 0.602$ for UN activity).

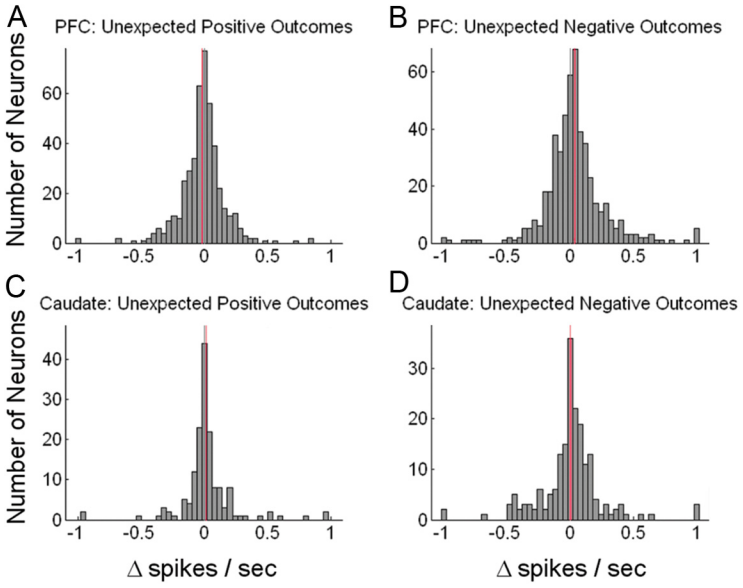

Figure 11. Association between neuronal activity and subsequent behavioral choice. ***A***, This histogram plots the difference in PFC neuronal activity (Δ spikes/s) between UP trials that were followed by a new behavioral choice (i.e., chosen direction) minus UP trials followed by the same behavioral choice (see Materials and Methods for details). Data from both subjects were included. The *y*-axis shows the number of neurons with the specified difference in neuronal activity indicated on the *x*-axis. If the activity of these UP-selective neurons was linked to behavior, greater activity on UP trials should more often lead to appropriate repetition of the same behavioral response, and so the plotted differences should generally be negative across the population. The slight leftward shift here (UP-then-new < UP-then-same) is consistent with this notion. The red line is the population mean ($p = 0.024$ by two-tailed *t* test). ***B***, Here, the difference between UN trials followed by a new behavioral choice minus UN trials followed by the same choice is plotted for neuronal activity in the PFC of both subjects. Greater activity on UN trials should be associated with the more appropriate selection of a new choice on the subsequent trial. The population's small rightward shift here (UN-then-new > UN-then-same) is consistent with this idea ($p = 0.010$ by two-tailed *t* test). ***C***, ***D***, There was no significant association between the level of neuronal activity on UP or UN trials in the caudate and subsequent behavioral choice, as shown in ***C*** and ***D***, respectively ($p > 0.5$ in both cases).

the particular behavioral modification required by a positive or negative RPE is distinct: the former should reinforce the action or strategy just undertaken, whereas the latter should drive behavior in an alternate direction.

Indeed, we observed a small but significant tendency for fluctuations of UP or UN activity in the PFC to be differentially linked to upcoming choice behavior (namely, repeating the same vs selecting a new option, respectively). This analysis was based on the notion that if the observed neuronal activity were relevant to decision making, even small trial-by-trial variations in single-neuron firing rates might be reflected in gross behavior. Simply assessing the effects of positive versus negative RPE activity on subsequent choice would have been confounded by the visual feedback stimuli and reward that differentiated these conditions. For this reason, examining the subtle variations within a condition (UP or UN, not UP vs UN) was necessary to avoid that pitfall. The magnitude of the observed effect, therefore, is of less interest than its mere presence as a marker of the link between neuronal activity and subsequent behavior. Furthermore, it is possible that the strength of the correlation between neuronal activity and subsequent choice depended on the stage-of-learning; because our assessment of this correlation was limited to the earliest trials within a block, where one particular strategy predominated, a stronger (or weaker) correlation in subsequent trials may have gone undetected. Nevertheless, that we were able to observe such a link between the "noisy" fluctuations of individual neurons and gross behavior was remarkable, and may imply a more robust correlation between outcome-related activity and consequent decisions more generally.

Our results concur with previous work by others that found immediately preceding outcomes were the most potent drivers of neuronal activity (Seo et al., 2007). This is unsurprising given the behavioral strategies used by the animals: they tended to repeat a spatial response that had led to an unexpected positive outcome, and unexpected negative outcomes led to a "guessing" strategy which did not accumulate counterfactual information from particular incorrect choices. Therefore, this simplified, local behavioral strategy is compatible with the short-sightedness of the neuronal representations (Fig. 6).

We did not observe an influence of task context (object- or spatial-learning) or specific task features (chosen object or chosen location) on the outcome-related activity of lateral PFC or caudate neurons. This conforms to the reinforcement learning notion that the RPE would be insensitive to particular behavioral details, but rather should serve to convey a broader evaluation of the general success of ongoing behavior.

The absence of dopamine may signal negative RPEs (Hollerman and Schultz, 1998; Frank et al., 2007). Consistent with this notion, patients with depleted levels of dopamine due to Parkinson's disease showed a bias to learn relatively more from negative RPEs rather than from positive RPEs; this bias was reversed when dopamine was repleted with medication (Frank et al., 2004). However, the limited range of less-than-baseline activity has led some to argue that dopamine may be a less potent messenger for unexpectedly negative events. Some recent work has found there may be midbrain dopamine neurons that, rather than being inhibited by negative events, increase firing rate in response to both positive and negative outcomes (Joshua et al., 2008; Matsumoto and Hikosaka, 2009). In addition, there may be other ascending neuromodulatory systems that complement the dopaminergic system by preferentially signaling unexpectedly negative outcomes (Daw et al., 2002).

We found that the latency to the appearance of positive RPEs tended to be shorter than the latency for negative RPEs in both the PFC (of one subject) and caudate nucleus (of both subjects). This is consistent with an extra processing step required to invert a negative signal, perhaps such as that provided by lower-than-baseline dopaminergic activity. Or it may reflect extra time needed to detect this inhibition, as the absence of a few spikes relative to low baseline activity is less immediately evident than the presence of a phasic burst.

The similarities between the representation of RPEs by the lateral PFC and the caudate were striking. Only one prior study had directly compared outcome-related activity across these two structures in the nonhuman primate (Histed et al., 2009); in that case, caudate neurons on average conveyed more information about the absolute correctness of an outcome, but no analysis of RPEs (correctness as a function of expectation) was reported. Our results demonstrate that RPE prevalence (Fig. 5), magnitude

(Fig. 6), and timing (Fig. 8) were essentially indistinguishable across these structures.

A negative RPE reflecting the absence of reward may not have the same neural substrate as a negative RPE reflecting the presence of an aversive outcome. To what extent this distinction has influenced the heterogeneity of results reported in the literature is difficult to disentangle from the wide variety of tasks and methodologies used. In our behavioral paradigm, a negative RPE corresponded to the unexpected absence of an appetitive stimulus. Further experiments, using both appetitive and aversive learning, would be required to determine whether an aversive stimulus produces similar results in the lateral PFC and caudate nucleus, and whether aversive stimuli engage different or additional circuits.

In summary, we found that most neurons within the lateral PFC and caudate nucleus were significantly modulated by trial outcomes, and particularly by unexpected outcomes. Both positive and negative unexpected outcomes were represented primarily by increases in firing rates, and these, in-turn, were linked to behavior in a predictable fashion that would facilitate learning. The delayed appearance of negative RPEs relative to positive RPEs provides clues about the underlying circuitry that conveys these critical signals to drive learning.

References

Abler B, Walter H, Erk S, Kammerer H, Spitzer M (2006) Prediction error as a linear function of reward probability is coded in human nucleus accumbens. Neuroimage 31:790–795.

Aosaki T, Kimura M, Graybiel AM (1995) Temporal and spatial characteristics of tonically active neurons of the primate's striatum. J Neurophysiol 73:1234–1252.

Asaad WF, Eskandar EN (2008) A flexible software tool for temporally-precise behavioral control in Matlab. J Neurosci Methods 174:245–258.

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

Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature 437:1158–1161.

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

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129–141.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B 57:289–300.

Brasted PJ, Wise SP (2004) Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. Eur J Neurosci 19:721–740.

Canales JJ, Capper-Loup C, Hu D, Choe ES, Upadhyay U, Graybiel AM (2002) Shifts in striatal responsivity evoked by chronic stimulation of dopamine and glutamate systems. Brain 125:2353–2363.

D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319:1264–1267.

Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. Neural Netw 15:603–616.

Dias R, Robbins TW, Roberts AC (1996) Dissociation in prefrontal cortex of affective and attentional shifts. Nature 380:69–72.

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943.

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A 104:16311–16316.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1993) Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas". J Neurosci 13:1479–1497.

Fusi S, Asaad WF, Miller EK, Wang XJ (2007) A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. Neuron 54:319–333.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J Neurosci 28:5623–5630.

Hayden BY, Heilbronner SR, Pearson JM, Platt ML (2011) Surprise signals in the anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci 31:4178–4187.

Histed MH, Pasupathy A, Miller EK (2009) Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. Neuron 63:244–253.

Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1:304–309.

Ichihara-Takeda S, Funahashi S (2006) Reward-period activity in primate dorsolateral prefrontal and orbitofrontal neurons is affected by reward schedules. J Cogn Neurosci 18:212–226.

Joshua M, Adler A, Mitelman R, Vaadia E, Bergman H (2008) Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. J Neurosci 28:11673–11684.

Kennerley SW, Wallis JD (2009) Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. Eur J Neurosci 29:2061–2073.

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940–947.

Kim H, Sul JH, Huh N, Lee D, Jung MW (2009) Role of striatum in updating values of chosen actions. J Neurosci 29:14701–14712.

Kimchi EY, Torregrossa MM, Taylor JR, Laubach M (2009) Neuronal correlates of instrumental learning in the dorsal striatum. J Neurophysiol 102:475–489.

Kojima S, Goldman-Rakic PS (1982) Delay-related activity of prefrontal neurons in rhesus monkeys performing delayed response. Brain Res 248:43–49.

Kubota K, Niki H (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. J Neurophysiol 34:337–347.

Liu X, Powell DK, Wang H, Gold BT, Corbly CR, Joseph JE (2007) Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. J Neurosci 27:4587–4597.

Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459:837–841.

Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci 10:647–656.

McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. Neuron 38:339–346.

Nieuwenhuis S, Slagter HA, von Geusau NJ, Heslenfeld DJ, Holroyd CB (2005) Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. Eur J Neurosci 21:3161–3168.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452–454.

Oyama K, Hernádi I, Iijima T, Tsutsui K (2010) Reward prediction error coding in dorsal striatal neurons. J Neurosci 30:11447–11457.

Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 433:873–876.

Paxinos G, Xu-Feng H, Toga AW (2000) The rhesus monkey brain in stereotaxic coordinates. San Diego: Academic.

Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532–552.

Rainer G, Asaad WF, Miller EK (1998) Memory fields of neurons in the primate prefrontal cortex. Proc Natl Acad Sci U S A 95:15008–15013.

Ramnani N, Elliott R, Athwal BS, Passingham RE (2004) Prediction error for free monetary reward in the human prefrontal cortex. Neuroimage 23:777–786.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: Classical conditioning 2: current research and theory (Black AH, Prokasy WF, eds), pp 64–69. New York: Appleton Century-Crofts.

Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. Nature 413:67–70.

Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G (2009) Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. J Neurosci 29:13365–13376.

Schmitzer-Torbert N, Redish AD (2004) Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. J Neurophysiol 91:2259–2272.

Schultz W, Apicella P, Ljungberg T, Romo R, Scarnati E (1993) Reward-related activity in the monkey striatum and substantia nigra. Prog Brain Res 99:227–235.

Schultz W, Tremblay L, Hollerman JR (1998) Reward prediction in primate basal ganglia and frontal cortex. Neuropharmacology 37:421–429.

Seo H, Barraclough DJ, Lee D (2007) Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. Cereb Cortex 17 [Suppl 1]:i110–i117.

Seymour B, Daw N, Dayan P, Singer T, Dolan R (2007) Differential encoding of losses and gains in the human striatum. J Neurosci 27: 4826–4831.

Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304:1782–1787.

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.

Tremblay L, Hollerman JR, Schultz W (1998) Modifications of reward expectation-related neuronal activity during learning in primate striatum. J Neurophysiol 80:964–977.

Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K (2009) Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 324:1080–1084.

Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43–48.

Williams ZM, Eskandar EN (2006) Selective enhancement of associative learning by microstimulation of the anterior caudate. Nat Neurosci 9:562–568.

Yacubian J, Gläscher J, Schroeder K, Sommer T, Braus DF, Büchel C (2006) Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. J Neurosci 26:9530–9537.