Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex

Christopher H Donahue1,4 & Daeyeol Lee1–3

Neurons in the dorsolateral prefrontal cortex (DLPFC) encode a diverse array of sensory and mnemonic signals, but little is known about how this information is dynamically routed during decision making. We analyzed the neuronal activity in the DLPFC of monkeys performing a probabilistic reversal task where information about the probability and magnitude of reward was provided by the target color and numerical cues, respectively. The location of the target of a given color was randomized across trials and therefore was not relevant for subsequent choices. DLPFC neurons encoded signals related to both task-relevant and irrelevant features, but only task-relevant mnemonic signals were encoded congruently with choice signals. Furthermore, only the task-relevant signals related to previous events were more robustly encoded following rewarded outcomes. Thus, multiple types of neural signals are flexibly routed in the DLPFC so as to favor actions that maximize reward.

Many studies have emphasized the role of the DLPFC in working memory1–6, but neurons in DLPFC also encode a highly heterogeneous set of signals, reflecting the stimulus features or cognitive processes related to the particular tasks performed by the animals7,8. DLPFC neurons can encode both the attended and remembered locations9,10, as well as experienced or expected outcomes11–13, temporally discounted values14,15 and other properties of rewards, such as their probabilities and magnitudes and the effort necessary to acquire them16. Although the heterogeneity of signals present in the DLPFC suggests that it may have an important function in integrating different sources of information to aid in decision making, this region has also been shown to encode information that is irrelevant for the tasks that animals are trained to perform17,18. This poses a problem as to how signals that are most relevant in a particular context might gain priority over other, less-relevant signals.

We trained monkeys on a probabilistic reversal task where they had to estimate the reward probabilities on the basis of their recent experience and combine them with magnitude information that varied independently across trials. Reward probabilities were associated with the colors of the targets whose spatial positions were randomized across trials. Therefore, the animals had to associate previous outcomes with the colors of chosen targets and ignore their spatial locations. Indeed, the monkeys showed a strong tendency to choose the target colors rewarded previously. We also found that reward in a given trial enhanced the encoding of task-relevant information in the immediately succeeding trial. In addition, only the information about the task-relevant events in the previous trial was encoded congruently with the choice signals in the DLPFC. In contrast, reward did not influence the encoding of task-irrelevant signals. Thus, task-relevant and irrelevant signals in the DLPFC might be differentially integrated such that task-relevant information can bias behavior toward the selection of actions that maximize reward.

RESULTS
Behavioral effects of rewarded and unrewarded outcomes

The animals were trained on a probabilistic reversal task in which the magnitude of reward available from each target was varied independently across trials (Fig. 1a; see Online Methods). After achieving fixation on the central target, the monkeys were presented with a red target and a green target (target period). One of the target colors was associated with a high reward probability (80%), and the other was associated with a low reward probability (20%). After an additional delay (0.5 s), small yellow tokens appeared around each target to indicate the magnitude of the available reward (magnitude period). The animals were then required to shift their gaze toward one of the targets. The reward probabilities for red and green targets were fixed within a block and underwent unsignaled reversals so that the animals had to estimate them through experience.

The task required the animals to associate the colors of previously chosen targets (relevant information) with their outcomes and ignore previously chosen locations (irrelevant information), as target locations were randomized across trials. To examine how the animals were influenced by task-relevant and task-irrelevant information, we examined the win-stay and lose-switch behavior in a color and spatial reference frame (Fig. 1b). We found that in 117/118 (99.2%) of the sessions, there was a significant difference between win-stay and lose-switch behavior in terms of chosen colors ($\chi^2$-tests, $P < 0.05$). In contrast, none of the sessions showed a significant influence of previously chosen location. These results indicate that the animals correctly used task-relevant information from previous trials to make their choices.

To determine how the monkeys integrated all of the information required to make their decisions, we compared a series of reinforcement learning models19 that addressed three important issues (Supplementary Table 1 and Online Methods). First, we examined how rewarded and unrewarded outcomes differentially affected

1Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut, USA. 2Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut, USA. 3Department of Psychology, Yale University, New Haven, Connecticut, USA. 4Present address: Gladstone Institute of Neurological Disease, San Francisco, California, USA. Correspondence should be addressed to D.L. (daeyeol.lee@yale.edu).

Received 10 October 2014; accepted 5 December 2014; published online 12 January 2015; doi:10.1038/nn.3918
learning on subsequent trials. Second, we examined whether the animals used information about the anticorrelated structure of reward probability reversals in this task. Finally, we examined whether the animals combined probability and magnitude information additively or multiplicatively. To systematically address these issues, we explored all possible models where these three factors were independently manipulated. The result was consistent for both animals. We found that the models with two learning rates in which target values were updated reciprocally (coupled) were superior in both monkeys. In both monkeys, the learning rates associated with rewarded trials was significantly larger than the learning rates associated with unrewarded trials (monkey O, $\alpha = 0.65$ and 0.05; monkey U, $\alpha = 0.64$ and 0.04, for rewarded and unrewarded outcomes, respectively, paired $t$-test, $P < 10^{-35}$). These results suggest that the animals updated their value functions more after rewarded trials than they did after unrewarded trials and that this updating took into account the anticorrelated structure of reward probability reversals. We also found that the additive model provided a better fit to the data than the model in which they were combined multiplicatively in every session for both monkeys. Thus, both animals might have relied more on a suboptimal additive strategy rather than on calculating expected values. When the performance of this additive model was evaluated using a leave-one-session-out cross-validation, the animal’s choice could be predicted correctly in 91.0 ± 2.3% (s.d.) of the trials. For comparison, the clairvoyant model, in which the animal always chooses the target with the higher expected value, correctly predicted the animal’s choices in 79.8 ± 2.8% of the trials.

**DLPFC activity related to multiple features**

Single-unit activity was recorded from a total of 226 neurons in the DLPFC of both monkeys (77 neurons in monkey O; 149 neurons in monkey U). We used a multiple linear regression model (Online Methods) to characterize neural signals related to various events in current and previous trials and to identify how task-relevant (i.e., previously chosen color, previous outcome) and task-irrelevant features (previously chosen target location) are encoded. Overall, substantial proportions of DLPFC neurons encoded both task-relevant and task-irrelevant features related to events from the previous trial (Fig. 2a). For example, during the interval surrounding the target onset (~250 to 250 ms relative to target onset), the proportion of neurons encoding information about the outcome in the previous trial (33.2%) and the previously chosen color (10.6%) was significantly greater than expected from chance (binomial test, $P < 10^{-3}$). Task-irrelevant information about the previously chosen location was even more strongly represented in the population than information about previously chosen color (28.3% versus 10.6%; two-proportion $z$-test, $P < 10^{-5}$) (Fig. 2b–d).

Neurons in DLPFC also encoded multiple types of information presented in the current trial (Fig. 3a). For example, in the epoch after target onset (250–750 ms after onset), 29.2% of neurons encoded the relative positions of the red and green targets. The neuron shown in Figure 3b increased its activity more when the red target appeared on the left than it appeared on the right ($t$-test, $P < 10^{-7}$). In the epoch after the onset of magnitude information (750–1,250 ms after target onset), many neurons also encoded the animal’s chosen color (15.0%), chosen target location (34.1%) and reward magnitudes (31.9%) (Fig. 3c–e).

We next examined how signals related to the previous trials’ events were integrated with information from the current trial. In particular, we tested how various task-related signals were combined to allow the animal to identify the location of the target with the higher reward probability (the high-value target location (HVL)) in the current trial. To identify the HVL, the animals need to integrate three pieces of information: the previously chosen color, the previous outcome and the configuration of target positions in the current trial. During the target period, we found that 18.1% (binomial test, $P < 10^{-12}$) of the neurons encoded a three-way interaction among these factors and therefore provided specific information about the HVL (Fig. 4a).
For example, the neuron shown in Figure 4b increased its firing rate when the animal had been rewarded in the previous trial and the previously chosen color was displayed on the right. In contrast, its firing rate decreased if the target color chosen and rewarded in the previous trial was on the left. During the magnitude period, activity was significantly influenced by the HVL in 29 neurons (12.8% binomial test, $P < 10^{-3}$), 16 of which (55.2%) showed significant effects of both HVL and choice ($\chi^2$-test, $P < 0.05$), suggesting that HVL and choice signals might coexist in the DLPFC.

Neurons that integrated individual elements of the HVL signal were also present. For example, before the onset of magnitude cues (~250 to 250 ms relative to target onset), 10.2% of the neurons integrated signals related to the previously chosen color and previous outcome. The single neuron example shown in Figure 4c reached its highest firing rate when the animal had previously chosen the red target and was rewarded. The integration and maintenance of such signals during the intertrial interval might provide the animals with the information needed to determine the HVL once the target location information is presented. Because we defined the HVL as the three-way interaction among the previously chosen color, previous outcome and current target configurations, we also used a decoding analysis to test whether HVL signals could be decoded from a population of neurons with multiple main effects of these three variables (three groups with different pairs of main effects and another group with all three main effects) or their two-way interactions, but without the significant three-way interaction (Online Methods). When corrected for multiple comparisons, none of these seven groups showed significant encoding of HVL signals, whereas, as expected, HVL signals were reliably decoded from the neurons with significant three-way interactions ($t$-test, $P < 10^{-5}$).

We found that information about the previous trial’s chosen location and outcome were combined to indicate the previously rewarded location (PRL), even though this information was irrelevant in the current task. In the epoch after target onset, 20.8% of the neurons encoded the PRL. The example neuron in Figure 4d increased its activity more when the animal had previously chosen the target on the right and was rewarded. Thus, although a substantial population of neurons in the DLPFC combined information at the single neuron level that was crucial for performing this task (HVL), many DLPFC neurons still encoded conjunctions that were not useful to the animals (PRL).

### Congruency of task-relevant information and choice signals

Given that task-relevant and irrelevant variables were both encoded in the DLPFC, we tested how each of these signals might be combined with the signals related to the animal’s upcoming choice. We found that the task-relevant information (HVL) tended to be encoded in the same population of neurons that later encoded the animal’s upcoming choice. Neurons encoding the HVL during the target period were more likely to encode the animal’s choice later during the magnitude period (24/41 neurons, 58.5%) than were the neurons without HVL signals (53/185 neurons, 28.6%; $\chi^2$ test, $P < 10^{-3}$). We next compared the regression coefficients for HVL defined in the epoch after target onset and the regression coefficients for upcoming choice defined in the epoch after the onset of magnitude cues. These regression coefficients were strongly correlated ($r = 0.55$, $P < 10^{-18}$), indicating that individual neurons encoded information about HVL and upcoming magnitude
choice consistently (Fig. 5a). Namely, neurons that increased their firing rates when the HVL was displayed on the right also tended to increase their firing rates later in the trial when the animal made a rightward choice.

We also found that the DLPFC neurons encoding task-irrelevant information about the previously chosen target location (PRL) tended to encode the upcoming choice later in the trial. Neurons encoding PRL during the target period were more likely to encode the animal’s upcoming choice during the magnitude period (24/47 neurons, 51.1%) than those without PRL signals (53 neurons, 29.6%; \( \chi^2 \) test, \( P < 0.01 \)). However, there was no significant correlation between regression coefficients for PRL and upcoming choice (Fig. 5b; \( r = 0.01; P = 0.93 \)), and this was significantly smaller than the correlation for HVL and choice signals (\( P < 10^{-10} \)). These results revealed an important difference in the way that task-relevant and irrelevant information are encoded. Although both types of signals are present in the DLPFC, only task-relevant information was encoded congruently by the neurons encoding the animal’s choice later. However, not all task-relevant information was treated in the same manner. We found that the information about the difference in reward magnitude was not encoded consistently by the choice-encoding neurons. During the magnitude period, the regression coefficients for the magnitude difference and choice were not significantly correlated (\( r = 0.02, P = 0.78 \)), nor were the regression coefficients for the magnitude difference and the HVL estimated during the target period (\( r = -0.02, P = 0.60 \)).

To further investigate how these signals were integrated in single neurons, we performed principal components analysis (PCA) on the regression coefficients (Online Methods) and examined the weights assigned to the regression coefficients for the animal’s upcoming choice, HVL and PRL. If these signals mix independently, we would expect the first principal component (PC1) to capture variability in the population activity related to only the variable associated with the largest changes in firing rate. We found that PC1 captured shared variance related to both choice and HVL (Fig. 5c). The maximum value of the weights related to HVL was significantly higher than observed after shuffling the relationship among the regression coefficients (\( P < 10^{-4} \)), but this was not true for the weights related to the task-irrelevant PRL (\( P = 0.11 \)). We repeated the same analysis on the absolute values of regression coefficients and found that weights for both HVL and PRL were significantly above the chance level (Fig. 5d; \( P < 10^{-4} \)). Thus, both HVL and PRL signals tended to combine in the same neurons that would eventually signal the animal’s choice, but only task-relevant HVL and choice were encoded in the same spatial reference frame.

**Effects of reward on DLPFC activity**

Because we observed a strong behavioral asymmetry after rewarded and unrewarded trials, we tested whether the outcome of the previous

![Figure 5a](image1.png)  ![Figure 5b](image2.png)  ![Figure 5c](image3.png)  ![Figure 5d](image4.png)

**Figure 5** Congruent coding of HVL and choice. (a) Relationship between regression coefficients for HVL versus choice (\( n = 226 \) neurons). (b) Relationship between regression coefficients for PRL versus choice (\( n = 226 \) neurons). (c, d) Time course of PC1 (see Online Methods) related to choice, HVL and PRL using signed (c) or unsigned (d) regression coefficients (\( n = 226 \) neurons). Arrows indicate critical values estimated from 10,000 random shuffles (\( P < 0.05 \)).

![Figure 6](image5.png)

**Figure 6** Effects of reward on task-relevant and task-irrelevant signals in the DLPFC. (a) Firing rate in example neuron encoding the HVL, shown separately for leftward and rightward choices in the current trial (t). (b) Decoding accuracy for the current location of the previously chosen color, previously chosen location and currently chosen location for previously rewarded and unrewarded trials (\( t = 1 \)) (\( n = 216 \) neurons that survived our criterion for cross-validation; Online Methods). Shaded areas represent s.e.m. The scatterplots (bottom) show the fraction of correct classifications for each neuron during the target period. Colors indicate whether zero, one or both decoding analyses applied to previously rewarded and unrewarded trials were significantly higher than chance (\( P < 0.05 \)), and large symbols indicate that the difference between rewarded and unrewarded trials was statistically significant (z-test, \( P < 0.05 \)).
trial also asymmetrically influenced the way in which different task-related variables were encoded in the DLPFC. Indeed, we found that DLPFC neurons tended to encode the current location of previously chosen color more robustly after a reward. For example, the neuron shown in Figure 6a increased its firing rate during the target period when the previously chosen and rewarded color was presented rightward. In contrast, it decreased its firing rate when the previously chosen and rewarded color was presented on the left. However, when the previous trial was unrewarded, the position of previously chosen color did not strongly affect the activity of this neuron. These effects were present regardless of whether the animal eventually chose the leftward or rightward target. Thus, the information pertaining to the location of the previously chosen color was more robustly encoded when the previous trial was rewarded (Fig. 4b). Furthermore, this neuron also increased its firing rate when the animal chose the rightward target in the current trial, suggesting that the HVL and choice were encoded in the same frame of reference as described above.

To investigate whether these effects were consistent across the population, we applied a linear discriminant analysis to decode the current location of the color chosen in the previous trial separately according to whether it had been rewarded or not (Fig. 6b). During the target period, we found that this information was encoded more robustly in both monkeys after a reward (paired t-test, \( P < 10^{-12} \)). The position of the previously chosen color could not be reliably decoded after unrewarded trials (paired t-test, \( P = 0.52 \)). We also found that the average decoding accuracy for the color of the target chosen in the previous reward trial (0.52) was significantly higher than that in the unrewarded trial (0.50; paired t-test, \( P < 0.05 \)). We next tested whether there were any differences in the ability to decode task-irrelevant information (namely, previously chosen target location) after rewarded and unrewarded outcomes, and found that there was no significant difference (Fig. 6b; paired t-test, \( P = 0.82 \)). In addition, we found a significant interaction between information type (location of previously chosen color versus previously chosen location) and previous outcome (two-way analysis of variance (ANOVA), \( P = 0.01 \)).

We also found that there was no difference in the way the neurons encoded upcoming choice information as a function of the previous outcome (paired t-test, \( P = 0.69 \); Fig. 6b). This confirmed that retrospective but not prospective task information was encoded more robustly after a rewarded outcome. Thus, task-relevant information (position of the color chosen in the previous trial) was encoded in a specific manner, becoming more robust following rewarded outcomes, whereas task-irrelevant information (previously chosen locations) was not influenced by the previous outcome. These results mirror the asymmetry in the behavior described above, indicating that rewarded outcomes encouraged the animals to choose the same color as in the previous trial and also enhanced the encoding of the corresponding information in the DLPFC.

### Anatomical distributions

Neurons encoding HVL, choice and PRL were broadly distributed throughout the DLPFC explored in this study (Supplementary Fig. 1). We performed one-way multivariate ANOVA (MANOVA) to see whether neurons encoding these variables were differentially distributed in either the anterior-posterior or dorsal-ventral axis. While neither HVL- nor PRL-encoding neurons showed a significant tendency to be located in any particular position (one-way MANOVA, \( P = 0.19 \) and \( P = 0.68 \), respectively), we found that neurons with choice-related activity tended to be located slightly more posteriorly (\( P = 0.03 \)). Neurons that conjunctively encoded both choice and HVL tended to be located more posteriorly and dorsally (\( P < 0.01 \)), and neurons that conjunctively encoded choice and PRL lacked any significant bias in their spatial distribution (\( P = 0.40 \)).

### Discussion

Although neurons in the DLPFC encoded a diverse collection of signals related to both relevant and irrelevant aspects of the task, we found clear differences in how this information was integrated at the single neuron level. DLPFC neurons tended to encode information about the current position of the previously rewarded target color and the animal’s upcoming choice congruently. In contrast, irrelevant information related to previously rewarded locations was encoded largely independently of upcoming choice locations. In addition, consistently with the animal’s behavior, we found that only the signals relevant to the task were enhanced following a rewarded outcome. Thus, DLPFC might be a crucial node in the network of brain areas that allows information from a variety of sources to be dynamically gated and transmitted to downstream neurons responsible for controlling the animal’s behavior.

### Behaviors related to the previous outcome

Although it is optimal to combine reward probabilities and magnitudes multiplicatively for calculating expected values, we found that both animals tested in this study used a heuristic additive strategy. Previous work has examined potentially different anatomical substrates for processing reward probability and magnitude information, but relatively little is known about how these signals are combined to calculate expected value. Although our results indicate that reward probability and magnitude signals are combined in the DLPFC, the animals did not use this information optimally. It is possible that the task design in this study did not provide sufficiently strong incentive to calculate the expected value of reward accurately, especially given that the animals also had to estimate reward probabilities.

We also found that rewarded outcomes had a much greater influence on the animals’ behavior than did unrewarded outcomes. This asymmetry has been reported across a wide range of behavioral paradigms, including a competitive game and a paired-associate task in primates, as well as in a free choice task using probabilistic outcomes in rodents. This suggests that rewarded and unrewarded outcomes might contribute differently to the neural process of updating the animal’s decision-making strategies.

Finally, we found that animals tended to update chosen and unchosen reward probabilities reciprocally in our task. This has potential implications for comparing model-free and model-based reinforcement learning. In simple model-free reinforcement learning, animals update only the action values associated with chosen actions. In our task, the objective reward probabilities were anticorrelated for the two targets. Exploiting knowledge of this structure would allow the animal to improve the speed and accuracy of learning by reciprocally adjusting the value of both targets. Different striatal and cortical circuits have been associated with model-free and model-based reinforcement learning. In addition, the prefrontal cortex might also have an important function in arbitrating the competition between multiple learning algorithms. However, how prefrontal activity related to previous outcomes might be used to adjust the estimates of multiple reward probabilities simultaneously is not known and should be investigated further.

### Integration of task-relevant signals in DLPFC

Decision-making in natural environments often requires animals to integrate various types of information over multiple timescales.
Indeed, persistent activity often carries information about actions and outcomes from previous trials in dynamic decision-making tasks\textsuperscript{13,39}, as well as in tasks that require the use of abstract response strategies depending on events in the previous trial\textsuperscript{41,42}. Neurons in the DLPFC are also capable of integrating information across multiple sensory modalities\textsuperscript{43}. In the present study, we found that many DLPFC neurons encode signals related to task-irrelevant events in previous trials. Similarly, some DLPFC neurons encode task-irrelevant features of sensory stimuli presented previously\textsuperscript{18}. Such encoding of task-irrelevant signals may be important for the behavioral flexibility often attributed to prefrontal cortex, as stimulus-outcome or action-outcome contingencies can change unexpectedly in a dynamic environment. In particular, residual coding of task-irrelevant information in the DLPFC might facilitate the detection of changes in the animal's environment. Consistently with this possibility, DLPFC neurons often encode task-irrelevant features during perceptual decision making\textsuperscript{17,44}. For example, in a recent study\textsuperscript{44}, monkeys were required to integrate noisy sensory evidence related to the color or motion direction of stimuli during a perceptual decision-making task. The relevant sensory evidence was selectively integrated in the prefrontal cortex along the axis in the state space corresponding to the animal's choice. In addition, the degree to which the neural activity in the state space deviates from the choice axis increased with the strength of sensory signals. Our results suggest that task-irrelevant information from memory, such as the animal's choices and their outcomes in preceding trials, might be encoded similarly in the DLPFC during value-based decision-making tasks. In particular, DLPFC signals related to task-relevant information were integrated better with the choice signals than signals related to task-irrelevant information. Therefore, similar mechanisms might exist in the prefrontal cortex for gating and selection of task-relevant information regardless of whether such information originates from sensory stimuli or previous experience. Coding of task-relevant signals in the DLPFC might change adaptively during learning to promote behaviors leading to more favorable outcomes in a particular context. Such flexible coding might not be unique to the DLPFC, as the directional tuning of neurons in the supplementary eye field (SEF) have been shown to shift while animals learn novel conditional visuomotor relationships\textsuperscript{15–47}.

We also found that DLPFC signals related to the task-relevant information, but not those related to task-irrelevant information, was selectively enhanced following a positive outcome. This tendency has been reported across a diverse set of brain regions, including DLPFC\textsuperscript{27,28}, SEF and lateral intraparietal area\textsuperscript{27}, striatum\textsuperscript{29} and hippocampus\textsuperscript{30}. Interestingly, the nature of the behavioral task may dictate precisely what types of signals are enhanced. For example, we have previously shown that signals related to past actions but not upcoming actions were encoded more robustly after rewards in a computer-simulated competitive game\textsuperscript{27}, whereas another study found similar effects related to signals encoding upcoming actions\textsuperscript{28}. In the present study, we found that signals related to neither past nor upcoming actions were influenced by the previous outcome. Instead, a more complex combination of signals corresponding to the location of the color chosen in the previous trial was enhanced following a rewarded outcome. This particular variable (HVL) indicated the position of the target with a high reward probability and was therefore relevant for the task in the present study. DLPFC neurons with HVL signals tended to encode the animal's choice once the information about the reward magnitude became available. Therefore, HVL signals might reflect the animal's tentative choice, which can be revised later. However, two lines of evidence suggest that HVL and choice signals are not equivalent. First, HVL signals were asymmetrically affected by the outcome of the animal's choice in the previous trial, such that information about the location of the previously chosen color was encoded in the DLPFC only after the rewarded trial. Second, even after the onset of magnitude cues, DLPFC neurons encoded HVL and choice signals concurrently, suggesting that the HVL signals were not merely temporary choice signals. Instead, reward-dependent enhancement of task-relevant signals in the DLPFC might be a general mechanism for learning to optimize the outcomes of the animal's behavior.

Our results suggest that task-relevant and task-irrelevant signals might be processed differently, despite the heterogeneity of signals represented in DLPFC. In addition, heterogeneous activity patterns in the DLPFC may provide computational advantages by broadening the possible range of information available in the DLPFC\textsuperscript{49}. This diversity may enable a substrate for representations to be updated flexibly on the basis of changing task demands\textsuperscript{49,50}.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank M. Hammond and P. Kurnath for technical support, and Z. Zhang for his help with the experiment. This study was supported by the National Institutes of Health (T32 NS007224 to C.H.D. and R01 DA029330 to D.L.).

**AUTHOR CONTRIBUTIONS**

C.H.D. and D.L. designed the experiments and wrote the manuscript. C.H.D. did the experiment and analyzed the data.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Funahashi, S., Bruce, C.J. & Goldman-Rakic, P.S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. 61, 331–349 (1989).
2. Lara, A.H. & Wallis, J.D. Executive control processes underlying multi-item working memory. Nat. Neurosci. 17, 876–883 (2014).
3. Romo, R., Brody, C.D., Hernández, A. & Lemus, L. Neuronal correlates of parametric working memory in the prefrontal cortex. Nature 399, 470–473 (1999).
4. Constantinidis, C., Franowicz, M.N. & Goldman-Rakic, P.S. The sensory nature of mnemonic representation in the primate prefrontal cortex. Nat. Neurosci. 4, 311–316 (2001).
5. Ō Scalaidhe, S.P., Wilson, F.A. & Goldman-Rakic, P.S. Areal segregation of face-processing neurons in prefrontal cortex. Science 276, 821–824 (1997).
6. Miller, E.K. & Cohen, J.D. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202 (2001).
7. Tanji, J. & Hoshi, E. Role of the lateral prefrontal cortex in executive behavioral control. Physiol. Rev. 88, 57–57 (2008).
8. Lebedev, M.A., Messinger, A., Kralik, J.D. & Wise, S.P. Representation of attended versus remembered locations in prefrontal cortex. PLoS Biol. 2, e365 (2004).
9. Messinger, A., Lebedev, M.A., Kralik, J.D. & Wise, S.P. Multitasking of attention and memory functions in the primate prefrontal cortex. J. Neurosci. 29, 5640–5653 (2009).
10. Watanabe, M. Reward expectancy in primate prefrontal neurons. Nature 382, 629–632 (1996).
11. Leon, M.I. & Shadlen, M.N. Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. Neuron 24, 415–425 (1999).
12. Rao, S.C., Rainer, G. & Miller, E.K. Integration of what and where in the primate prefrontal cortex. Science 278, 1135–1138 (1997).
13. Miller, E.K. & Cohen, J.D. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202 (2001).
14. Tanji, J. & Hoshi, E. Role of the lateral prefrontal cortex in executive behavioral control. Physiol. Rev. 88, 57–57 (2008).
15. Messinger, A., Lebedev, M.A., Kralik, J.D. & Wise, S.P. Multitasking of attention and memory functions in the primate prefrontal cortex. J. Neurosci. 29, 5640–5653 (2009).
16. Watanabe, M. Reward expectancy in primate prefrontal neurons. Nature 382, 629–632 (1996).
17. Leon, M.I. & Shadlen, M.N. Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. Neuron 24, 415–425 (1999).
ONLINE METHODS

Animal preparation. Two male rhesus monkeys (O and U, body weight 9–11 kg) were used. Their ages at the time of recording were 5.8 years (monkey O) and 5.1 years (monkey U). Monkey O had been previously trained on a manual joystick task but had not been used for electrophysiological recordings before this experiment. Monkey U had not been used for any prior experiments. Both animals were socially housed throughout these experiments. Eye movements were monitored at a sampling rate of 225 Hz with an infrared eye tracker (ET49, Thomas Recording, Germany). All procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at Yale University.

Behavioral task. The animals fixated a small white square at the center of a computer monitor to initiate each trial (Fig. 1a). Following a 0.5-s fore-period, two peripheral targets were simultaneously presented along the horizontal meridian. One of the targets was green, the other target was red, and their positions were randomized across trials. After a 0.5-s interval (target period), a set of small yellow tokens was presented around each target, indicating the magnitude of available reward (magnitude period). The central fixation cue was extinguished after a random interval ranging from 0.5 to 1.2 s, chosen according to a truncated exponential distribution with a mean of 0.705 ms. The animals were then free to shift their gaze toward one of the two peripheral targets. After fixating the chosen target for another 0.5 s, the animals received visual feedback. A red or green ring around the chosen target indicated that the animals would be rewarded, while a gray or blue ring indicated that they were not to be rewarded. An additional delay of 0.5 s (feedback period) separated the onset of the feedback ring from reward delivery. In trials where the animals were rewarded, they received the amount of apple juice associated with the chosen target. Each token corresponded to one drop of juice (0.1 ml).

The reward magnitudes associated with each target color were drawn from the following ten possible pairs: {1,1}, {1,2}, {1,4}, {1,8}, {2,1}, {2,4}, {4,1}, {4,2}, {4,4}, [8,1]. Each magnitude pair was counter-balanced across target locations, yielding 20 unique trial conditions (10 magnitude combinations × 2 target locations). At the beginning of a trial, a condition was randomly selected without replacement from the pool of 20. After completing 20 trials, the set of trial conditions was replenished to assure even sampling.

The red and green targets were associated with different reward probabilities, which were fixed within a block of trials and alternated across blocks that consisted of 20 or 80 trials. One target color was associated with a high reward probability (80%), and the other was associated with a low reward probability (20%). These reward probabilities underwent unsigned reversals across blocks so that the animals had to learn them through experience. Target reward probabilities were anticorrelated such that the reversals for the two targets always occurred simultaneously. We did not find any systematic differences in either animals’ behavior or in neural activity related to different block lengths. Therefore, we combined the data from all blocks in the analyses.

Analysis of behavioral data. The optimal strategy in this task is to multiply estimated reward probabilities with magnitudes to calculate expected values. To investigate how estimated reward probabilities and magnitudes were combined, we first tested a series of models to empirically estimate reward probabilities. Specifically, we examined variations of reinforcement learning (RL) models19 in which the reward probabilities were updated with either one learning rate used otherwise). Similarly, in the RL model with double learning rates (RL2), where $P_{tc}$ is the learning rate associated with rewarded outcomes, and $\alpha_{unr}$ is the learning rate associated with unrewarded outcomes. $\delta(x) = 1$, if x = 0, and 0 otherwise.

We also tested models to assess whether the animals might have had knowledge that the reward probabilities of the targets were anticorrelated. That is, we tested models where only the reward probability of the chosen target, $P_c(t)$, was updated (uncoupled) against models where the chosen target was updated as above, and the reward probability for the unchosen target was determined as $1 - P_c(t)$.

Using these estimates of reward probabilities, we proceeded to test whether they were combined multiplicatively with the reward magnitude, as would be expected if the animal made its choice according to the expected value of reward. To achieve this aim, we directly compared several variations of models in which probabilities and magnitudes were combined additively, multiplicatively or both. In the additive model ($P+M$),

$$\logit p(\text{Red}) = \beta_0 + \beta_1 \Delta P_{RG}(t) + \beta_2 \Delta M_{RG}(t)$$

where $\Delta P_{RG}(t) = P_{\text{red}}(t) - P_{\text{green}}(t)$ is the difference in estimated reward probabilities of the red and green targets, and $\Delta M_{RG}(t) = M_{\text{red}}(t) - M_{\text{green}}(t)$ is the difference in magnitudes of the targets. In the multiplicative or expected-value model (EV),

$$\logit p(\text{Red}) = \beta_0 + \beta_1 \Delta E_{RG}(t)$$

where $\Delta E_{RG}(t) = P_{\text{red}}(t) \cdot M_{\text{red}}(t) - P_{\text{green}}(t) \cdot M_{\text{green}}(t)$ is the difference in estimated expected values of the targets on trial t. The full model ($P+M+EV$) combined additive and multiplicative components:

$$\logit p(\text{Red}) = \beta_0 + \beta_1 \Delta P_{RG}(t) + \beta_2 \Delta M_{RG}(t) + \beta_3 \Delta E_{RG}(t)$$

Maximum likelihood estimates for model parameters were estimated using the fminsearch function in Matlab (Mathworks) by combining trials across all sessions. Because the models differed in the number of free parameters, we used the Bayesian information criterion (BIC) for model comparison, which penalized the use of additional parameters. We tested all 12 combinations of models on the basis of the three different factors described above—namely, the number of learning rates (RL1 or RL2), whether the animals exploited the correlation in reward probabilities (coupled or uncoupled) and how probability and magnitude information was combined ($P+M$, EV, or $P+M+EV$). These models were separately fit to the behavioral data in each monkey.

Neurophysiological recording. Activity of individual neurons in the DLPFC was recorded extracellularly (left hemisphere in both monkeys) using a 16-channel multielectrode recording system (Thomas Recording, Germany) and a multichannel acquisition processor (Plexon, TX). On the basis of magnetic resonance images, the recording chamber was centered over the principal sulcus and located anterior to the genu of the arcuate sulcus (monkey O, 4 mm; monkey U, 10 mm). All neurons selected for analysis were located anterior to the frontal eye field, which was defined by eye movements evoked by electrical stimulation in monkey O (current <50 μA). The recording chamber in monkey U was located sufficiently anterior to the frontal eye field, so stimulation was not performed in this animal. Each neuron in the data set was recorded for a minimum of 320 trials (77 and 149 neurons in monkeys O and U, respectively), and on average for 518.8 trials (s.d. = 147.2 trials). We did not preselect neurons on the basis of activity, and all neurons that could be sufficiently isolated for the minimum number of trials were included in the analyses.

Analysis of neural data. Linear regression analyses. A linear regression model was used to investigate how each neuron encoded both task-relevant and task-irrelevant information from the current and previous trials. Some of this information relied on the integration of simpler, more basic elements. For example, task-irrelevant information about the location of the rewarded target in the previous trial required knowledge of both the previously chosen location and previous outcome. Therefore, to analyze these higher-order features, we used the
following multiple linear regression model, which included simpler potentially confounding terms:

\[
y = \beta_0 + \beta_1 C_{LR}(t) + \beta_2 C_{RG}(t) + \beta_3 R(t-1) + \beta_4 PRL(t-1) + \beta_5 POS_{RG}(t) + \beta_6 M_{LR} + \beta_7 M_{RG} + \beta_8 M_{ML} + \beta_9 M_{LH} + \beta_{10} C_{RG}(t-1) + \beta_{11} PRL(t-1) + \beta_{12} POS_{RG}(t-1) + \beta_{13} C_{RG}(t-1) \times R(t-1) + \beta_{14} PRL(t-1) + \beta_{15} HVL(t)
\]

where \(C_{LR}(t)\) is the position of the chosen target on trial \(t\) (1 if right, −1 otherwise), \(R(t)\) is the outcome on trial \(t\) (1 if rewarded, −1 otherwise), \(POS_{RG}(t)\) is the positions of the red and green targets on trial \(t\) (1 if red is on the right, −1 otherwise), \(M_{LR}(t)\) and \(M_{RG}(t)\) are the magnitudes of rewards available from the right and left targets on trial \(t\), respectively, and \(C_{RG}(t)\) is the color chosen by the animal on trial \(t\) (1 if red, −1 if green). In this model, task-irrelevant information related to the previously rewarded location, abbreviated hereafter as PRL(t), is given by the interaction between the previously chosen location and previous outcome, namely \(PRL(t) = C_{LR}(t-1) \times R(t-1)\). Task-relevant information corresponding to the high-value target location, HVL(t), is defined as the current position of the target color that was rewarded in the previous trial. This corresponds to a three-way interaction among the previously chosen color, previous outcome and current color locations, namely, \(HVL(t) = C_{RG}(t-1) \times R(t-1) \times POS_{RG}(t)\). To test whether signals related to reward magnitude and choice are encoded consistently across different DLPCF neurons, we have also tested the same regression model in which the two terms related to reward magnitude \(M_{LR}(t)\) and \(M_{RG}(t)\) were replaced by their difference \(\Delta M = M_{LR} - M_{RG}\).

Correlation and principal components analyses. As described in Results, partially overlapping populations of neurons encoded task-relevant information related to HVL and task-irrelevant information related to PRL. To investigate how these populations were related to the neurons that encoded the animal’s choice location, we examined the correlation between the regression coefficients calculated in non-overlapping 500-ms time bins. The use of nonoverlapping bins was necessary for avoiding biases related to correlations between the regressors themselves. Therefore, the regression coefficients corresponding to HVL and PRL were estimated for the epoch following target onset (250–750 ms after target onset), and the regression coefficients for choice were taken from the period following magnitude onset (750–1,250 ms relative to target onset).

We performed principal components analysis (PCA) to examine this further. PCA is an orthogonal linear transform used to find the axes that maximize the variance explained in the data. We found that the regression coefficients corresponding to HVL were strongly correlated with choice signals, which suggests that a mixture of signals related to HVL and choice may be present in the first principal component (PC1). Therefore, we constructed a matrix based on the regression coefficients derived from each variable. Rows were made up of each neuron, and columns were comprised of each variable’s regression coefficient as a function of time. We applied PCA to this matrix and plotted the time course of loadings for PC1 corresponding to choice, HVL and PRL. To assess the statistical significance of observed loadings revealed by this analysis, we randomly shuffled labels corresponding to the identities of different neurons independently for different variables 10,000 times. Next, we determined where the maximum value corresponding to each regression coefficient was relative to the shuffled distribution. Values corresponding to \(\alpha = 0.05\) were used to determine whether the loadings for the PC1 were statistically significant or not.

We tested whether the regression coefficients for magnitude difference (ΔM) during the magnitude period were correlated with those for the HVL during the target period. We also examined how the regression coefficients for magnitude difference and choice were correlated during the magnitude period. However, the value of correlation coefficients between regression coefficients can be biased owing to the correlation between the regressors when they are estimated from the same data. Therefore, to test the statistical significance of the correlation between the regression coefficients estimated for the same time interval, we randomly divided the trial results into two groups and examined the correlation coefficient between the regression coefficients estimated from two separate groups of trials.

Decoding analysis. To test whether signals related to the position and color of the target chosen by the animal was enhanced after a reward or not, we applied a linear discriminant analysis with fivefold cross-validation separately for each neuron. As in our previous study, the training set was balanced according to the current location of the previously chosen target, the upcoming choice location, and the previous outcome (yielding eight different categories). To ensure an unbiased classifier, we randomly removed trials until each category had the same number of trials. We first used the classifier to see whether we could extract information about the firing rates of neurons in the current trial for previously rewarded and unrewarded trials separately. We applied a 500-ms sliding window advanced in 50-ms increments, from 500 ms before target onset to 1,500 ms after target onset. Trials were randomly assigned to five different subgroups (for fivefold cross-validation). Each subgroup served as a test set, and the combined trials from the four remaining subgroups served as a training set, which we used to find the firing rate boundary that maximally separated rightward and leftward categories. We then calculated how well we could classify the trials in the test set given this firing rate boundary. The process was repeated for each subgroup and the results averaged to obtain a time course for decoding accuracy for each neuron. Population means were constructed by averaging over all neurons. The same decoding procedure was repeated for decoding PRL and the animal’s upcoming choice. In the process of balancing the decoder, the number of trials available was not sufficient (<5 trials in one category) for a small fraction of neurons (10/226) to allow for proper cross-validation. We removed these neurons from the analysis and ran the decoder on the remaining 216 neurons that survived this criterion (77/77 neurons in monkey O; 139/149 neurons in monkey U).

Statistical analysis. Binomial tests (one-tailed) were used to determine whether the fraction of neurons encoding a particular signal was significantly greater than chance (5%). Paired t-tests (two-tailed) were used to test whether means for firing rates or decoding accuracy significantly differed between experimental conditions. χ²-tests were used to test whether neurons independently encoded choice and HVL or PRL. The statistical significance of specific regressors in the multiple linear regression analysis was assessed with a two-tailed t-test. Data distribution was assumed to be normal but this was not formally tested.

All neurons that could be isolated for at least 320 trials were included in this data set, and there was no blinding for data collection or analysis. Trials were pseudo-randomly presented such that one trial condition was randomly chosen without replacement from a pool of 20 possible conditions, balancing across target magnitudes and positions. After successful completion of all trial conditions, the set was replenished to assure even sampling. The sample size in the present study was not based on any statistical test but is comparable to that of our previous studies.

A Supplementary Methods Checklist is available.