nature neuroscience

Neuronal basis of sequential foraging decisions in a patchy environment

Benjamin Y Hayden[1,2], John M Pearson[1,2] & Michael L Platt[1–3]

Deciding when to leave a depleting resource to exploit another is a fundamental problem for all decision makers. The neuronal mechanisms mediating patch-leaving decisions remain unknown. We found that neurons in primate (*Macaca mulatta*) dorsal anterior cingulate cortex, an area that is linked to reward monitoring and executive control, encode a decision variable signaling the relative value of leaving a depleting resource for a new one. Neurons fired during each sequential decision to stay in a patch and, for each travel time, these responses reached a fixed threshold for patch-leaving. Longer travel times reduced the gain of neural responses for choosing to stay in a patch and increased the firing rate threshold mandating patch-leaving. These modulations more closely matched behavioral decisions than any single task variable. These findings portend an understanding of the neural basis of foraging decisions and endorse the unification of theoretical and experimental work in ecology and neuroscience.

Resources are rarely distributed uniformly in the environment. Food, water and other vital commodities more often occur in spatially localized and temporally ephemeral patches[1]. Patchy environments force animals to balance the benefits of staying in a depleting patch and leaving for a richer one[2]. According to the marginal value theorem (MVT) of behavioral ecology, animals should leave patches when their intake rate diminishes to the average intake rate for the overall environment[2,3]. Organisms as diverse as worms, bees, wasps, spiders, fish, birds, seals and even plants obey the MVT[3–6]. Ethnographic evidence demonstrates that human subsistence foragers also obey the predictions of the MVT in their hunting behavior[7], and laboratory findings suggest that monkeys may do so as well[8]. The generality of the MVT solution to the patch-leaving problem suggests that the underlying mechanism is fundamental to the way organisms make decisions[4]. The neuronal basis of patch-leaving decisions, however, remains unknown.

Building on recent progress toward understanding the neuronal mechanisms mediating perceptual decisions[9], we hypothesized that the brain maintains a decision variable specifying the current relative value of leaving a patch. Conceptually, a decision variable is an analog quantity that incorporates all sources of information—in this case, reward size, handling time, search time and travel time—evaluated by the decision policy to generate a behavioral choice[9]. The hypothesized decision variable gives rise to a decision via comparison with a specific threshold. For simplicity, we assume that this process is analogous, although not necessarily isomorphic, to the neural integrate-to-threshold processes thought to mediate perceptual judgments[9–13]. We further conjecture that travel time between patches influences leaving decisions by changing the rate at which the decision variable grows, the threshold or both[10,11].

To test these hypotheses, we developed a virtual foraging task in which rhesus monkeys chose one of two targets. One target corresponded to remaining in the patch, and choosing it yielded a juice reward that declined each time it was chosen. The other target corresponded to leaving the patch, and choosing it yielded only a delay before the opportunity to choose again at a replenished patch. Monkeys' behavior closely matched the predictions of the MVT. We then recorded activity of neurons in the dorsal anterior cingulate cortex (dACC) while they performed the task.

dACC has been linked to reward outcome monitoring and behavioral adjustment[14–16], as well as to signaling reward outcomes and predicting changes in behavior[17–24]. Notably, ACC dysfunction attends clinical disorders that are associated with difficulty in abandoning maladaptive patterns of behavior or cognition, including depression, addiction, obsessive-compulsive disorder and Tourette Syndrome[25–27].

We found that dACC neurons responded each time monkeys made a choice and that these responses increased with time spent in the current patch. Monkeys abandoned a patch when neuronal responses reached a threshold associated with a particular travel time. When travel time between patches was high, the gain of neuronal responses with each decision to remain in the patch was smaller and the threshold for patch abandonment was higher than when travel time was short. Overall, neuronal response gain and threshold jointly predicted patch-leaving decisions. These findings suggest that dACC mediates patch-leaving decisions using a common integrate-to-threshold mechanism.

RESULTS

For each choice there were two options (**Fig. 1a**). Choosing the stay (short blue) target led to a juice reward in 0.4 s (handling time). The value of this reward declined by 19 μl ± ε (s.e.m., ε = 1.9 μl) each time it was chosen, mirroring the diminishing returns common to patchy foraging environments (**Fig. 1b**). Choosing the leave (tall gray) target led to no reward and a long delay that was fixed in a patch and varied

[1]Department of Neurobiology, Duke University School of Medicine, Durham, North Carolina, USA. [2]Center for Cognitive Neuroscience, Duke University, Durham, North Carolina, USA. [3]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA. Correspondence should be addressed to B.Y.H. (benhayden@gmail.com).

Received 9 November 2010; accepted 28 March 2011; published online 5 June 2011; doi:10.1038/nn.2856

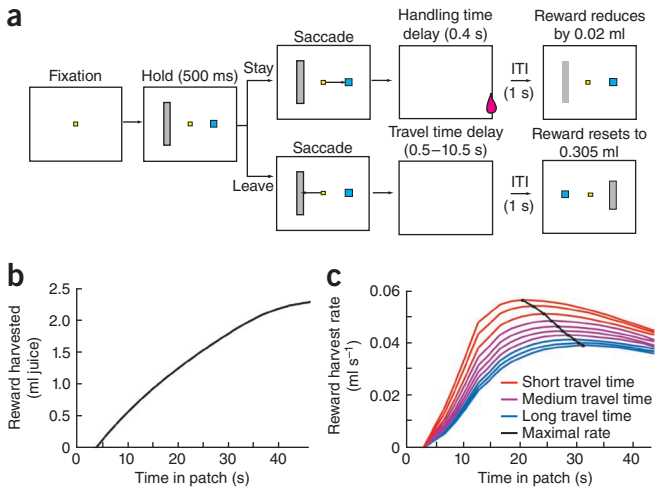

Figure 1 Patch-leaving task. (**a**) Task design. After fixation, two eccentric targets, a large gray and a small blue rectangle, appear. Monkey chooses one of two targets by shifting gaze to it. Choice of blue rectangle (stay in patch) yields a short delay (0.4 s, handling time) and reward whose value diminishes by 19 µl per trial. Choice of gray rectangle (leave patch) yields no reward and a long delay (travel time) whose duration is indicated by the height of the bar, and resets the value of the blue rectangle at 306 µl. Travel time varies randomly from patch to patch and ranges from 0.5 to 10.5 s. (**b**) Plot of the cumulative reward available in this task as a function of time in patch, given the search times associated with animals' performance in the task (black line). Data are generated on the basis of average times associated with performance. (**c**) Plot of reward intake rate derived from a range of patch residence times (*x* axis: range of residence times). Data are shown for each of ten travel times (1-s intervals from 0.5 to 10.5 s). Rate-maximizing time in patch (the curves' maxima, shown by the black line) increases with increasing travel time. Data are generated based on average times associated with actual animal performance.

between patches, and reset the value of the stay target to its initial high value (306 µl). We defined search time as any additional time spent in the patch not explicitly waiting and residence time as total time from arrival at a new patch, including handling times, search time and intertrial intervals (ITIs). Travel time was explicitly cued on all trials by the height of the gray bar and was reset to a new random value each time it was chosen (0.5 to 10.5 s, uniform distribution). The blue and gray bars alternated sides each time the leave target was chosen; any potential laterality in neural tuning functions was assumed to average out (**Supplementary Data 1**). In contrast to some natural foraging decisions, there was no physical travel during the travel time, nor was any action required during this delay; the only explicit cost of delay was opportunity cost. These simplifications are, we believe, not critical, as most other laboratory foraging tasks eschew effort requirements (for example, see refs. 8,28).

Monkeys approximate rate maximization according to MVT

As travel time between patches grows, so does the rate-maximizing residence time (**Fig. 1c**). Consistent with the MVT, monkeys' patch-residence times rose with increasing travel time and were nearly rate maximizing ($P < 0.0001$, $\beta = 1.247$, regression of residence time (s) against travel time (s); **Fig. 2a**). These effects were found in both monkeys individually ($P < 0.0001$ for both individuals, $\beta = 1.11$ for monkey E, $\beta = 1.47$ for monkey O; **Supplementary Data 2** and **Supplementary Fig. 1**). Overall, both monkeys obtained 97.2% of the reward obtained by the best-fit rate-maximization algorithm (note that this is a measure of reward obtained versus maximal obtainable, not a measure of variance in behavior explained). Both monkeys remained in patches slightly longer than predicted by the MVT (mean 2.2 s longer, $P < 0.01$, Student's *t* test). This slight over-staying may reflect a weak preference for immediate small rewards over delayed large rewards[29–31], a slight over-estimate of travel times or even a status quo bias[32]. Leaving time was not influenced by travel time on the previous patch (regression of residence time against previous travel time, $P = 0.44$; **Supplementary Data 3** and **Supplementary Fig. 2**).

Monkeys attempting to maximize local intake rates over the long-term should consider handling time as well as travel time[2,3]. To confirm that monkeys do so, we performed an additional behavioral experiment in which handling times, but not travel times, were varied from patch to patch (11 sessions, 6 in monkey E, 5 in monkey O). In each patch, handling time took one of ten values: 0.1, 0.4 , 0.6, 1.1, 1.6, 2.1, 2.6, 3.1, 3.6 or 4.1 s. We cued handling time by varying the height of the blue rectangle; travel time was held constant at 5 s. We performed these experiments after the monkeys had learned the task, but before we began physiological data collection; monkeys received one full day of training each. As predicted, patch residence times declined with increased handling time (regression, $\beta = -3.71$ for both, −3.81 for E and −3.62 for O, $P < 0.001$ in all cases; **Fig. 2b**). Leaving times did not differ systematically from the rate-maximizing predictions at any of the ten points ($P > 0.05$ in all cases, Student's *t* test). As an additional control, during these sessions, we interleaved standard fixed handling time patches with variable travel times. The average residence time on these trials was consistent with those obtained in the handling time control (**Fig. 2b**).

A natural question is whether the monkeys' foraging behavior may be explained in a delay-discounting framework[30,33]. In such a framework, each leave/stay decision is regarded as a choice between a smaller-sooner stay reward and larger-later leave reward (that is, the first reward in the new patch). Such a decision model would naturally account for the monkeys' observed tendency to stay longer in patches when faced with longer travel times, as the delay associated with patch-leaving would lead to discounting of the larger-later reward. To test this idea, we compared an empirically derived sequential foraging model inspired by the MVT against a standard delay-discounting model in which the hyperbolic discount parameter *k*

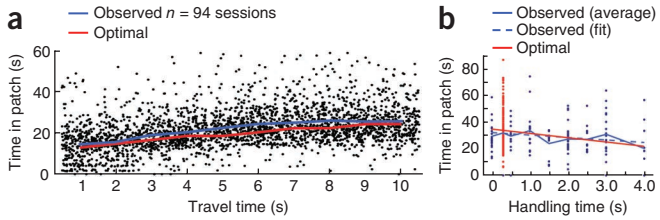

Figure 2 Monkeys obey the marginal value theorem in a virtual patchy foraging task. (**a**) Monkeys remain in the patch longer as travel time rises, as predicted by the marginal value theorem (MVT). Each dot indicates a single patch-leaving decision ($n = 2{,}834$ patch-leaving events). The time at which the monkey chose to leave the patch (*y* axis) was defined relative to the beginning of foraging in that patch. Travel time was kept constant in a patch (*x* axis). Data from both monkeys is shown. Behavior (average is traced by the blue line) closely followed the rate-maximizing leaving time (red line), albeit delayed by 0–2 s. (**b**) Performance of two monkeys on handling time variant of patch-leaving task. In this control experiment, travel time was held constant (5 s) and handling time was randomly reset between each patch to have one of ten values. Patch residence time fell as handling time rose, consistent with the MVT. Observed times are shown with black dots; averages are shown with solid blue line. Best-fit line (dashed blue line) is nearly identical to rate-maximizing (red line). Average patch residence time on the interleaved standard travel time version of the task was consistent with this curve as well (red dots).

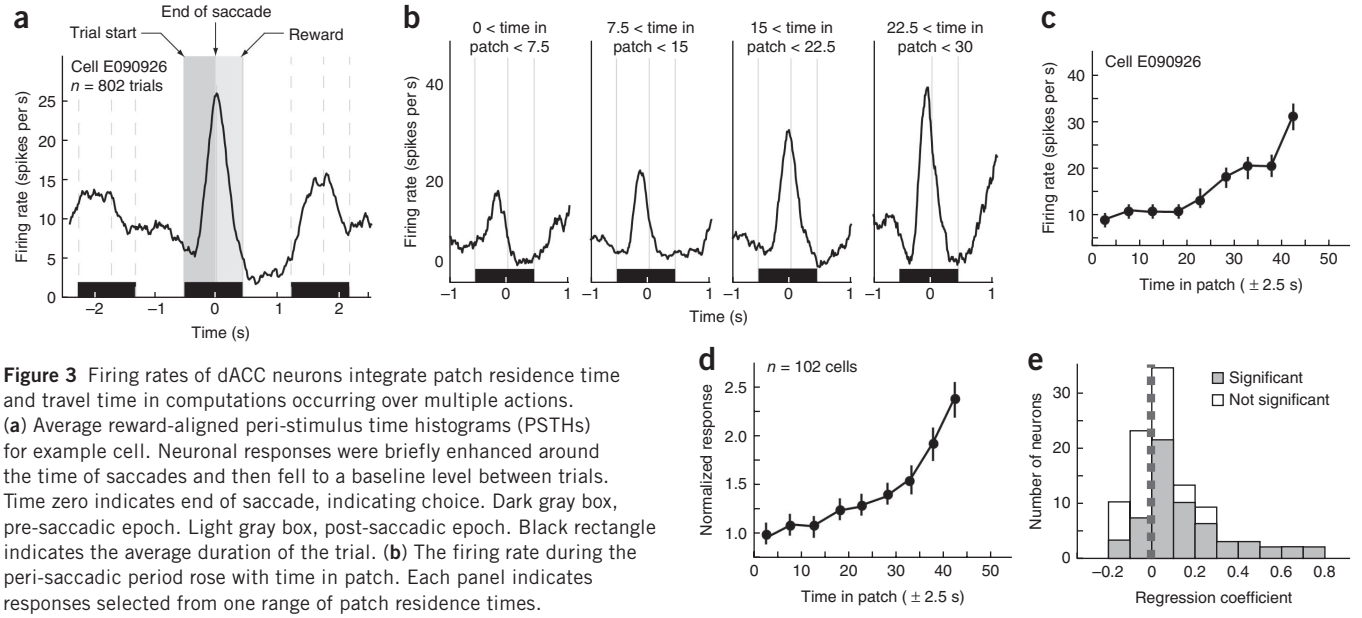

Figure 3 Firing rates of dACC neurons integrate patch residence time and travel time in computations occurring over multiple actions. (**a**) Average reward-aligned peri-stimulus time histograms (PSTHs) for example cell. Neuronal responses were briefly enhanced around the time of saccades and then fell to a baseline level between trials. Time zero indicates end of saccade, indicating choice. Dark gray box, pre-saccadic epoch. Light gray box, post-saccadic epoch. Black rectangle indicates the average duration of the trial. (**b**) The firing rate during the peri-saccadic period rose with time in patch. Each panel indicates responses selected from one range of patch residence times. (**c**,**d**) Average responses of example neuron (**c**) and population of neurons (**d**) occurring in a series of 5-s analysis epochs (gray box in **a**). Firing rates increased as time in patch increased. Error bars represent s.e.m. (**e**) Histogram of regression coefficients relating firing rate in pre-saccadic epoch to time in patch for each neuron in the population ($n = 102$). Significant effects are indicated with gray boxes ($P < 0.05$).

was estimated by maximum likelihood (best fit, $k = 1.26\ s^{-1}$). The Akaike weights for the two models (a measure of goodness of fit that accounts for different numbers of parameters in different models; w_{MVT} = weight for MVT model, w_{DD} = weight for delay-discounting model) clearly favored the sequential-trial foraging model, endorsing the MVT account of decision making in our task ($w_{MVT}/w_{DD} = 1.31 \times 10^{52}$) (**Supplementary Data 4** and **Supplementary Fig. 3**).

Neurons integrate information over multiple actions

We recorded the activity of 102 single neurons in dACC in two monkeys performing this task (52 neurons in monkey E, 50 in monkey O; **Fig. 3**). For an example neuron, neural activity was aligned to the end of the choice saccade (time zero; **Fig. 3a**). Firing rate rose to a peak around the time of the choice saccade and then returned to a baseline value between trials. Such brief, peri-saccadic responses, often modulated by reward size and task context, are characteristic of neurons in dACC[18,21,22,24]. We focused on neuronal activity in the 500-ms epoch preceding saccade onset (pre-saccadic epoch). For most analyses, we focused on neural data associated with choosing to remain in the patch and excluded neural data associated with choosing to leave the patch (exceptions are noted). Data for individual subjects matched the combined data (**Supplementary Data 5** and **Supplementary Fig. 4**).

We next examined the responses of the example neuron from four time periods relative to the beginning of foraging in the patch ($t <$ 7.5 s, $7.5 < t < 15$, $15 < t < 22.5$, and $22.5 < t$; **Fig. 3b**). For this neuron, responses rose with cumulative time spent foraging in the patch. To quantify this enhancement, we measured pre-saccadic responses in a series of non-overlapping 5-s time bins (**Fig. 3c**). We included in each time bin all of the decisions in which the end of the saccade occurred in that bin. We found that firing rates rose with increasing patch residence time ($\beta = 0.31$, $P < 0.0001$, linear regression of firing rate (spikes per s) against time in patch (s)). The same effects were observed in the population average firing rates ($\beta = 0.18$, $P < 0.0001$, regression; **Fig. 3d**, **Supplementary Data 6** and **Supplementary Fig. 5**). We observed a significant ($P < 0.05$) positive regression coefficient in 49 neurons (average $\beta = 0.24$ in significantly modulated cells), a significant negative slope in 10 (average $\beta = -0.09$) and no significant slope in the remainder ($P > 0.05$, $n = 43$, average $\beta = 0.041$). The 49 neurons with positive slopes constitute the focus of subsequent analyses (**Supplementary Data 7** and **Supplementary Figs. 6** and **7**).

We next performed the same analysis on two later epochs. In the post-saccadic epoch, we measured firing rates during the 400-ms handling time period beginning at saccade termination and ending with the reward. In the ITI epoch, we measured firing rates during the 1-s period beginning after reward delivery and ending when the next set of choice options was presented to the monkey. For the post-saccadic epoch, we observed a positive regression coefficient in 44 neurons ($P < 0.05$, average $\beta = 0.15$ in significantly modulated cells), a negative coefficient in 8 (average $\beta = -0.1$), and no significant effect in the remainder ($P > 0.05$, 50 neurons, average $\beta = 0.010$). Nor was there much evidence of an effect in the ITI epoch at the population level; we observed a significant correlation between firing rates in the ITI and patch residence time in only four neurons (average $\beta = 0.02$), all positive in sign. These numbers are similar to what would be expected by chance (**Supplementary Data 8**). This result suggests that dACC does not maintain a representation of time spent foraging in a patch across multiple actions; the locus of this trace in the brain remains to be determined. Overall, we observed weak or no effect of saccade direction on responses (**Supplementary Data 1**).

Threshold-crossing of dACC firing predicts patch-leaving

The gradual rise in neural responses across decisions to stay in a patch resembles the within-trial rise-to-threshold processes observed in lateral intraparietal area, frontal eye fields (FEFs) and superior colliculus during motor preparation and decision-making[10–12,34,35]. We wondered whether a similar rise-to-threshold model might also account for the relationship between firing rates in dACC and patch-leaving decisions. To test this idea, we performed an analysis modeled on a previously developed method[11] for probing the relationship between the firing rates of FEF neurons and saccade initiation. Although FEF firing rates in that study[11] rose gradually to a fixed threshold on a single trial, the analogous rise in our study

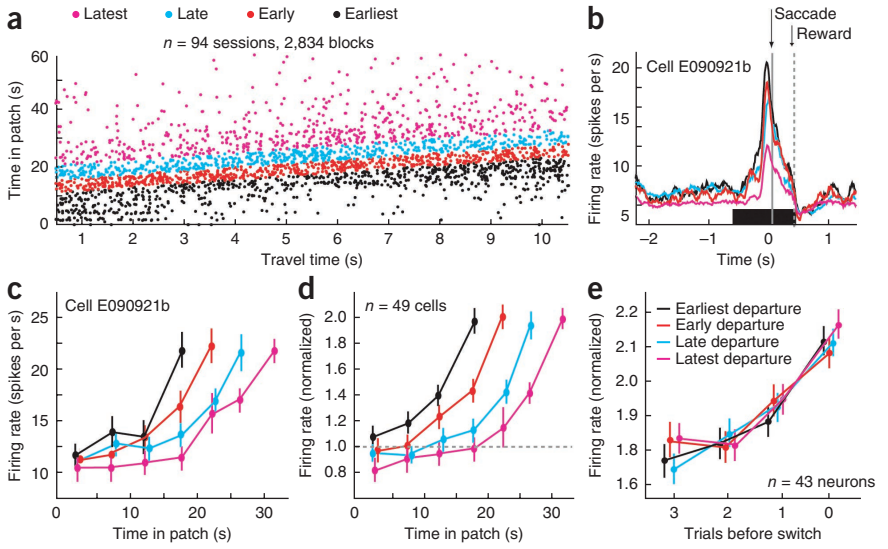

Figure 4 Firing rates of dACC neurons rise to a threshold associated with patch abandonment. (**a**) Plot of patch-leaving times, separated by whether they were earlier or later than the average leaving time. We divided patch-leaving decisions into four categories: earliest (black), early (red), late (cyan) and latest (magenta). These variables are independent of travel time and time in patch, meaning that, for example, earliest trials are equally likely to occur at any travel time (*x* axis) and any time in patch (*y* axis). (**b**) PSTH for an example neuron separated by earliness level. dACC neurons responded sooner and more strongly on earlier trials than on later trials. Black rectangle indicates the average duration of the trial. (**c**,**d**) Average firing rates of example neuron (**c**) and population (**d**) separated by earliness level. Firing rates rose faster for earlier patches but asymptoted at the same level. Error bars represent s.e.m. (**e**) Plot of average firing rate of population of neurons, aligned to final trial in patch (*x* = 0 on graph) and showing the final three trials before switch (*x* = 1, 2 and 3). Firing rates rose to the same level on final trial, as well as preceding trials. Error bars represent s.e.m.

was associated with changes in firing rate occurring in discrete bouts over multiple actions. Moreover, firing rates of FEF neurons gradually rise to threshold over tens of milliseconds, whereas the amplitudes of discrete dACC neuronal responses in our task increased over tens of seconds, orders of magnitude longer. Nonetheless, these analytical methods in principle generalize readily to our task. Our analysis asked two questions. First, does variability in patch-leaving times correlate with variability in the rate at which neural activity rises? Second, do firing rates rise to the same level regardless of the precise time monkeys choose to leave the patch for a given travel time?

We first divided all patch-leaving choices into residence time quartiles in each travel time (medians of each set for the aggregate response, earliest = 14.1 s, early = 19.2 s, late = 23.5 s, latest = 32.2 s; **Fig. 4a**). We refer to this classification of trials as the 'earliness' for each patch. We repeated the classification of leaving times into four different earliness bins for each neuron separately. By design, earliness is orthogonal to travel time, and so neural correlates of earliness and travel time are independent. For our example neuron, firing rates were significantly higher for the earliest than for the early patches (difference = 1.8 spikes per s, $P = 0.02$; **Fig. 4b**), significantly higher for the early than for the late patches (difference = 2.0 spikes per s, $P < 0.01$) and significantly higher for the late than for the latest patches (difference = 4.7 spikes per s, $P < 0.01$).

We next calculated the average neural responses in a series of 5-s time bins encompassing multiple choices separated by earliness level. For both the example neuron and the focal population of neurons ($n = 49$ of 102, see above), we found that the rate of rise of firing rates was positively correlated with earliness (**Fig. 4c**,**d**). We quantified these effects by calculating the regression weight for firing rate as a function of patch residence time separately for each of the four earliness bins. The slope for the earliest patches ($\beta = 0.71$) was greater than the slope for the early patches ($\beta = 0.52$, $P < 0.01$, bootstrap test), which was greater than the slope for the late patches ($\beta = 0.44$, $P < 0.01$), which was in turn greater than the slope for the latest patches ($\beta = 0.39$, $P < 0.01$). The same effects were observed in the population average responses ($P < 0.005$ for each comparison). Slopes decreased monotonically from earliest to latest quartiles for 32 of 49 neurons (65%). The remainder ($n = 17$) showed no such effect.

We then examined whether firing rates rose to the same threshold in each of the four earliness bins. For this analysis, we only examined firing rates for patch-leaving choices, which we had ignored in previous analyses. We took these firing rates to be a proxy for the patch-leaving threshold. Firing rates did not depend on earliness for the example neuron (regression of firing rate against earliness level, $P = 0.45$; **Fig. 4c**) or the population ($P = 0.88$; **Fig. 4d**), which is consistent with the threshold hypothesis. We observed no relationship between earliness and threshold for the population average response (regression, $P = 0.79$). We observed a significant effect of earliness on threshold for only a small number of cells, 6 out of 49 significantly modulated neurons (12.2%, $P < 0.05$). Finally, we found that neuronal responses aligned to the patch-leaving trial for different earliness levels overlapped (**Fig. 4e**). For the population of neurons, firing rates rose to approximately the same level on the last decision to stay in the patch before choosing to leave the patch (regression, $\beta = -0.003$, $P = 0.51$).

Travel time influences gain and threshold of neuronal responses

Assuming that patch-leaving decisions are governed by a threshold process, variation in travel times should influence the threshold. There are three basic mechanisms by which travel time could influence the accumulation-to-threshold process (**Fig. 5a**). First, travel time could increase the rate of rise of the decision variable. Second, travel time could adjust the threshold level. Third, travel time could influence the baseline. Our next analysis was designed to determine which of these processes are implemented in dACC.

We first examined the relationship between travel time and response gain (**Fig. 5b**,**c**). For each neuron, we divided patches into ten travel time deciles (equal-sized sequentially classified bins, 1 to 10 ± 0.5 s). In each decile, we calculated the slope of firing rate versus time in patch. We found a significant negative correlation between travel time and regression slope for our example neuron (regression, $\beta = -0.04$, $P < 0.01$; **Fig. 5b**) and for the 49 cells in the analysis population ($\beta = -0.033$, $P < 0.01$; **Fig. 5c**). We observed significant negative effects in 29 of 49 of the focal population of neurons, positive correlations in 6, and no effect in 14 ($P < 0.05$). Across the entire population, we found a significant negative correlation in 41 of 102 neurons ($P > 0.05$), a positive correlation in 7 neurons and no effect in 54 neurons.

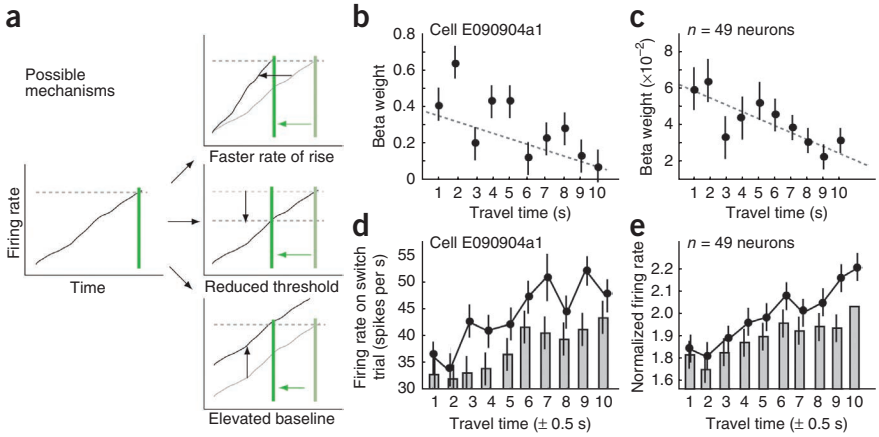

Figure 5 Travel time governs both neuronal response gain and threshold. (**a**) Schematic of three possible mechanisms by which exogenous factors may govern a rise-to-threshold process. Shorter travel times can hasten patch-leaving (leftward movement on *x* axis) by increasing the rate of rise, reducing the threshold or elevating the baseline. (**b**,**c**) Evidence that travel times change rate of rise. Example neuron (**b**) and population (**c**) average regression slopes (beta weights) for firing rate as a function of time in patch. Beta weights fell as travel times rose, indicating that shorter travel time increases neuronal response gain. Error bars represent s.e.m. (**d**,**e**) Evidence that travel time influences firing threshold for patch abandonment. Firing rate on patch-leaving trial was taken as a proxy for threshold level. Example neuron (**d**) and population (**e**) show increasing firing rates on patch-leaving trial as travel time increases (black dots). Firing rate on penultimate trial also rose with travel time, consistent with a multi-trial integration process. Error bars represent s.e.m.

We next examined the relationship between travel time and threshold (**Fig. 5d**,**e**). As above, we assumed that the firing rate in the epoch immediately preceding patch-leaving provides a proxy for threshold. For our example neuron, the threshold rose with travel time (β = 1.53, $P < 0.01$, regression of firing rates (spikes per s) against travel time (s); **Fig. 5d**). There was also a significant, but weaker, correlation between travel time and firing rate on the last choice to stay in the patch before patch-leaving, consistent with the idea of a gradual rise-to-threshold process (β = 1.32, $P < 0.025$); the whole population showed similar effects (**Fig. 5e**). For the focal population (n = 49), threshold rose with travel time (β = 0.92, $P < 0.01$), as did previous-trial responses (β = 0.69, $P < 0.001$). We observed a rise in threshold in 37 of 102 neurons (regression, $P < 0.05$, a negative effect in one neuron, no effect in the rest).

If travel time influences threshold levels, do thresholds actually remain constant for different leaving times, after accounting for travel time? To answer this question, we measured neural responses on patch-leaving trials in five groups of travel times (0.5 to 2.5, 2.5 to 4.5, 4.5 to 6.5, 6.5 to 8.5 and 8.5 to 10.5 s). Two facts suggest that they were constant across earliness level (see above) after accounting for travel time. First, in each of these travel time groups, earliness did not influence firing rate in the pre-saccadic epoch of the patch-leaving trial ($P > 0.2$ in all cases). Second, there was no significant interaction between earliness and travel time (P = 0.67, two-way ANOVA, four levels of earliness and five travel times, main effect of earliness = 0.06 spikes per s per bin).

Finally, we considered whether travel times influence baseline firing. We reasoned that such an effect would be most apparent on the first few choices in a new patch. We found no effects of travel time on firing rate responses occurring in the first 5 s of a patch (ANOVA of normalized firing rate against travel time, P = 0.41) and a significant effect in nine neurons individually ($P < 0.05$). The frequency of these effects is not much more than would be expected by chance (n = 5.1 neurons, P = 0.067, binomial test), and the size of the effect was weak (average difference between 1- and 10-s travel times was 0.09 spikes per s). Moreover, of these nine neurons, six showed increasing firing rates with longer travel times. Our data therefore do not endorse the idea that travel times influence baseline neuronal firing rates in ACC.

DISCUSSION

Choosing when to leave a depleting resource patch is a ubiquitous natural decision problem that is central to foraging theory and behavioral ecology[2,3]. Although the brains of animals have undoubtedly been shaped by evolutionary pressures for foraging efficiency, the neural processes that mediate the simple decision to give up on one patch and move to another remain obscure. Our findings suggest that, during foraging, the primate brain computes a decision variable whose magnitude corresponds to the relative value of leaving a patch, that this value rises to a threshold associated with patch leaving, and that travel time between patches governs both the threshold and the rate at which this decision variable rises. This decision variable is represented in the firing rates of neurons in the dACC, a frontal lobe structure associated with reward monitoring and behavioral adjustment[14–16].

If the thresholding process that we propose were to occur at the level of dACC or its inputs, we would expect to see the outcome of the threshold process in the responses of dACC neurons. Instead, we observed a signal that varied continuously with time in patch, suggesting that the thresholding process occurs downstream of dACC, perhaps in FEFs or some other premotor structure. Notably, our results imply that a downstream neuron cannot judge whether it is time to leave the patch solely by querying the output of dACC; it needs to have information about travel time, likely in the form of the change in its threshold. We speculate that such control may be implemented through neuromodulatory inputs to the dACC, perhaps via dopamine or norepinephrine[36,37].

Responses of dACC neurons do not uniquely represent any single variable in the MVT equations. The MVT is a description of foraging behavior at the computational level, whereas our data support a particular mechanism by which the decision process could be implemented[38]. Thus, although we claim that our hypothesized decision variable encodes the relative value of leaving a patch, we could just as easily argue that it encodes the negative value of staying (indeed, this would be consistent with error-related theories of ACC function[16]). Because these functional mechanisms resemble those known to support basic perceptual and mnemonic decisions[10–13,34], our findings endorse the idea that the brain uses a small suite of common mechanisms to solve diverse problems in multiple domains.

The broad applicability of the MVT to such a wide array of organisms underscores the fact that dACC is unlikely to be the sole neural locus of the decision process, even in organisms with brains similar to ours. Indeed, these mechanisms may not even be limited to brains. Many organisms that lack brains, including amoebas, slime molds and plants, exhibit behavior that is consistent with the MVT[6,39,40]. We conjecture that such organisms solve the patchy foraging problem in much the same way that monkeys do, namely by maintaining and

controlling a representation of the relative value of leaving a patch. Thus, diverse organisms may solve common problems using similar algorithms that are implemented very differently[38]. Even in organisms with brains similar to ours, given the high redundancy of decision signals across brain areas in primates[41], we predict that similar signals might be observed in other regions, including the dorsolateral prefrontal cortex, lateral intraparietal area and posterior cingulate cortex, although such signals might be convolved with other information such as target location or movement metrics.

Relation to previous studies

Our findings are broadly consistent with prior studies showing that dACC monitors reward information from many sources and signals the need to adjust behavior in some manner[17,18,22,23,42–46]. Our results corroborate these earlier results and extend them in four important ways. First, we found that dACC responses vary continuously with the extent to which circumstances favor the decision to move on, even if leaving does not occur. This observation supports the idea that dACC neurons represent a scalar decision variable reflecting the relative value of leaving. The relative value of switching behavior was not manipulated in a previous study in which all non-switch trials were, in all important respects, the same[23]. Second, we found and measured a specific threshold at which leaving, and by extension switching, occurs. Third, we identified two mechanisms by which exogenous factors govern patch-leaving behavior. Finally, we found that neuronal activity in dACC promotes disengagement in a relatively natural task that is directly modeled on real-world foraging situations. These results directly link dACC neuronal activity to behavior in a naturalistic context and may extend to situations outside the laboratory in which dACC dysfunction has been implicated, including addiction, depression, obsessive-compulsive disorder and Tourette syndrome[25–27].

At first glance, our results appear to contradict those obtained by an earlier study that reported increasing firing rates of dACC neurons with increasing proximity to reward[24], whereas we found increasing firing rates in anticipation of sequentially smaller rewards. We believe that the two sets of findings are fully concordant. In our study, firing rates rose as the monkey approached the decision to abandon the current patch for a new one. In the earlier study, firing rates increased as the monkey neared the rewarded action. In both cases, firing rates of dACC neurons marked progression through a sequence of actions toward a salient behavioral event: the reward in their task, patch-leaving in ours. Together, our results suggest a broader view, namely that dACC neurons do not signal reward value *per se*, but rather that their responses encode an abstract decision variable that is suitable for guiding a variety of different modifications in behavior, whether generated endogenously or exogenously.

Our findings may also initially appear to contradict results from our earlier studies of dACC neurons[21,22]. Previously, we found that the firing rates of dACC neurons reflected both real and fictive rewards, and generally did so with higher firing rates for larger rewards[22]. In that study, however, large rewards, both real and fictive, promoted a behavioral strategy that led to potentially larger rewards. Indeed, trial-to-trial variations in firing rate in that study positively covaried with likelihood of adjusting behavior on the next trial. In other words, higher firing rates in both studies predicted the likelihood that the monkey would successfully incorporate new information about the world into an ongoing decision to change behavior.

In another study, we found that neuronal activity in dACC showed weak, but significant, selectivity for saccade direction, in addition to anticipated real and fictive reward size[21]. In contrast, here we found no evidence for spatial selectivity in neuronal responses. These differences likely reflect task design. We used eight targets in the previous study, but only two saccade targets here, thus weakening our sensitivity to spatial selectivity, especially bimodal tuning common in dACC[21]. Also, the task used in the prior study demanded that monkeys carefully distinguish adjacent, physically similar targets to evaluate the associated reward outcomes, whereas the two targets were widely separated and physically distinct in the current study. We hypothesize that the greater demand for attentional resources associated with spatial locations in that task accentuated spatial tuning in the earlier study.

Conclusion

Our virtual foraging task is an ersatz idealization of a real patchy foraging environment. Given that foraging often involves physical effort, these results are only a first step on the path to understanding real foraging decisions. Because of the clear links between ACC function and effortful choices, dACC seems particularly well positioned to guide real-world foraging choices and is likely involved in these choices. Thus, we believe that our results provide a useful advance toward understanding natural value–based decisions and forge a critical link between systems neurobiology and behavioral ecology.

Animals' bodies are have evolved to efficiently exploit the resources in their environments. Natural selection has also acted on the nervous systems of these animals to enable the adaptive action of their bodies. Few studies have linked neural computations to specific types of naturally occurring foraging decisions. Our study portends a more general understanding of prey selection, diet selection and more complex foraging problems[3,47]. Ultimately, these results endorse the unification of theoretical and experimental work in the ecological and neural sciences[48].

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank S. Heilbronner for comments on design, analysis and writing. This research was supported by US National Institutes of Health grant R01EY013496 (M.L.P.), a Fellowship from the Tourette Syndrome Association (B.Y.H.) and US National Institutes of Health grant K99 DA027718-01 (B.Y.H.).

AUTHOR CONTRIBUTIONS

B.Y.H. designed the experiment and collected the data. B.Y.H. and J.M.P. contributed to data analysis. B.Y.H., J.M.P. and M.L.P. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Prins, H.H.T. *Ecology and Behavior of the African Buffalo: Social Inequality and Decision Making* (Chapman and Hall, London, 1996).
2. Charnov, E.L. Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* **9**, 129–136 (1976).
3. Stephens, D.W. & Krebs, J.R. *Foraging Theory* (Princeton University Press, Princeton, NJ, 1986).
4. Bendesky, A., Tsunozaki, M., Rockman, M.V., Kruglyak, L. & Bargmann, C.I. Catecholamine receptor polymorphisms affect decision-making in C. elegans. *Nature* **472**, 313–318 (2011).
5. Thompson, D. & Fedak, M.A. How long should a dive last? A simple model of foraging decisions by breath-hold divers in a patchy environment. *Anim. Behav.* **61**, 287–296 (2001).
6. McNickle, G.G. & Cahill, J.F. Jr. Plant root growth and the marginal value theorem. *Proc. Natl. Acad. Sci. USA* **106**, 4747–4751 (2009).
7. Smith, E.A. & Winterhalder, B. *Evolutionary Ecology and Human Behavior* (de Gruyer, New York, 1992).

8. Agetsuma, N. Simulation of patch use by monkeys using operant conditioning. *J. Ethology* **16**, 49–55 (1999).
9. Gold, J.I. & Shadlen, M.N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
10. Gold, J.I. & Shadlen, M.N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36**, 299–308 (2002).
11. Hanes, D.P. & Schall, J.D. Neural control of voluntary movement initiation. *Science* **274**, 427–430 (1996).
12. Schall, J.D. On building a bridge between brain and behavior. *Annu. Rev. Psychol.* **55**, 23–50 (2004).
13. Carpenter, R.H.S. *Movements of the Eyes* (Pion, London, 1988).
14. Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J. & Rushworth, M.F. Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* **9**, 940–947 (2006).
15. Rushworth, M.F. & Behrens, T.E. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* **11**, 389–397 (2008).
16. Holroyd, C.B. & Coles, M.G. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* **109**, 679–709 (2002).
17. Seo, H. & Lee, D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27**, 8366–8377 (2007).
18. Quilodran, R., Rothe, M. & Procyk, E. Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* **57**, 314–325 (2008).
19. Ito, S., Stuphorn, V., Brown, J.W. & Schall, J.D. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* **302**, 120–122 (2003).
20. Williams, Z.M., Bush, G., Rauch, S.L., Cosgrove, G.R. & Eskandar, E.N. Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat. Neurosci.* **7**, 1370–1375 (2004).
21. Hayden, B.Y. & Platt, M.L. Neurons in anterior cingulate cortex multiplex information about reward and action. *J. Neurosci.* **30**, 3339–3346 (2010).
22. Hayden, B.Y., Pearson, J.M. & Platt, M.L. Fictive reward signals in the anterior cingulate cortex. *Science* **324**, 948–950 (2009).
23. Shima, K. & Tanji, J. Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* **282**, 1335–1338 (1998).
24. Shidara, M. & Richmond, B.J. Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* **296**, 1709–1711 (2002).
25. Baler, R.D. & Volkow, N.D. Drug addiction: the neurobiology of disrupted self-control. *Trends Mol. Med.* **12**, 559–566 (2006).
26. Stern, E. *et al.* A functional neuroanatomy of tics in Tourette syndrome. *Arch. Gen. Psychiatry* **57**, 741–748 (2000).
27. Devinsky, O., Morrell, M.J. & Vogt, B.A. Contributions of anterior cingulate cortex to behavior. *Brain* **118**, 279–306 (1995).
28. Stephens, D.W. & Anderson, D. The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. *Behav. Ecol.* **12**, 330–339 (2001).
29. Evans, T.A. & Beran, M.J. Delay of gratification and delay maintenance by rhesus macaques (*Macaca mulatta*). *J. Gen. Psychol.* **134**, 199–216 (2007).
30. Kim, S., Hwang, J. & Lee, D. Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron* **59**, 161–172 (2008).
31. Louie, K. & Glimcher, P.W. Separating value from choice: delay discounting activity in the lateral intraparietal area. *J. Neurosci.* **30**, 5498–5507 (2010).
32. Kahneman, D., Knetsch, J.L. & Thaler, R.H. Anomalies: the endowment effect, loss aversion, and the status quo bias. *J. Econ. Perspect.* **5**, 193–206 (1991).
33. Mazur, J.E. An adjusting procedure for studying delayed reinforcement. in *Quantitative Analyses of Behavior, vol 5. The Effect of Delay and Intervening Events on Reinforcement Value* (eds. Commons, M.L., Mazur, J.E., Nevin, J.A. & Rachlin, H.) (Erlbaum, Mahway, New Jersey, 1987).
34. Roitman, J.D. & Shadlen, M.N. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* **22**, 9475–9489 (2002).
35. Horwitz, G.D., Batista, A.P. & Newsome, W.T. Representation of an abstract perceptual decision in macaque superior colliculus. *J. Neurophysiol.* **91**, 2281–2296 (2004).
36. Tobler, P.N., Fiorillo, C.D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005).
37. Yu, A.J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–692 (2005).
38. Marr, D.C. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, New York, 1982).
39. Latty, T. & Beekman, M. Food quality affects search strategy in the acellular slime mold, *Physarum polycephalum*. *Behav. Ecol.* **20**, 1160–1167 (2009).
40. Bonser, R., Wright, P.J., Bament, S. & Chukwu, U.O. Optimal patch use by foraging workers of *Lasius fuliginosus*, *L. niger* and *Myrmica ruginodis*. *Ecol. Entomol.* **23**, 15–21 (1998).
41. Wallis, J.D. & Kennerley, S.W. Heterogeneous reward signals in prefrontal cortex. *Curr. Opin. Neurobiol.* **20**, 191–198 (2010).
42. Sallet, J. *et al.* Expectations, gains, and losses in the anterior cingulate cortex. *Cogn. Affect. Behav. Neurosci.* **7**, 327–336 (2007).
43. Kennerley, S.W., Dahmubed, A.F., Lara, A.H. & Wallis, J.D. Neurons in the frontal lobe encode the value of multiple decision variables. *J. Cogn. Neurosci.* **21**, 1162–1178 (2008).
44. Matsumoto, M., Matsumoto, K., Abe, H. & Tanaka, K. Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* **10**, 647–656 (2007).
45. Amiez, C., Joseph, J.P. & Procyk, E. Reward encoding in the monkey anterior cingulate cortex. *Cereb. Cortex* **16**, 1040–1055 (2006).
46. Hayden, B.Y., Heilbronner, S.R., Pearson, J.M. & Platt, M.L. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *J. Neurosci.* **31**, 4178–4187 (2011).
47. Stephens, D.W., Brown, J.S. & Ydenberg, R.C. *Foraging: Behavior and Ecology* (University of Chicago Press, Chicago, 2007).
48. Wilson, E.O. *Consilience: The Unity of Knowledge* (Knopf, New York, 1998).

ONLINE METHODS

Surgical procedures. All procedures were approved by the Duke University Institutional Animal Care and Use Committee and were designed and conducted in compliance with the Public Health Service's Guide for the Care and Use of Animals. Two male rhesus monkeys (*Macaca mulatta*) served as subjects. Initially, a head-holding prosthesis was implanted in both animals using standard surgical techniques. Six weeks later, animals were habituated to laboratory conditions and trained to perform oculomotor tasks for liquid reward. A second surgical procedure was then performed to place a plastic recording chamber (Crist Instruments) over dorsal anterior cingulate cortex. Animals received analgesics and antibiotics after all surgeries. Throughout both behavioral and physiological recording sessions, the chamber was kept sterile with regular antibiotic washes and sealed with sterile Teflon caps.

Behavioral task. Monkeys were placed on controlled access to fluid outside of experimental sessions to motivate behavior. Horizontal and vertical eye positions were sampled at 1,000 Hz by an infrared eye-monitoring camera system (SR Research). Stimuli were controlled by a computer running MATLAB (MathWorks) with Psychtoolbox and Eyelink Toolbox[49,50]. Visual stimuli were small colored rectangles on a computer monitor placed directly in front of the animal and centered on his eyes. A standard solenoid valve controlled the duration of juice delivery. We estimated the precision of fluid volume delivered by the solenoid across the range of open time commands used in this study. Given that the s.e.m. in volume increased with the mean, we calculated the coefficient of variation as a measure of precision, and this value was 0.29. Because of this small uncertainty in liquid volume delivered, the reported values should be taken as approximate.

Our task was designed to mimic the critical elements of the patch-leaving task studied in foraging theory[2,3]. On each trial, a small yellow square appeared in the center of the monitor (<0.5 degrees of visual angle). Following its appearance, the monkey aligned gaze with the square to indicate his readiness to begin the trial. Once the monkey acquired fixation (± 0.5 degrees), two eccentric targets appeared, one on the left and one on the right (300 pixels from fixation). One target was a small blue rectangle (remain in patch option) and the other was a gray rectangle (leave patch option) that was the same width (80 pixels) and taller (precise height depending on condition). One pixel is approximately equal to 0.024 degrees of visual angle.

Following a 500-ms delay, the central fixation square extinguished and the monkey was free to select either of the two targets by shifting gaze to it (±2 degrees from the center of the rectangle). Following choice of either target, the rectangle began to shrink at a constant rate (65 pixels per s) until it disappeared, a reward was given (if the blue 'stay' target was chosen), and the ITI began (1 s). Because the rate of bar shrinking was constant, the height of the bar provided an unambiguous cue to the delays associated with the two options on every trial.

The delay associated with the blue stay (that is, remain in patch) rectangle occurred before the reward and was isomorphic to the handling time in foraging decisions (set at 400 ms in the task used for recording, that is, the variable travel time task). The delay associated with the gray 'switch' (that is, leave the patch) rectangle was analogous to the travel time in foraging decisions (ranging from 0.5 to 10.5 s in this experiment). It was set at a random value on each patch, but did not vary in a patch. The fixed delay (ITI) between trials was uncued, but was always the same (1 s). This fixed delay is isomorphic to a fixed search time in foraging theory. Errors (14% of trials) consisted of either early fixation breaks during the brief hold period (89% of the error trials) or failures to initiate the trial within 30 s (the remaining 11% of error trials). In either case, an error was followed by the presentation of a dark green square on the center of the screen (an error cue) and a 3-s timeout period. We defined search time as the average total time per trial in each patch, excluding handling time. Search time included the ITI, the saccade time, and any other sources of delay or variability.

Following the first choice of the blue stay rectangle in each patch, the monkey received 306 µl of water. On subsequent choices of the 'stay' target, the reward decreased by 19 µl (although we introduced a small variance in this amount, ε = s.e.m. of 1.9 µl). If the monkey continued to choose the blue stay option, its value would eventually reach 0 and remain 0 thereafter. On choosing the gray switch rectangle, the location of the blue and gray rectangles would alternate and the value of the blue rectangle would reset to 306 µl. On choosing the gray rectangle, the size of the gray rectangle and the associated travel time would reset to a new value, chosen from a uniform distribution between 0.5 and 10.5 s.

Rewards were not completely deterministic in this task, although the variability was quite small. Stochasticity came from two sources. First, there was an inherent variability in the reward amounts, added to keep the task engaging, and to encourage the monkeys to actively monitor rewards, rather than simply learn to count a certain number of trials. Second, there is necessarily some uncontrolled, but small, variability associated with the juice dispenser. We estimate that this variability is less than that generated by the variability that we deliberately added.

To plot residence time as a function of travel time (**Fig. 2**), we computed residence time for each patch. Residence time includes all time in patch from arrival to the decision to leave, and thus included handling time and search time and its constituents, saccade latencies, reward duration, etc. On the occasional trial in which the monkey chose to leave the patch immediately on the first trial, we assigned the residence time a value of 0 s.

The critical equation in the MVT is equation (2) from ref. 3.

$$E_{\mathrm{n}} = \frac{\sum P_i \times g_i(T_i) - t \times E_T}{t + \sum P_i \times T_i}$$

In this equation, E_{n} is net energy intake (the quantity to be maximized), P_i is the proportion of patches to be visited of a given type, g_i is the assimilated energy corrected for the cost of searching (assumed to be zero in this task), T is the hunting time (in this case, patch residence time), t is the travel time, and E_T is the energy cost per unit time for traveling.

We adopted a simplified version of the more general situation encompassed by the MVT. Two simplifying assumptions are especially important. First, all patches are identical. Second, there is no energy cost associated with travel, only a time cost.

Thus, intake rate for a given patch is given by $E_{\mathrm{n}} = \frac{g(T)}{t+T}$. According to the MVT, intake rate is maximized when $E^*_{\mathrm{n}} = \frac{\delta g(T)}{\delta T}$, where E^*_{n} refers to reward intake rate when patch residence time is maximized. At this point, the marginal intake rate matches the average intake rate for the habitat.

Total patch residence time depends strongly on several factors that are beyond the control of the monkeys. These include handling time and ITI, but also the monkey's own deliberation time, reaction times, saccade latency, etc. As these variables potentially lie outside the control of the monkeys, we assumed that they treat them as fixed quantities. Reward as a function of average patch residence time is shown in **Figure 1b**. Because the reward function was somewhat stochastic, we used a repeated randomized algorithm to calculate rate-maximizing behavior. Specifically, we simulated the total reward harvested over 5,000 trials for each of 100 patch residence times, ranging from 0 to 75 s. To reduce any possible effects of noise, we repeated each simulation 1,000 times. We determined the rate-maximizing value by locating the peak of the resulting intake curve. This randomization process eliminates biases emanating from the variability in reward values.

Microelectrode recording techniques. Single electrodes (Frederick Haer) were lowered using a microdrive (Kopf) until the waveform of a single (1–3) neuron(s) was isolated. Individual action potentials were identified by standard criteria and isolated on a Plexon system (Plexon). Neurons were selected for study on the basis of the quality of isolation, and not on task-related response properties.

We approached dACC through a standard Teflon recording grid (Crist Instruments). dACC was identified by structural magnetic resonance images taken before the experiment. Neuroimaging was performed at the Center for Advanced Magnetic Development at DUMC, on a 3T Siemens Medical Systems Trio MR Imaging Instrument using 1-mm slices. We confirmed that electrodes were in dACC using stereotactic measurements, as well as by listening for characteristic sounds of white and gray matter during recording. Our recordings were likely to have come from area 24, and especially the dorsal and ventral banks of the anterior cingulate sulcus (see **Supplementary Fig. 8** for reconstructions of recording sites).

Prior to beginning formal experiments, we performed several exploratory recording sessions to map the physiological response properties of neurons accessible through our recording chamber. We were able to distinguish white from gray matter by the presence and absence of neural activity. During these mapping sessions, we were able to identify both the dorsoventral and mediolateral

doi:10.1038/nn.2856

extent of the cingulate sulcus. Neurons were recorded in the same animals and at the same grid positions as used in two previous studies examining the role of dACC in representing information about rewards[22] and action[21]. It is therefore likely that neurons in both studies come from the same brain region. Moreover, as in these previous studies, we performed no pre-selection on neurons, aside from the natural biases in single-unit recordings toward larger and higher firing rate neurons[21].

Analysis. PSTHs were constructed by aligning spike rasters to trial events and averaging firing rates across multiple trials. Firing rates were calculated in 10-ms bins, but were generally analyzed in longer epochs. For display, PSTHs were smoothed using a 100-ms running boxcar. Neuronal activity was temporally aligned to the end of the saccade indicating the animal's choice (time zero). To normalize neuronal activity, we calculated the average firing rate for each neuron during a pre-trial epoch, defined as a 0.5-ms period beginning 0.75 s before the beginning of each trial. We then divided neural activity by this value for each neuron.

49. Brainard, D.H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
50. Cornelissen, F.W., Peters, E. & Palmer, J. The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behav. Res. Methods Instrum. Comput.* **34**, 613–617 (2002).

Current Biology 19, 1532–1537, September 29, 2009 ©2009 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2009.07.048

Report

Neurons in Posterior Cingulate Cortex Signal Exploratory Decisions in a Dynamic Multioption Choice Task

John M. Pearson,[1,*] Benjamin Y. Hayden,[1] Sridhar Raghavachari,[1] and Michael L. Platt[1,2]
[1]Department of Neurobiology, Duke University School of Medicine and Center for Neuroeconomic Studies
[2]Center for Cognitive Neuroscience and Department of Evolutionary Anthropology
Duke University, Durham, NC 27710, USA

Summary

In dynamic environments, adaptive behavior requires striking a balance between harvesting currently available rewards (exploitation) and gathering information about alternative options (exploration) [1–4]. Such strategic decisions should incorporate not only recent reward history, but also opportunity costs and environmental statistics. Previous neuroimaging [5–8] and neurophysiological [9–13] studies have implicated orbitofrontal cortex, anterior cingulate cortex, and ventral striatum in distinguishing between bouts of exploration and exploitation. Nonetheless, the neuronal mechanisms that underlie strategy selection remain poorly understood. We hypothesized that posterior cingulate cortex (CGp), an area linking reward processing, attention [14], memory [15, 16], and motor control systems [17], mediates the integration of variables such as reward [18], uncertainty [19], and target location [20] that underlie this dynamic balance. Here we show that CGp neurons distinguish between exploratory and exploitative decisions made by monkeys in a dynamic foraging task. Moreover, firing rates of these neurons predict in graded fashion the strategy most likely to be selected on upcoming trials. This encoding is distinct from switching between targets and is independent of the absolute magnitudes of rewards. These observations implicate CGp in the integration of individual outcomes across decision making and the modification of strategy in dynamic environments.

Results

To probe the neuronal processes mediating the strategic balance of immediate reward and information acquisition, we recorded the activity of single cingulate cortex (CGp) neurons in two rhesus macaques performing a "restless" variant of the four-armed bandit for juice rewards [3, 5] (Figure 1). This variant provides a high level of environmental variability with a behaviorally tractable number of options. On each trial, monkeys chose one of four targets whose payoffs were randomly selected from distributions centered about their values on the previous trial. Once a target was chosen, monkeys in principle had perfect knowledge of its present value (there was no added variance in payouts), though the values of all targets changed each trial. As a result, monkeys had to select an option to learn its current value and integrate this information with their statistical knowledge of the environment to predict its relative value on upcoming trials.

Both monkeys were highly adept at optimizing reward. They earned 92% and 91%, respectively, of the total reward that would have been earned by an omniscient observer. Nevertheless, despite this high level of performance, a perfectly greedy decision maker, focused on the option with highest immediate value, would have harvested more, though not all, available reward (see Supplemental Data available online).

More importantly, nothing intrinsic to the task design serves to distinguish exploratory from exploitative decisions. On each trial, both monkeys simply selected among the four available options and received a reward. As a result, individual decisions must be classified as exploratory or exploitative according to a model-based analysis of each monkey's behavior, with model parameters chosen to maximize the likelihood of observed choices. We report here only results based on our best-fitting Kalman filter model, though results were similar for other models as well (see Supplemental Data).

We analyzed the firing rates of 83 single neurons in CGp in both monkeys performing the four-armed bandit task (59 from monkey N and 24 from monkey B). We focused on two trial epochs, a 2 s decision epoch (DE; 1 s before trial initiation extending to juice delivery) and a 2 s postreward evaluation epoch (EE; from the offset of juice delivery through the intertrial interval). Analyses based on mean firing rates in each epoch readily identified neurons that discriminated between the two strategies (14%, n = 12/83, DE; 16%, n = 13/83, EE; $p < 0.05$, Mann-Whitney U test), with 22% of neurons doing so in at least one epoch (n = 18/83; $p < 0.025$, Bonferroni-corrected Mann-Whitney U test).

Figure 2A depicts the average firing rate of a single neuron on trials classified as either exploratory or exploitative. Responses on exploit trials were significantly higher in both decision and evaluation epochs (Mann-Whitney U test, $p < 0.01$). In contrast, the neuron whose activity is plotted in Figure 2B was more responsive on exploratory trials in both epochs (Mann-Whitney U test, $p < 0.01$). Although the population as a whole exhibited slightly higher firing on exploratory trials in both epochs (modulation index = 0.0084 [DE], 0.0026 [EE]), the population of cells with significant modulation was mixed (modulation index = 0.046 [DE], $\geq$0.011 [EE]), indicating heterogeneity in single-cell responses to the different strategies. Thus, firing rates in CGp distinguish between upcoming exploratory and exploitative decisions in the epoch leading up to selection and continue to reflect that choice during the postreward delay.

We previously reported that responses of CGp neurons predict impending switches from one option to another on the next trial in a simple two-alternative task [11]. Based on that finding, we hypothesized that CGp neurons would also carry predictive information about more general impending choices of strategy. To test this hypothesis, we regressed the probability of exploration on the upcoming decision as a function of observed firing rate for each neuron. Of the 83 neurons in our sample, about 16% showed significant correlations between firing rate during the decision epoch and the probability of exploration on the ensuing choice [n = 13, $p < 0.05$,

*Correspondence: pearson@neuro.duke.edu

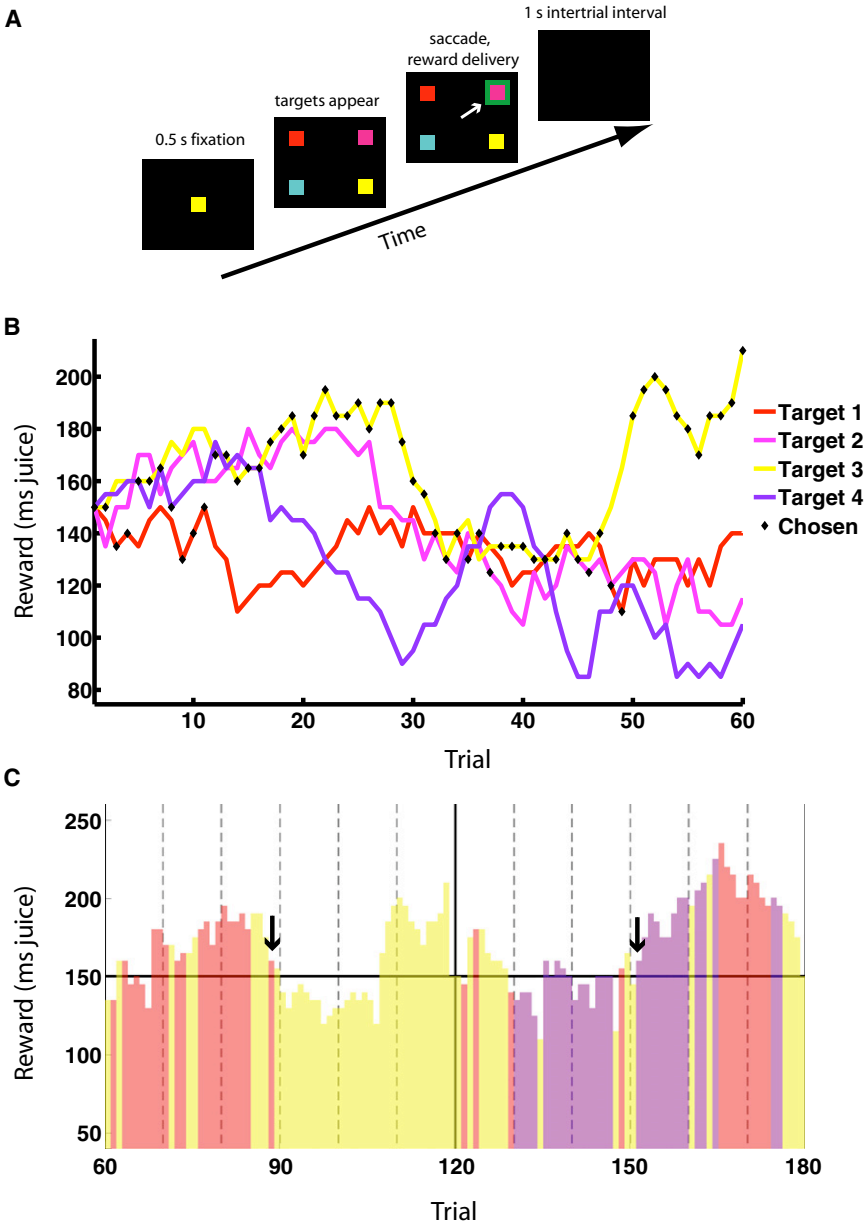

Figure 1. Task and Example Reward Schedule Used to Study the Explore/Exploit Dilemma

(A) Schematic of the four-armed bandit task. Following a 0.5 s fixation period, the central cue disappears, replaced by four colored targets. Subjects indicate choices by shifting gaze to targets, after which the chosen target is highlighted in green for 1 s and a juice reward is delivered. Consecutive trials are separated by a 1 s intertrial interval. Every 60 trials, a block change cue appears, and all target values are reset to the mean reward value.

(B) Sample payouts and choices for the four options over a single block. Reward values for each target follow a random walk with fixed standard deviation for step size, biased toward the mean of 150 ms. Black diamonds indicate choices made by the monkey during the given block.

(C) Sample monkey B choice behavior over two blocks of the four-armed bandit task. Bar colors indicate target chosen; bar heights indicate the values of rewards received. The horizontal line indicates the mean reward value. Monkeys exhibit bouts of exploitation of favorable targets with exploration of alternatives. Arrows indicate trials that might plausibly be classified as either exploratory or exploitative, depending on the behavioral model used. Both involve a change in target selected (action switch) but also a return to a target with high remembered value and so might be classified as exploitative.

Mann-Whitney U test; p(n > 12) < 0.001, binomial test]. Even more importantly, 16% of neurons showed a correlation between firing during EE and the probability of exploration on the following trial [n = 13 p < 0.05, Mann-Whitney U test; p(n > 12) < 0.001, binomial test], suggesting that CGp differentially signals the probability of strategic decisions within a block of trials. Figures 2C and 2D depict the separate population averages for the subsets of cells whose activity correlated negatively (n = 6) and positively (n = 7) with exploration. Average response for each of these two groups of neurons strongly predicted probability of impending strategic choices in a graded fashion.

One potential confound of these results arises from the link between exploitation and the likelihood of increased reward. Because we might expect that exploitative choices, on average, yielded higher rewards, a possible alternative interpretation of the present data is that effects of strategy on neuronal activity are reducible entirely to neuronal sensitivity to reward value. To investigate this possibility, we calculated and fit reward size tuning curves for each of our 83 neurons (Figures 3A and 3B). Consistent with previous studies [11, 18], we found that the firing rates of CGp neurons during the decision epoch varied with the amount of reward received on the previous trial (DE, n = 39, p < 0.05, F test for quadratic regression) and that firing rates in the evaluation epoch varied with the amount just received (EE, n = 44, p < 0.05, F test). Over the range of experienced reward values (50–350 μl), we found a heterogeneity of tuning curves: some were linear (n = 14 positive, 6 negative, DE; n = 18 positive, n = 9 negative, EE; p < 0.05 nonzero regression coefficient), whereas others were U-shaped, both concave up (n = 8, DE; n = 8, EE; p < 0.05 nonzero regression coefficient) and concave down (n = 11, DE; n = 9, EE; p < 0.05 nonzero regression coefficient). We therefore restricted our next series of analyses to those trials where monkeys adopted different strategies but received the same amount of reward. We found that 12% [n = 10; p(n > 9) < 0.001, binomial test] of recorded neurons still showed different mean firing rates on explore versus exploit trials (p < 0.01, Bonferroni-corrected Mann-Whitney U test). Data for an example neuron showing this effect are shown in Figures 3D–3F. Here, the middle third of reward values have been subdivided into three categories (medium-low, medium-medium, and medium-high), and neuronal firing is plotted as a function of time for both explore and exploit trials, controlled for received reward. This neuron, like many others in our population, showed clear

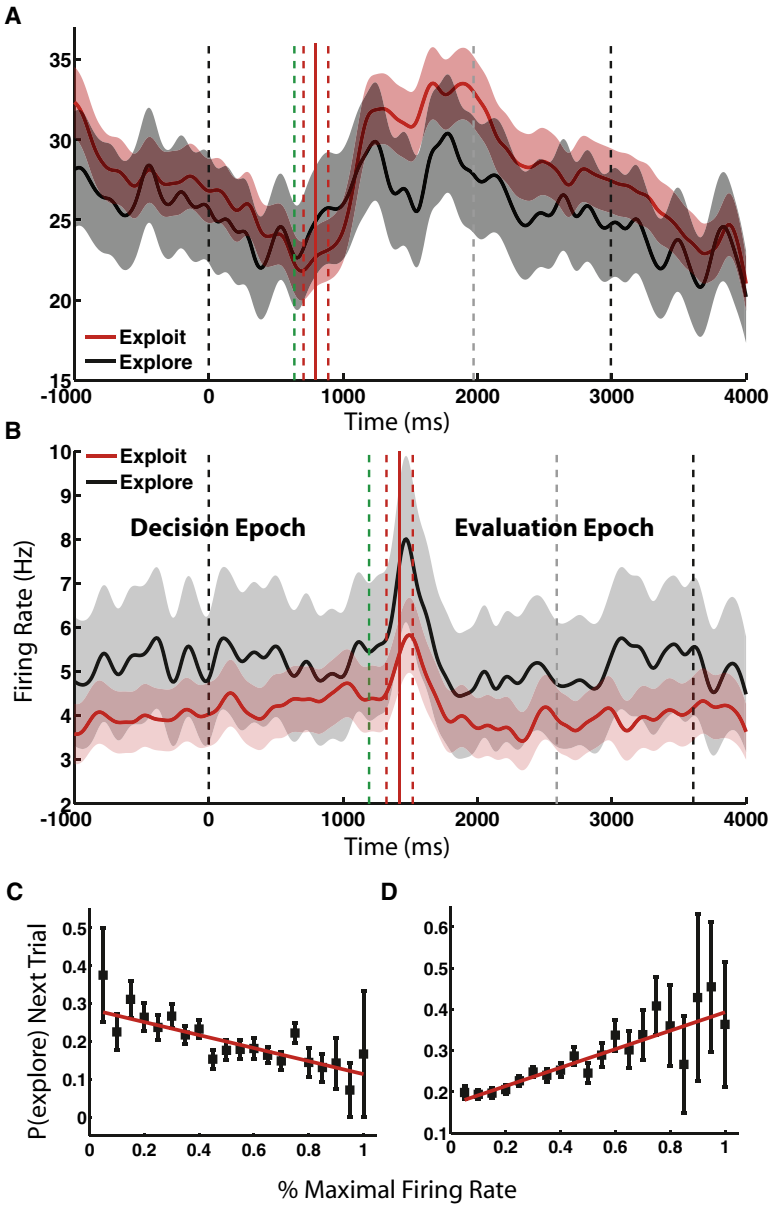

Figure 2. Neurons in CGp Distinguish between Exploration and Exploitation

(A and B) Sample neurons (neurons 62 and 12) with significant differences in firing during explore and exploit trials for both the decision and evaluation epochs, with firing aligned to trial onset. Individual cells might prefer either explore (black) or exploit (red) trials. The task begins at time 0. Onset of "go" cue (dashed green line), reward delivery (solid red line), beginning of intertrial interval (dashed gray line), and end of trial (second dashed black line) are mean times. Dashed red lines indicate ± one standard deviation in reward onset. Shaded areas represent standard error of the mean (SEM).

(C and D) Neurons in CGp encode probability of exploring on the next trial. Points are probabilities of exploring next trial as a function of percent maximal firing rate in the decision epoch, averaged over negatively and positively tuned populations of neurons (C and D, respectively). Error bars represent SEM.

sensitivity to strategy even when we controlled for the value of the reward the monkeys received.

Two other possible confounds arise from the known spatial tuning of CGp and the close relationship between exploration and simply switching between targets. As reported previously [21], we found that 63% of neurons were tuned for the location of the target chosen ($n = 52/83$, $p < 0.05$, one-way analysis of variance of mean firing rates for each target over all trials; see Supplemental Data). Across the population, 39% of neurons were significantly tuned for both reward size and target location [18], whereas 23% were tuned for neither (EE; 34% and 24%, respectively, in DE). However, the population as a whole showed no consistent target tuning across trials ($p > 0.2$, one-sample t test for contralateral and upper-hemifield tuning indices). In the case of target switching, because repeatedly choosing a poor target is not necessarily exploitative (if higher reward has recently been sampled elsewhere), there is not a strict one-to-one correspondence between exploitation and perseveration or between exploration and switching. As a result, our task design allows for the possibility of disambiguating these phenomena.

Indeed, a partial correlation analysis of neuronal firing rates and upcoming decisions (firing rate in DE for decision in current trial; firing rate in EE for decision in following trial) that controlled for the effects of reward tuning, spatial tuning, target switching, and previous explore/exploit decision revealed significant correlations in 12% of neurons [$n = 11$, DE; $n = 10$, EE; $p < 0.05$ Spearman partial correlation; $p(n > 9) < 0.01$, binomial test]. Thus, even when all known effects on firing rate were accounted for, a significant number of neurons still exhibited clear predictive correlations with upcoming strategy. Collectively, these results indicate not only that single neurons in CGp receive information about both previous rewards and previous choices [19] and maintain that information across trials [11, 19], as reported previously, but that these same neurons also carry signals related to dynamic changes in choice strategy in a multiplexed format.

Discussion

We found that, when choosing among multiple targets whose relative values changed dynamically, neurons in posterior cingulate cortex signaled the distinction between trials on which monkeys pursued an exploratory rather than an exploitative strategy. This signal was robust against classifications of trials based on differing models of behavior, including a perfectly greedy strategy and a simple heuristic based on comparison to a reward threshold (see Supplemental Data). More importantly, single neurons signaled in graded fashion the probability of pursuing each strategy on upcoming trials.

Previous work has shown that CGp neurons are sensitive to reward [18], risk [19], and option switching [11] and integrate this information across multiple trials [11], but the present study generalizes the decision environment to one in which exploration and exploitation are distinguishable from a simple "win-stay/lose-shift" heuristic [11, 22] based only on the most recent reward received and in which outcomes must be evaluated in light of multiple options with dynamically changing rewards. As a result, relatively bad outcomes in rich environments might be acceptable under circumstances where all

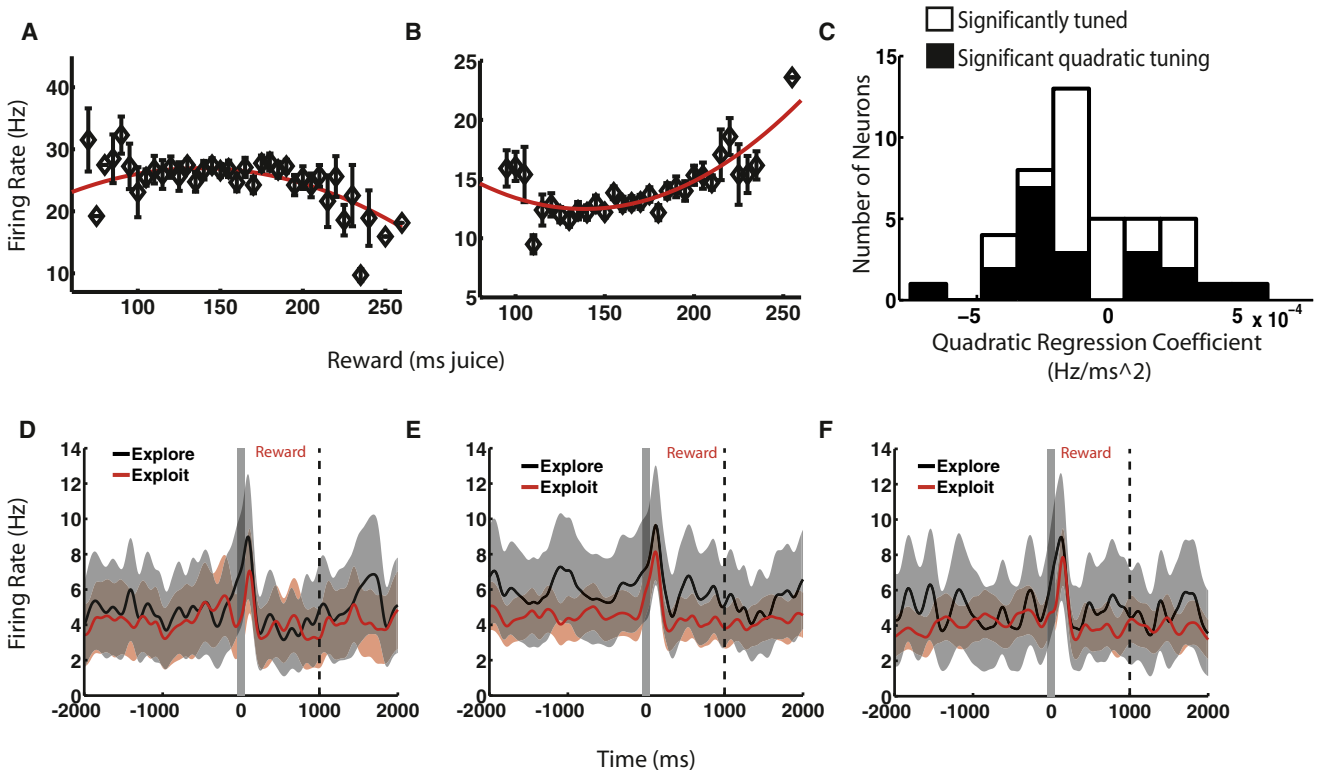

Figure 3. Reward Sensitivity of CGp Neurons Does Not Explain Sensitivity to Strategy

(A and B) Sample tuning curves from single neurons, showing firing as a function of reward received. Neurons display both positively and negatively monotonic tuning, as well as parabolic tuning (both concave up and concave down).

(C) Histogram of quadratic regression coefficients. Positive values are concave up, as in (B). White bars represent neurons with significant overall regressions, as determined by F test. Black bars represent the subset of this population with significant coefficients for quadratic tuning.

(D–F) Sample neuron (cell 12) firing rates on explore and exploit trials, controlled for low, medium, and high reward and aligned to reward onset. Bins are defined as follows: low, 115–135 ms; medium, 140–160 ms; high, 165–185 ms. In (D), the low case, the difference in mean reward sizes between explore and exploit conditions is 5 ms; in (E) and (F), the difference is less than 2 ms. Shaded regions represent SEM.

alternatives are poor and searching for better options necessitates choosing among several competing alternatives, each with a distinct reward history. Strategic decisions in such an environment thus require greater abstraction and integration of information than in comparatively static contexts because no single variable (or single trial) contains sufficient information on which to base a decision.

Together, these results invite the hypothesis that CGp is part of a network that monitors the outcomes of individual decisions and integrates that information into higher-level strategies spanning multiple choices. Thus, although the prolonged time courses of firing-rate changes in CGp are unlikely to be responsible for decisions on individual trials, their tonic activity levels might be responsible for encoding the gradual accumulation of information that gives rises to changes in strategy. However, we do not expect that this integration of single-trial history with strategic information is limited to CGp, nor is this its sole function. For example, Daw et al. [5] found that exploratory decisions are associated with activation in frontal polar cortex and intraparietal sulcus, whereas exploitative decisions are associated with activity in striatum. We speculate that information about individual rewards and reward predictions is initially computed in striatum, orbitofrontal cortex, and medial prefrontal cortex; subsequently combined with recent reward outcome and choice history and maintained online in CGp; and finally passed to the anterior cingulate cortex (ACC), where it is utilized in the selection of appropriate actions. Moreover, reciprocal connections between ACC and CGp might play a role in learning which combinations of single-trial variables are most relevant when deciding among strategies to maximize reward.

In this framework, recent reward history, computational difficulty, stimulus novelty, memory load, and the statistics of the environment are distilled into a small number of task-related decision variables for the purposes of encoding and selecting among potential actions. Thus, individual neurons that compose this network would be expected to display sensitivity to the many single-trial variables like risk, reward, and spatial location that serve as its inputs. We found that these variables are represented multimodally by neurons in CGp. As a result, we suspect that CGp might play a key role in the process of learning the combinations of stimuli and accumulated statistics most relevant to making decisions, analogous to its role in simple conditioning [23–25]. This is in keeping with our observation of increased firing rates in response to block boundaries, near-threshold decisions, and aborted trials in the bandit task, as also observed in ACC [12]. If so, CGp dysfunction might be related to deficiencies in memory-guided learning and action selection observed in disorders like Alzheimer's disease and obsessive-compulsive disorder, and its proper function might be crucial to the flexible adaptation of strategy in response to changing environments.

Experimental Procedures

Surgical Procedures

All procedures were approved by the Duke University Institutional Animal Care and Use Committee and were conducted in compliance with the Public Health Service's Guide for the Care and Use of Animals. Two rhesus monkeys (*Macaca mulatta*) served as test subjects for recording. A small prosthesis and a stainless steel recording chamber were attached to the calvarium. The chamber was placed over CGp at the intersection of the interaural and midsagittal planes. Animals were habituated to laboratory conditions and trained to perform oculomotor tasks for liquid reward. Animals received analgesics and antibiotics after all surgeries. The chamber was kept sterile with antibiotic washes and sealed with sterile caps.

Behavioral Techniques

Monkeys were familiar with the task. Eye position was sampled at 1000 Hz (camera, SR Research). Data were recorded by a computer running MATLAB (The Mathworks) with Psychtoolbox [26] and Eyelink [27]. Visual stimuli were squares (6° wide) on a computer monitor 50 cm away. A solenoid valve controlled juice delivery. Juice flavor was the same for each target.

On every trial, a central cue appeared and stayed on until the monkey fixated it. Fixation was maintained within a 1°–2° window. After a brief delay, the central cue disappeared and the four targets were displayed in the corners of the screen. Targets appeared in the same location each trial. After selection of a target, its border was illuminated and reward was delivered, followed by a 1 s intertrial interval. Rewards varied from 40 ms to 280 ms of solenoid open time in 5 ms increments (50–350 μl, in 7.5 μl increments). Juice volumes were linear in solenoid open time, and we have previously shown that monkeys discriminate juice volumes as small as 20 μl [18]. All target values began at 200 μl and reset each block. Blocks were 60 trials long and were cued by the appearance of a gray square in the center of the screen. Reward values for all targets changed each trial according to a biased random walk (see Supplemental Data).

Microelectrode Recording Techniques

Single electrodes (Frederick Haer Co.) were lowered under microdrive guidance (Kopf) until the waveforms of one to three individual neurons were isolated. Individual action potentials were identified by standard criteria and isolated on a Plexon system. Neurons were selected on the basis of the quality of isolation, but not on selectivity for the task. Recordings were made in areas 23 and 31 in the cingulate gyrus and ventral bank of the cingulate sulcus, anterior to the intersection of the marginal and horizontal rami.

Statistical Methods

We used an alpha of 0.05 as a criterion for significance. Peristimulus time histograms (PSTHs) were constructed by aligning spikes to trial events, averaging across trials, and smoothing by a Gaussian filter with 50 ms standard deviation. Shaded regions in PSTHs represent the standard error of the mean (± SEM), also Gaussian smoothed. Firing-rate modulation indices were calculated in each epoch as $m = (f_{explore} - f_{exploit})/(f_{explore} + f_{exploit})$, where f is the firing rate averaged over the relevant subset of trials. Behavioral parameters were fitted by custom scripts written with the MATLAB Optimization Toolbox (The Mathworks). Details of modeling can be found in Supplemental Data.

Firing-Rate Analysis

Analyses utilized a binary, model-based classification of choices on each trial as exploratory or exploitative (see Supplemental Data). We tested for significant differences in firing rates in both the decision and evaluation epochs as a function of the explore/exploit classification of the decision on the current trial (that is, effects on DE_n and EE_n as a function of x_n, where x_n is the binary explore/exploit variable). We also tested for predictive correlations between firing in one epoch and upcoming decision (DE_n with x_n and EE_n with x_{n+1}). To do this, we binned firing rates for each neuron into deciles of percent maximal firing and examined the percentage of exploratory decisions made subsequent to epochs with firing rates in each bin. This allowed us to construct a probability of exploration as a function of percent maximal firing, which we averaged across significant cells of each tuning.

Our reward controls were performed by grouping the 45 distinct reward values into nine bins and comparing firing rates within each bin during the evaluation epoch on explore and exploit trials. Significance levels utilize a Bonferroni correction for the number of tests performed, which varied (not all bins contained an explore or exploit trial). Our reward-controlled plots grouped the 15 middle rewards into three groups of five, denoted medium-high, medium-medium, and medium-low.

Our partial correlation analyses correlated (raw, unbinned) firing rate in a given epoch (DE_n or EE_n) with the upcoming explore/exploit decision (x_n and x_{n+1}, respectively). In each case, the correlation is controlled for spatial location (split into two variables, one for upper versus lower hemifield and one for left versus right hemifield, each taking values ± 1), previous received reward (r_{n-1} and r_n, respectively), chosen target switch (binary; s_n and s_{n+1}, respectively), and previous explore/exploit choice (x_{n-1} and x_n, respectively). Correlations were calculated as Spearman rank correlations and so allow for generic monotonic relations among variables.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, six tables, and five figures and can be found with this article online at http://www.cell.com/current-biology/supplemental/S0960-9822(09)01474-2.

Acknowledgments

This work was supported by National Institute on Drug Abuse postdoctoral fellowship 023338-01 (B.Y.H.), National Institutes of Health grant R01EY013496 (M.L.P.), and the Duke Institute for Brain Studies (M.L.P.). We thank K. Watson for assistance in training the animals and A. Long for comments on the manuscript.

Received: April 3, 2009
Revised: July 8, 2009
Accepted: July 9, 2009
Published online: September 3, 2009

References

1. Stephens, D.W., and Krebs, J.R. (1986). Foraging Theory (Princeton, NJ: Princeton University Press).
2. Gittins, J.C. (1989). Multi-armed Bandit Allocation Indices (Chichester, NY: Wiley).
3. Whittle, P. (1988). Restless bandits: Activity allocation in a changing world. J. Appl. Probab. *25*, 287–298.
4. Berry, D.A., and Fristedt, B. (1985). Bandit Problems: Sequential Allocation of Experiments (London: Chapman and Hall).
5. Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. Nature *441*, 876–879.
6. Wittmann, B.C., Daw, N.D., Seymour, B., and Dolan, R.J. (2008). Striatal activity underlies novelty-based choice in humans. Neuron *58*, 967–973.
7. Walton, M.E., Devlin, J.T., and Rushworth, M.F. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. Nat. Neurosci. *7*, 1259–1265.
8. Yoshida, W., and Ishii, S. (2006). Resolution of uncertainty in prefrontal cortex. Neuron *50*, 781–789.
9. Shima, K., and Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. Science *282*, 1335–1338.
10. Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., and Rushworth, M.F. (2006). Optimal decision making and the anterior cingulate cortex. Nat. Neurosci. *9*, 940–947.
11. Hayden, B.Y., Nair, A.C., McCoy, A.N., and Platt, M.L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. Neuron *60*, 19–25.
12. Quilodran, R., Rothe, M., and Procyk, E. (2008). Behavioral shifts and action valuation in the anterior cingulate cortex. Neuron *57*, 314–325.
13. Rudebeck, P.H., Behrens, T.E., Kennerley, S.W., Baxter, M.G., Buckley, M.J., Walton, M.E., and Rushworth, M.F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. J. Neurosci. *28*, 13775–13785.
14. Kobayashi, Y., and Amaral, D.G. (2003). Macaque monkey retrosplenial cortex: II. Cortical afferents. J. Comp. Neurol. *466*, 48–79.
15. Vogt, B.A., Finch, D.M., and Olson, C.R. (1992). Functional heterogeneity in cingulate cortex: The anterior executive and posterior evaluative regions. Cereb. Cortex *2*, 435–443.
16. Vogt, B.A., and Gabriel, M. (1993). Neurobiology of Cingulate Cortex and Limbic Thalamus: A Comprehensive Handbook (Boston: Birkhauser).

17. Vogt, B.A., Gabriel, M., Vogt, L.J., Poremba, A., Jensen, E.L., Kubota, Y., and Kang, E. (1991). Muscarinic receptor binding increases in anterior thalamus and cingulate cortex during discriminative avoidance learning. J. Neurosci. *11*, 1508–1514.
18. McCoy, A.N., Crowley, J.C., Haghighian, G., Dean, H.L., and Platt, M.L. (2003). Saccade reward signals in posterior cingulate cortex. Neuron *40*, 1031–1040.
19. McCoy, A.N., and Platt, M.L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. Nat. Neurosci. *8*, 1220–1227.
20. Dean, H.L., and Platt, M.L. (2006). Allocentric spatial referencing of neuronal activity in macaque posterior cingulate cortex. J. Neurosci. *26*, 1117–1127.
21. Dean, H.L., Crowley, J.C., and Platt, M.L. (2004). Visual and saccade-related activity in macaque posterior cingulate cortex. J. Neurophysiol. *92*, 3056–3068.
22. Barraclough, D.J., Conroy, M.L., and Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. Nat. Neurosci. *7*, 404–410.
23. Gabriel, M., Foster, K., and Orona, E. (1980). Interaction of laminae of the cingulate cortex with the anteroventral thalamus during behavioral learning. Science *208*, 1050–1052.
24. Gabriel, M., Sparenborg, S.P., and Stolar, N. (1987). Hippocampal control of cingulate cortical and anterior thalamic information processing during learning in rabbits. Exp. Brain Res. *67*, 131–152.
25. Gabriel, M., Kubota, Y., Sparenborg, S., Straube, K., and Vogt, B.A. (1991). Effects of cingulate cortical lesions on avoidance learning and training-induced unit activity in rabbits. Exp. Brain Res. *86*, 585–600.
26. Brainard, D.H. (1997). The Psychophysics Toolbox. Spat. Vis. *10*, 433–436.
27. Cornelissen, F.W., Peters, E.M., and Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. Behav. Res. Methods Instrum. Comput. *34*, 613–617.

Fictive Reward Signals in the Anterior Cingulate Cortex

Benjamin Y. Hayden, *et al.*

Science **324**, 948 (2009);
DOI: 10.1126/science.1168488

Fictive Reward Signals in the Anterior Cingulate Cortex

Benjamin Y. Hayden,[1]* John M. Pearson,[1] Michael L. Platt[1,2]

The neural mechanisms supporting the ability to recognize and respond to fictive outcomes, outcomes of actions that one has not taken, remain obscure. We hypothesized that neurons in the anterior cingulate cortex (ACC), which monitors the consequences of actions and mediates subsequent changes in behavior, would respond to fictive reward information. We recorded responses of single neurons during performance of a choice task that provided information about the reward values of options that were not chosen. We found that ACC neurons signal fictive reward information and use a coding scheme similar to that used to signal experienced outcomes. Thus, individual ACC neurons process both experienced and fictive rewards.

People routinely recognize and respond to fictive outcomes, which are rewards or punishments that have been observed but not directly experienced (1–3). Fictive thinking affects human economic decisions (4) and is disrupted in disorders such as anxiety and impulsivity (5). Moreover, monkeys respond to information about rewards that they have not directly experienced (6) or were received by other monkeys (7). To understand the neural mechanisms that mediate these processes, we investigated how fictive reward information is encoded in the anterior cingulate cortex (ACC), part of a neural circuit that mediates outcome-contingent changes in behavior (8–10) and processes fictive information in humans (11). The ACC is interconnected with the orbitofrontal cortex, which mediates fictive thinking in humans (5, 12).

[1]Department of Neurobiology, Duke University School of Medicine, Center for Neuroeconomic Studies, Center for Cognitive Neuroscience, Duke University, Durham, NC 27701, USA. [2]Department of Evolutionary Anthropology, Duke University, Durham, NC 27701, USA.

*To whom correspondence should be addressed. E-mail: hayden@neuro.duke.edu

In our task, monkeys chose from among eight white targets arranged in a circle (13). Seven low-value (LV) targets provided small rewards (100 μL), whereas the eighth target [high-value (HV)] provided a variable reward with a larger expected value (EV). Its value on each trial was selected randomly from six possibilities (0, 200, 267, 300, 333, and 367 μL). Once the monkey selected a target, the values associated with all eight of the targets, represented by their colors, were revealed (Fig. 1, A and B). After a half-second delay, the monkey received the reward associated with the chosen target. On the next trial, the position of the HV target either remained in the same position (60% probability) or moved one position clockwise (40% probability).

We analyzed only those trials in which monkeys maintained fixation (90.6% of trials). Because the HV target had a greater EV than the LV targets (243 μL versus 100 μL), we expected that monkeys would prefer the HV target. Indeed, in a control task that explicitly cued HV location, monkeys chose it on 93.4% of trials. In the standard task, monkeys chose the HV target (45.6% of trials) more often than chance ($P < 0.005$, binomial test, Fig. 2A). Monkeys earned 165.0 μL per trial, 88.5% of the amount earned by an omniscient observer with access to information about the value of all targets on all preceding trials (13). Monkeys chose targets adjacent to potential HV targets more often (37.7% of trials) than more distal targets (16.7% of trials, $P < 0.005$, binomial test, Fig. 2A), suggesting that they understood the probabilistic relation between the HV target on the current trial and its likely location on the next.

Large fictive rewards promote gambling in humans (14, 15); thus, we hypothesized that monkeys would likewise preferentially choose HV options after large fictive rewards. We observed this pattern in our experiment (Fig. 2B, black line, correlation coefficient $r = 0.300$, $P < 0.001$). This effect may reflect an increased willingness to switch from to a new target, as the likelihood of switching increased with larger fictive outcomes (Fig. 2C, $r = 0.199$, $P < 0.001$). One alternative explanation for these effects is that HV targets may have positive associations that influence behavior. This explanation is unlikely for several reasons. First, obtained rewards never depended on unselected targets on that trial, so any associations between these fictive stimuli and reward values would be eliminated over the thousands of training trials that preceded recording. Second, immediately after making choices, monkeys were no more likely to make a second saccade (Fig. 2D, $r = -0.02$, $P > 0.2$) nor faster to shift gaze (Fig. 2E, $r = 0.008$, $P > 0.2$) to HV fictive targets than to LV fictive targets, indicating that attention and motivation were roughly similar after all fictive outcomes. Third, we performed a control task in which the HV target remained white and a colored square appeared in the center of the monitor during the delay after the trial. This square's color did not indicate what reward could have been received (and, thus, it provided no fictive information) but had the same associations as the fictive targets.

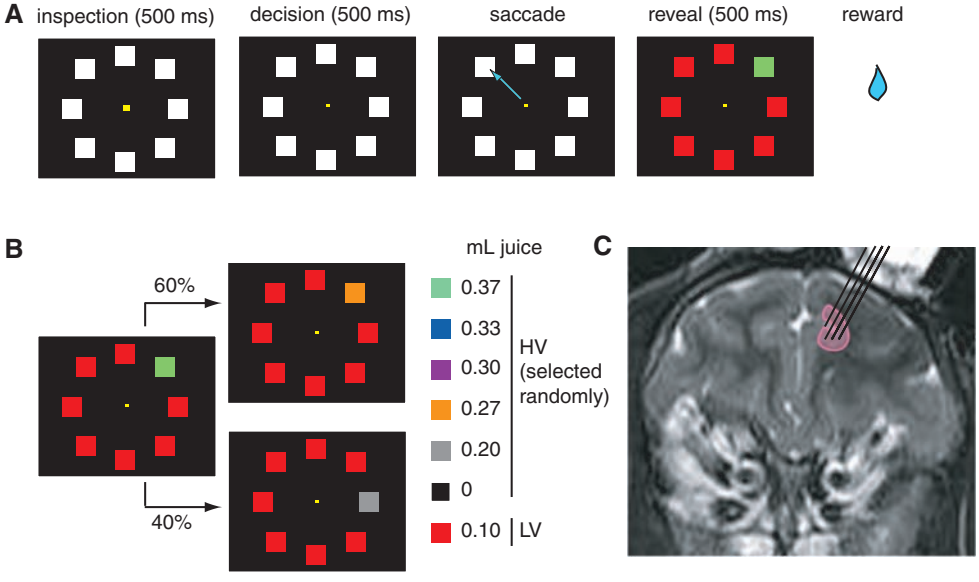

Fig. 1. Task and recording location. (**A**) Schematic of standard task. Fixation point and eight white squares appear; 500 ms after fixation, a monkey chooses one target, and all targets change color, revealing their value. A half-second later, a reward is given. (**B**) Between trials, the HV target either remains at the same position (60% chance) or moves to an adjacent position (40% chance). (**C**) Magnetic resonance image of monkey E. Recordings were made in the ACC sulcus.

Monkeys' choices on subsequent trials did not depend on the color of this stimulus (Fig. 2F, $r = 0.005$, $P > 0.6$).

An example ACC neuron showed clear phasic responses around the time of gaze shifts to targets; the amplitude of these responses was correlated with the size of both the experienced reward (Fig. 3A, $r = 0.056$, $P < 0.001$, the six rewards are grouped into four categories to simplify presentation) and the size of fictive outcomes on trials when the monkey chose the LV target (Fig. 3B, $r = 0.037$, $P < 0.001$). The amplitude of phasic responses of most neurons reflected experienced reward size [$n = 46$ out of 68 (46/68) neurons, 67.7%] and was usually greater for larger rewards ($n = 39/46$, 84.8%). Responses of 50% of neurons reflected fictive reward size ($n = 34/68$); these responses were usually greater for larger fictive rewards ($n = 30/34$, 88.2%, $P < 0.05$). A substantial proportion of neurons (35.5%, $n = 24/68$) showed tuning for both experienced and fictive outcomes; most were tuned in the same direction for experienced and fictive rewards (91.7%, $n = 22/24$). The majority of neurons showed matching tuning for experienced and fictive outcomes (97.0%, $n = 66/68$). For the population, the average response strength

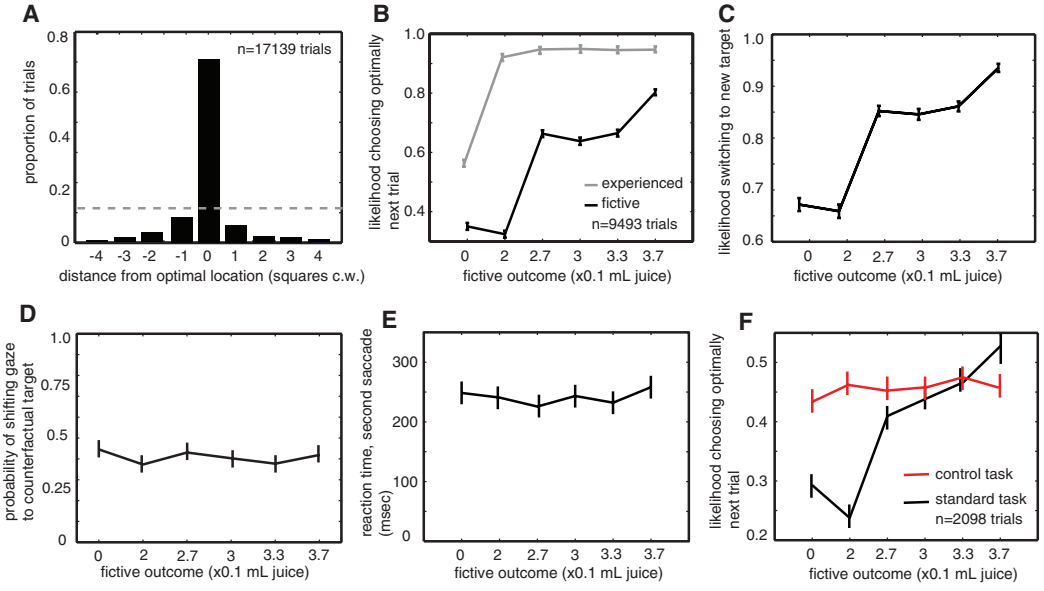

Fig. 2. Fictive outcomes influence behavior. (**A**) Histogram of distance between monkeys' choices and optimal target, measured in squares clockwise (c.w.). The dashed line indicates chance performance. (**B**) Likelihood of choosing optimally increases as a function of both fictive and experienced reward outcome on the previous trial. Black, trials after choice of LV; gray, trials after choice of HV. (**C**) Likelihood of switching to new target increases with size of fictive outcome on previous trial. (**D** and **E**) Likelihood and latency of immediately shifting gaze to HV location are not affected by fictive reward. (**F**) Likelihood of choosing optimally is not influenced by a colored square presented during the delay between trials (red line).

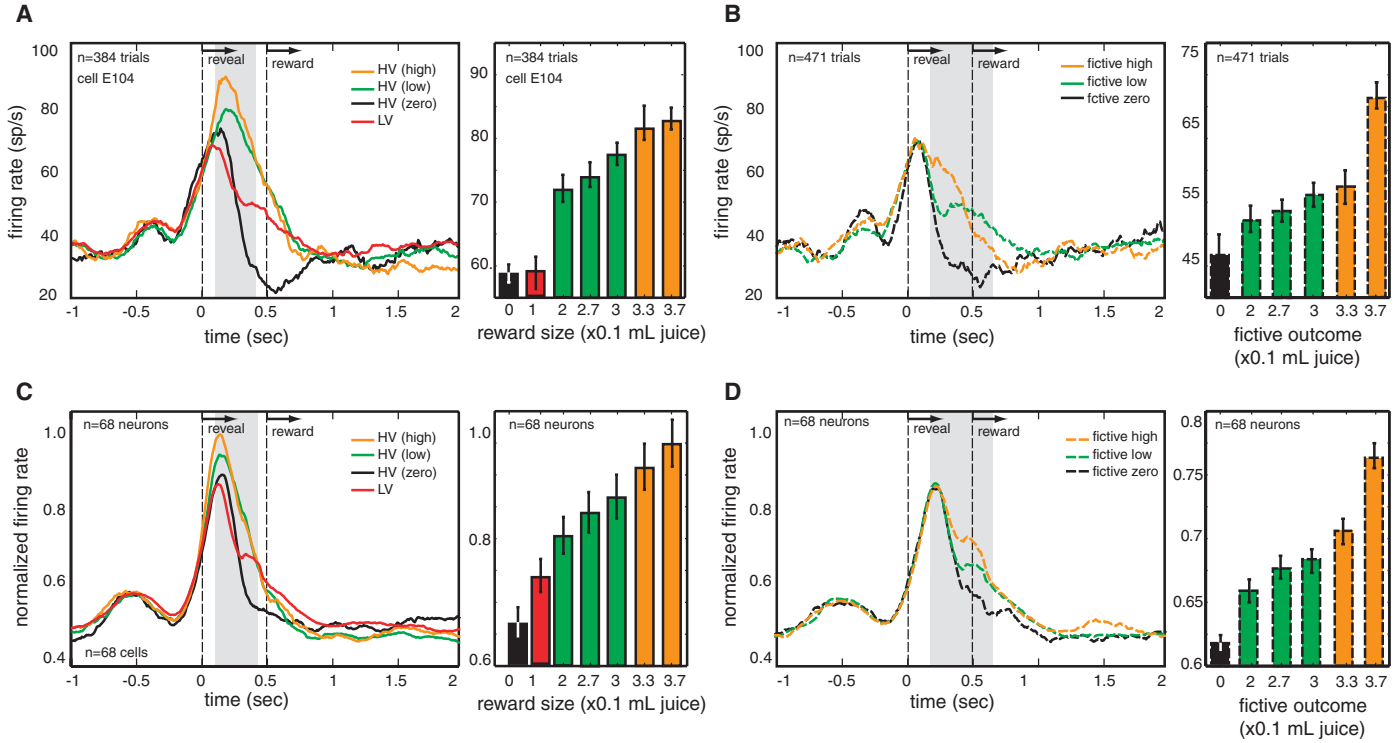

Fig. 3. ACC neurons signal both experienced and fictive rewards. (**A**) (Left) Peristimulus time histogram showing responses of example neuron after choice of HV target. Response grows with reward size. Vertical dashed lines indicate, successively, the time that outcomes are revealed and reward is given. The shaded gray region indicates the epoch used for the bar graph showing average (±1 SE) response of neuron for each experienced reward size. sp/s, spikes per second. (**B**) Responses of the same neuron for fictive rewards. Experienced reward was identical (100 μL) in all cases. (**C**) Population response ($n = 68$ neurons) for experienced rewards, normalized to the maximal firing rate for each neuron. (**D**) Population response for fictive rewards.

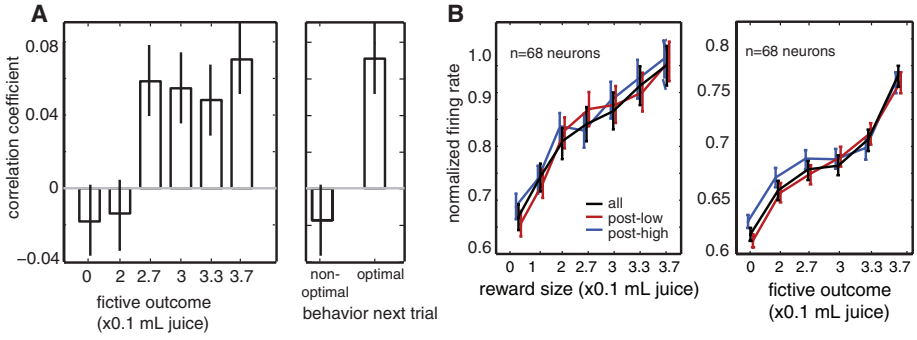

Fig. 4. Neuronal responses signal both fictive rewards and subsequent adjustments in behavior. (**A**) Firing rates after LV trials predict optimal choice on the next trial for four of the six fictive outcomes. (**B**) Neuronal responses to experienced rewards are identical on the trial that follows low (0 μL, red line) and high (≥300 μL, blue line) fictive outcomes and, thus, do not signal reward prediction errors. Error bars indicate 1 SE.

was greater for experienced rewards than for fictive reward outcomes ($P < 0.01$, bootstrap t test). These phasic neural responses are tightly coupled to gaze shifts to visual targets. These responses may thus reflect visual stimulation, reafferent oculomotor signals, or attention to the cue. The amplitude of these phasic responses carries information about the value of fictive outcomes.

To test the hypothesis that responses to fictive rewards may contribute to behavioral adjustment, we calculated the trial-by-trial correlation between firing rate and likelihood of choosing the optimal target after LV trials for all neurons (Fig. 4A). To control for the different neuronal responses to different fictive outcomes, we analyzed data separately for each fictive reward. We found a positive correlation for four of the six fictive outcomes ($P < 0.001$) and no correlation for the remaining two ($P > 0.05$). These results raise the possibility that the firing rate signals subsequent changes in behavior and not fictive outcomes (*16*). However, a second analysis revealed that firing rates were correlated with fictive outcome preceding trials in which monkeys chose optimally ($P < 0.001$). This analysis controls for any adjustment signal, and confirms that ACC neurons do not merely predict behavioral switching. Finally, reaction times did not correlate with likelihood of choosing the optimal target across all recording sessions ($P > 0.5$, correlation test). This analysis controls for the possibility that the correlation between firing rate and adjustment merely reflects uncontrolled variations in arousal.

One alternative explanation for these data is that, by influencing behavior and thus future rewards (Fig. 2B), fictive outcomes serve as the first predictive cue of the reward on the next trial. We find this alternative explanation unlikely for several reasons. First, a choice intervenes between the time of the fictive cue and the reward at the end of the next trial, which is itself probabilistic. The value of the subsequent reward is therefore not strictly predicted by the fictive cue. Second, the reward signal would have to skip the next salient/rewarding event (the reward on the present trial) and signal the subsequent one (the reward on the next trial); such a signal would be highly unusual and has not, to our knowledge, been observed in the ACC or any other brain area. Third, if fictive outcomes are perceived as reward-predicting cues, they should elicit faster reaction times and greater accuracy on the next trial (*17*). We did not observe these effects ($P > 0.5$ for both reaction time and accuracy, Student's t test). Fourth, if the reward on the next trial is larger than the value cued by the fictive outcome on this trial, we should see positive deflections in the neuronal response. Similarly, if the reward on the next trial is smaller than the value cued on this trial, we should see negative deflections in the neuronal response. However, we did not observe any dependence of HV neuronal response on previous fictive value ($P > 0.3$, Fig. 4B). Collectively, these data indicate that the behavioral and physiological correlates of fictive rewards are not an artefactual consequence of simple extended reward associations.

In summary, the most parsimonious explanation for monkeys' behavior in this task is that they recognize and respond to fictive outcomes, and responses of ACC neurons are sufficient to guide such fictive learning. Neural markers of fictive outcomes have so far been limited to non-invasive measures. Hemodynamic activity in the ventral caudate, which is connected with the ACC, reflects fictive learning signals (*15*), and ACC activity tracks the correlation between craving for cigarettes and fictive learning (*11*). The error-related negativity, an event-related potential component with a possible source in the ACC, tracks fictive outcomes (*18*). Here we show that the same neural circuit carries information about fictive outcomes in monkeys. Moreover, information about both experienced and fictive outcomes is encoded by the same neurons and is represented with the use of a similar coding scheme. The correlation between firing rate and behavior suggests that these neurons do not simply tag the incentive salience of a stimulus (*19*, *20*), but also reflect neuronal processes that translate outcomes into behavior. Thus, the ACC may integrate information about obtained rewards [probably signaled by the dopamine system (*21*, *22*)] with information about observed rewards [presumably computed in the prefrontal cortex (*23*)] to derive a model of the local reward environment in the near future. These findings are consistent with the idea that the ACC represents both real and fictive reward outcomes to dynamically guide changes in behavior (*9*, *24*–*27*). Such a mechanism may be crucial in complex social environments, where the behavior of others provides a rich supply of fictive information (*15*, *28*).

References and Notes

1. R. M. Byrne, *Trends Cogn. Sci.* **6**, 426 (2002).
2. K. Epstude, N. J. Roese, *Pers. Soc. Psychol. Rev.* **12**, 168 (2008).
3. N. J. Roese, *Psychol. Bull.* **121**, 133 (1997).
4. G. Loomes, R. Sugden, *Econ. J.* **92**, 805 (1982).
5. S. Ursu, C. S. Carter, *Cogn. Brain Res.* **23**, 51 (2005).
6. D. Lee, B. P. McGreevy, D. J. Barraclough, *Cogn. Brain Res.* **25**, 416 (2005).
7. F. Subiaul, J. F. Cantlon, R. L. Holloway, H. S. Terrace, *Science* **305**, 407 (2004).
8. S. Ito, V. Stuphorn, J. W. Brown, J. D. Schall, *Science* **302**, 120 (2003).
9. S. W. Kennerley, M. E. Walton, T. E. Behrens, M. J. Buckley, M. F. Rushworth, *Nat. Neurosci.* **9**, 940 (2006).
10. J. G. Kerns *et al.*, *Science* **303**, 1023 (2004).
11. P. H. Chiu, T. M. Lohrenz, P. R. Montague, *Nat. Neurosci.* **11**, 514 (2008).
12. N. Camille *et al.*, *Science* **304**, 1167 (2004).
13. Materials and methods are available as supporting material on *Science* Online.
14. R. L. Reid, *J. Gambling Behav.* **2**, 32 (1986).
15. T. Lohrenz, K. McCabe, C. F. Camerer, P. R. Montague, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9493 (2007).
16. K. Shima, J. Tanji, *Science* **282**, 1335 (1998).
17. M. R. Roesch, C. R. Olson, *Science* **304**, 307 (2004).
18. J. P. Goyer, M. G. Woldorff, S. A. Huettel, *J. Cogn. Neurosci.* **20**, 2058 (2008).
19. S. M. McClure, N. D. Daw, P. R. Montague, *Trends Neurosci.* **26**, 423 (2003).
20. K. C. Berridge, T. E. Robinson, *Brain Res. Rev.* **28**, 309 (1998).
21. W. Schultz, *Annu. Rev. Psychol.* **57**, 87 (2006).
22. P. R. Montague, G. S. Berns, *Neuron* **36**, 265 (2002).
23. M. F. Rushworth, T. E. Behrens, *Nat. Neurosci.* **11**, 389 (2008).
24. C. Amiez, J. P. Joseph, E. Procyk, *Cereb. Cortex* **16**, 1040 (2005).
25. S. W. Kennerley, A. F. Dahmubed, A. H. Lara, J. D. Wallis, *J. Cogn. Neurosci.* **21**, 1162 (2009).
26. R. Quilodran, M. Rothe, E. Procyk, *Neuron* **57**, 314 (2008).
27. M. Matsumoto, K. Matsumoto, H. Abe, K. Tanaka, *Nat. Neurosci.* **10**, 647 (2007).
28. P. H. Rudebeck, M. J. Buckley, M. E. Walton, M. F. S. Rushworth, *Science* **313**, 1310 (2006).

29. This work was supported by a postdoctoral fellowship to B.Y.H. (National Institute on Drug Abuse 023338), an R01 to M.L.P. (National Eye Institute 013496), and the Duke Institute for Brain Studies). We thank K. Watson for help in training the animals and S. Heilbronner for useful discussions on the tasks. The authors declare no conflicts of interest.

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5929/948/DC1
Materials and Methods
SOM Text
Figs. S1 to S9
References

13 November 2008; accepted 12 March 2009
10.1126/science.1168488