Published in final edited form as:
Neuron. 2008 May 8; 58(3): 451–463.

Value representations in the primate striatum during matching behavior

Brian Lau* and **Paul W. Glimcher**
Center for Neural Science, New York University

Summary

Choosing the most valuable course of action requires knowing the outcomes associated with the available alternatives. The striatum may be important for representing the values of actions. We examined this in monkeys performing an oculomotor choice task. The activity of phasically active neurons (PANs) in the striatum covaried with two classes of information: action-values and chosen-values. Action-value PANs were correlated with value estimates for one of the available actions, and these signals were frequently observed before movement execution. Chosen-value PANs were correlated with the value of the action that had been chosen, and these signals were primarily observed later in the task, immediately before or persistently after movement execution. These populations may serve distinct functions mediated by the striatum: some PANs may participate in choice by encoding the values of the available actions, while other PANs may participate in evaluative updating by encoding the reward value of chosen actions.

Keywords

basal ganglia; value; reward; reinforcement learning; expectation

Introduction

Neural activity in a number of brain areas is related to the values of rewards humans or animals gain, as well as the choices they make using their estimates of these values (Schultz, 2000; Sugrue et al., 2005; Daw and Doya, 2006). A growing body of evidence suggests that the basal ganglia is important for maintaining value representations to guide actions (Hikosaka et al., 2006). Phasically active neurons in the dorsal striatum can be modulated by reward properties (Cromwell and Schultz, 2003; Hassani et al., 2001), by changes in reward contingencies (Kawagoe et al., 1998; Lauwereyns et al., 2002a) as well as association learning (Pasupathy and Miller, 2005; Tremblay et al., 1998; Williams and Eskandar, 2006), and respond in a manner consistent with a role in biasing actions (Lauwereyns et al., 2002b; Watanabe et al., 2003; Samejima et al., 2005). These data suggest that striatal PANs promote the selection of valuable actions by modulating activity in the thalamus and midbrain.

However, in the oculomotor caudate, a nucleus of the striatum, it is not known how the activity of PANs reflects the values of actions during choice behavior. Simultaneous measurements of

Address correspondence to: Brian Lau, Center for Neural Science, New York University, 4 Washington Place, Room 809, New York, NY 10003, Voice: +1.212.998.3904, Fax: +1.212.995.4011, Email: blau@cns.nyu.edu.
*Current address: Department of Neuroscience, Columbia University

striatal activity and estimates of subjective values would thus be useful for testing whether PANs encode a subject's estimates of the values of actions. In addition, such measurements would allow us to more precisely define the types of value-related information these neurons encode. PANs could, for example, encode action-values, the values associated with potential actions. Basal ganglia models often posit that striatal PANs encode a quantity like action-value that biases the selection of actions associated with more valuable outcomes (Doya, 2000). Alternatively, striatal PANs could encode chosen-values, the value of the option the decision-maker selects (Morris et al., 2006; Padoa-Schioppa and Assad, 2006). A neuron encoding chosen-value cannot support action selection because its activity is contingent on the action ultimately executed. Chosen-value representations may be useful for evaluating the outcomes of actions to promote learning (Morris et al., 2006; Niv et al., 2006; Sutton and Barto, 1998) or modifying movements to reflect the value of the action (e.g., reaction time). Do PANs in the oculomotor caudate reflect one or both of these value representations when animals choose amongst actions associated with changing reward values?

To answer this question, we recorded from PANs in the oculomotor caudate while monkeys performed a choice task that elicited matching behavior (Herrnstein, 1961). We estimated the action-values and chosen-values associated with each action both at the level of the sequential blocks of trials presented to the monkeys as well as at the level of individual choices, and used these estimates to determine whether PANs tracked action-values or chosen-values. We found that the activity of a significant number of PANs was correlated with action-values and that the activity of a second group of PANs was correlated with chosen-values. Action-value related activity was more prominent prior to movement execution while chosen-value related activity was more prominent following movement execution. These results support the idea that some striatal neurons bias action selection, and provide evidence that a second novel group of striatal neurons may have an evaluative role, reporting the reward values associated with chosen actions.

Results

We recorded from PANs in the caudate nucleus of monkeys performing an oculomotor choice task where the values of the two available alternatives varied (Figure 1A). The choice task was based on the concurrent variable-interval schedules used to study Herrnstein's matching law (Herrnstein, 1961), which describes how many animals, including humans, choose amongst alternatives that differ in value (Davison and McCarthy, 1988; Williams, 1988). Monkeys allocate their choices in proportion to the relative probability or magnitude of rewards in this type of task (Corrado et al., 2005; Lau and Glimcher, 2005). Here, we varied the relative magnitude of rewards in blocks of roughly 130 trials while keeping the average probabilities of reward for each alternative equal. Once a reward was arranged for an alternative, it remained available until it was next chosen, similar to the reinforcement schedules used to elicit matching behavior in free operant experiments (Nevin, 1969). We found that monkeys matched their choices to the relative magnitude of rewards obtained from each alternative (Figure 1B). Moreover, their choice behavior following transitions to different relative magnitudes of reward quickly stabilized (Figure 1C). These results contrast with the behavior animals exhibit under variable-ratio schedules, where rewards are not held between choices; under those contingencies animals often learn to exclusively choose the better alternative, with little or no variation in relative choice as a function of the relative value of that alternative (Herrnstein and Vaughan, 1980; Samejima et al., 2005). That choice was lawfully related to relative reward magnitude during matching indicates that the monkeys acquired and maintained information about the consequences of their actions. This is consistent with the idea that their choices were based on the relative values they placed on the two actions.

We hypothesized that caudate PANs encode the values associated with specific actions. However, a correlation between neuronal activity and value does not necessarily mean that a neuron participates in the action selection. Identifying neurons that could be related to action selection requires distinguishing between what we term action-value and chosen-value. The difference between these two value representations is illustrated in Figure 2A. Action-values represent the potential outcomes available to the decision-maker, and can be used to select actions associated with these outcomes. Chosen-values cannot support action selection since they do not unambiguously reflect the value of one of the available actions. However, neurons encoding chosen-values may be useful for both evaluating and executing chosen actions. Hypothetical responses action-value or chosen-value neurons are illustrated in Figure 2B, where the responses are separated by the action chosen and whether the contralateral action-value was greater than the ipsilateral action-value. Action-value neurons reflect the value associated with a particular action irrespective of which action is selected. Chosen-value neurons, on the other hand, reflect the reward value associated with the selected action. Padoa-Schioppa and Assad (2006) first noted chosen-value activity in the orbitofrontal cortex (OFC), where they found that OFC chosen-value neurons encoded the chosen-value of whichever option was selected. It is important to note that the fundamental property of chosen-value activity is value sensitivity that depends on choice. Thus, the hypothetical chosen-value neuron illustrated in Figure 2B is only one of the possible types of chosen-value neuron. For example, another chosen-value neuron might respond most for contralateral choices only when that action is most valuable (high response in upper left quadrant) but be insensitive to the value of the ipsilateral target when it is chosen (low response in all other quadrants). Both because movement selectivity is common in the caudate nucleus and because we had no a priori reason to exclude the possibility that different types of chosen-value neuron might occur, we tested the possibility that different types of chosen-value neurons might be found in the caudate.

Relating caudate activity to blockwise estimates of value

We used the programmed reward magnitudes set by the reinforcement schedule to generate estimates of the chosen-values and action-values for each session (Figure 2A), which we refer to as *blockwise value estimates*. Although these value estimates do not vary from trial to trial as the animals' internal estimates probably do, they have the advantage of being essentially model free.

Individual PANs were active at idiosyncratic and highly repeatable times during each trial (Hikosaka et al., 1989a; Lau and Glimcher, 2007). To analyze this activity, we identified the time of peak activity for each neuron, and used the first times to half-maximal response preceding and following this peak to define an analysis window. We used multiple linear regression including blockwise value estimates as covariates to categorize individual neurons into three exclusive populations: 1) non-value, 2) action-value or 3) chosen-value. We also refer to non-value neurons that responded differentially according to movement direction as choice-only (Figure 2B).

Neurons categorized as action-value and chosen-value are plotted in Figure 3. For each category, the firing rates for two example PANs are plotted in each row, sorted by chosen action and whether the contralateral action-value was greater than the ipsilateral action-value. Both example action-value neurons are more active when the reward associated with the contralateral target is larger than that associated with the ipsilateral target; however, the neuron in Figure 3A exhibits this difference before the onset of the choice cues, whereas the neuron in Figure 3B exhibits this difference after the onset of the choice cues. The two example chosen-value neurons (Figure 3C–D) are more active when the ipsilateral target is chosen and that target is associated with the larger reward. In contrast to action-value neurons, the neuron in Figure 3C is also more active when the contralateral target is chosen and that target is associated

with the larger reward; this neuron reflects the value of whichever action was selected. The neuron in Figure 3D also differs from action-value neurons; its activity reflects the value of the ipsilateral action when it is chosen ($p<0.05$, *t*-test) but does not reflect value when the contralateral action is chosen ($p>0.10$, *t*-test). Figure 3E and 3F display neurons that were direction selective ($p<0.05$, *F*-test) but not value sensitive ($p>0.05$, *t*-test). The examples in Figure 3A–D show that different neurons can encode value at different points in a trial, from before cue presentation to after saccade execution. They are also representative in illustrating that neurons sensitive to value can exhibit additive changes in firing rate due to choice (Figure 3B).

We included a number of additional covariates in our regression analysis to protect against potential confounds: the direction, latency and speed of movements (Itoh et al., 2003; Watanabe et al., 2003) as well as reward outcome (Apicella et al., 1991). The parameters for these additional covariates were estimated alongside the value covariates, and summarized in Table 1. A significant number of PANs encoded information about choice (i.e. movement direction) and obtained reward in addition to value (Table 1).

We found that 62% of task-related PANs covaried significantly with action-value or chosen-value ($p<0.05$, *F*-test). To summarize the regression analysis and examine how substantially value influenced neuronal activity, we quantified the magnitude of firing rate changes due to changes in value. For each neuron, we subtracted from the raw firing rate on each trial the predicted effect due to all variables except value. This partial residual isolates the effect of value on firing rate by holding constant the effects due to all other covariates. We then averaged the partial residuals across trials and neurons for each category. Since neurons could increase or decrease firing rates in response to value, we altered the sign of the effect for each neuron so that the mean effect for all neurons increased for increasing contralateral value. To facilitate comparisons with hypothetical responses (Figure 2B), we computed separate averages according to the chosen action as well as to whether the contralateral action-value was greater than the ipsilateral action-value (Figure 4A). Figure 4A also shows the mean effects for non-value neurons, further separated into choice-only neurons (partial residuals computed for choice rather than value), and neurons that did not respond differentially to choice (partial residuals computed for value). The choice-only neurons are a useful reference since some PANs respond selectively during saccade execution; the quadrant plot in Figure 4A shows a contralateral bias in these neurons, which has been observed in the caudate (Hikosaka et al., 1989a), and provides a scale against which the effects of value can be compared. At the population level, blockwise variations in reward value change firing rates by approximately the same amount as changes in firing rate due to different saccade directions. These results indicate that a significant number of PANs were correlated with the reward value associated with particular actions.

To further explore the differences between action-value and chosen-value neurons we examined the partial residuals for individual neurons in each category. We computed an index based on the mean effects used to generate the average quadrant plots in Figure 4A by summing the absolute values of the row-wise differences of the quadrant plot for each neuron. This is a simple summary of the column asymmetry of the quadrant plot for each neuron, which reflects the degree to which a neuron with value sensitivity exhibits a dependence on the chosen action. Action-value neurons should produce column asymmetry indices close to zero whereas chosen-value neurons should produce column asymmetry indices greater than zero (Figure 2B). Note, however, that it is possible for chosen-value neurons to exhibit smaller asymmetry indices for weaker correlations with chosen-value. Figure 4B displays the asymmetry index for each neuron separated by value representation. The distributions for action-value neurons and chosen-value neurons show relatively little overlap, supporting the distinction between these value representations. Also illustrated in Figure 4B are those chosen-value neurons that had

significant coefficients for both chosen-values. We found that, unlike neurons in the OFC (Padoa-Schioppa and Assad, 2006), 58% (15/26) of chosen-value PANs in the caudate were significantly correlated with the chosen-value of only one action, while the remaining 42% (11/26) were significantly correlated with the chosen-values of both actions ($p<0.05$, *t*-test for both chosen-values). Thus, chosen-value PANs in the caudate are not homogenous, although both of these types of chosen-value neuron are distinct from action-value neurons.

Relating caudate activity to trial-by-trial estimates of value

In the analysis above we used the blockwise reward magnitudes to approximate the monkeys' internal value estimates. While simple, this approach has two limitations: 1) it ignores the behavioral dynamics that occur within a block and 2) while it allows us to identify action-value neurons, we cannot distinguish positive covariation with the contralateral action-value from negative covariation with the ipsilateral action-value. This ambiguity arises because the two reward magnitudes available to the monkeys in each block summed to a constant value by design in our experiment (see Experimental Procedures). We addressed both of these limitations by developing trial-by-trial estimates of the subjects' internal valuations for each action from the monkeys' choice behavior. To accomplish this we fit a reinforcement learning model to choice behavior to generate dynamic value estimates for each action (cf. Barraclough et al., 2004; Dorris and Glimcher, 2004; Sugrue et al., 2004; Samejima et al., 2005). We used these estimates to further explore the categorization we arrived at using the blockwise value estimates described above, and to extend our observations by examining whether action-value and chosen-value neurons covary with value estimates on a trial-by-trial basis.

Our behavioral model had three terms (Lau and Glimcher, 2005): 1) a linearly weighted sum of past rewards 2) the magnitude of each of the currently available rewards and 3) a linearly weighted sum of past choices. We have shown that this model accurately captures fluctuations in behavior driven by stochastic reward delivery in our matching task (Lau and Glimcher, 2005). The function of the first term is to identify the weight a subject places on each previous reward as a function of how long ago that reward was received. In practice, we have found that this linear weighting function on rewards takes an exponentially decaying form with recent rewards most strongly influencing current value estimates. This is what would be expected if the monkeys used a simple prediction-error learning rule to estimate the value of each alternative (e.g., Bayer and Glimcher, 2005). The second term simply encodes the magnitude of the rewards available in each block of trials, and allows the model to predict a simple bias (across the block) for the action associated with a larger reward. The third term captures the influence of previous choices on a current choice. In practice, this linear weighting on past choices captures features like the strong tendency of monkeys to alternate actions independent of rewards. Importantly, incorporating past choices into our behavioral model allowed us to accurately estimate the behavioral influence of past rewards and currently available rewards. Furthermore, we are able to separate the behavioral effects of reward value (first two model terms) from those of past choices (third term), which allowed us to test the hypothesis that PANs encode the dynamic action-values and chosen-values associated with the available choice alternatives.

We used the coefficients from this behavioral model related to reward value (the first two terms described in the preceding paragraph)—fit separately to the choice data pooled across sessions for each monkey—to generate dynamic reward-value estimates associated with each alternative (Figure 5A). These estimates represent the fluctuating subjective preferences of the monkeys due to stochastic rewards and are directly related to the probability that the subject will make a particular choice on each trial. In a manner similar to the blockwise value estimate analysis described above, we constructed dynamic action-values and dynamic chosen-values for each alternative. We then performed a second regression for each neuron of the type

described in the previous section, where we used the trial-by-trial estimates of action-value and chosen-value instead of the blockwise estimates (see Experimental Procedures). We used the neuronal categorizations from the blockwise analysis above to determine whether to correlate the activity of a particular neuron with dynamic estimates of action-value or chosen-value. Because we did not intentionally decouple blockwise values from dynamic values, these value estimates were strongly correlated, which means we are unable to determine whether dynamic values can produce fundamentally different categorizations of striatal neurons. Therefore, the following analysis is conditional on our categorization using blockwise values, which do not depend on a behavioral model.

Two example neurons, one action-value and one chosen-value, are shown in Figure 5B–C. For each neuron the partial residuals for value are plotted against the dynamic value estimates for each action, separated by chosen action. The distinction between action-value and chosen-value made previously applies here; action-value neurons reflect the value of a particular action irrespective of the chosen action, whereas chosen-value neurons exhibit value sensitivity that depends on the action selected. The action-value neuron in Figure 5B is significantly correlated with the contralateral action-value (left panel), but not the ipsilateral action-value (right panel). Importantly, a correlation with contralateral action-value exists for both ipsilateral and contralateral choices. The activity of the chosen-value neuron in Figure 5C is plotted as a function of action-values rather than chosen-values to better illustrate the difference between these value representations. Since chosen-values are equal to action-values when a particular action is chosen (and zero otherwise), chosen-value neurons will only exhibit sensitivity for the action-value associated with the action chosen. Thus, in Figure 5C, activity positively covaries with contralateral value when it is chosen (red points in the left panel), but is not sensitive to contralateral action-value when the ipsilateral target is chosen (green Xs in the left panel). Instead, activity when the ipsilateral target is chosen positively covaries with ipsilateral value (green Xs in the right panel), and not with action-value when the contralateral target is chosen (red points in the right panel).

Across PANs, the results using dynamic value estimates are consistent with the categorization obtained using blockwise value estimates. The great majority of action-value neurons identified by our blockwise analysis (86%, 31/36) significantly covaried with dynamic action-values ($p<0.05$, *t*-test). Importantly, since the dynamic action-values do not sum to a constant due to stochastic variations in reward delivery (Figure 5A), we were able to identify whether individual neurons responded to the value associated with a particular action. We found that 81% (25/31) of these action-value neurons selectively represented the value associated with only one action ($p<0.05$ for only one action-value, *t*-test), similar to the example in Figure 5B. For these neurons we defined the preferred action-value as whichever was significant. The remaining six action-value neurons had significant coefficients for both action-values (for four neurons one coefficient was roughly twice as large as the other but of the same sign, two neurons had coefficients of opposing sign); for these neurons we defined the preferred action-value as whichever had the larger absolute coefficient. We found that 61% (19/31) of action-value neurons preferred contralateral action-values while the remainder preferred ipsilateral action-values (not different from 50%, $p>0.10$, binomial test). Individual neurons either increased or decreased their responses according to action-value, and we observed that 65% of action-value neurons increased firing rate for the preferred action-value, while the remainder decreased firing rate (not different from 50%, $p>0.10$, binomial test).

We found that 85% (22/26) of chosen-value neurons significantly covaried with dynamic chosen-values ($p<0.05$, *t*-test). More of these neurons (77%, 17/22) significantly covaried with the value associated with only one particular action ($p<0.05$ for either chosen-value, *t*-test), for example, the chosen-value neurons in Figure 3. The remaining five neurons significantly covaried with both chosen-values ($p<0.05$ for both chosen-values, *t*-test), for example, the

chosen-value neuron in Figure 5C. The majority (77%, 17/22) of chosen-value neurons preferred contralateral chosen-values while the remainder preferred ipsilateral chosen-values (significantly different from 50%, $p<0.05$, binomial test). We observed that 55% of chosen-value neurons increased firing rate for the preferred chosen-value, while the remainder decreased firing rate (not different from 50%, $p>0.10$, binomial test).

A population summary is plotted in Figure 6, where each row represents the mean effect of dynamic value on firing rate, averaged over neurons within a category. Before averaging, the effect for those neurons with negative value coefficients was sign-reversed so that all the data are presented as positive increases in firing rate with increasing preferred value. Figure 6 shows that there is a robust encoding of dynamic value across our population of PANs for firing rate changes associated with the preferred dynamic value for both action-value and chosen-value neurons, and the data agree with the slope predicted by the regression coefficients for the preferred value (black line = median coefficient from the dynamic value regression). The relationship appears less clear for chosen-value neurons as a function of the non-preferred value (right panel, middle row), a feature that arises in the plot from the fact that there is more than one type of chosen-value neuron. The left panel of Figure 6B displays the mean effect as a function of the preferred action-value. Since most of the chosen-value neurons reflect the dynamic value for only one choice, and we aligned the data to the preferred action-value in the left panel, this is reflected in a clearer mean effect across the population as a function of the preferred action-value. The right panel of Figure 6B displays the mean effect as a function of the non-preferred action-value. Again, since the bulk of caudate chosen-value neurons reflect the value for only one choice, this is reflected in a smaller mean effect across the population in the right panel.

The results presented so far are consistent with the idea that PAN activity tracks value fluctuations due to stochastic reward delivery. However, since the dynamic value estimates are linearly weighted versions of past and currently available rewards, they are inherently correlated with the blockwise value estimates. Thus, demonstrating trial-by-trial covariation requires further showing that PAN responses are not fully explained by the blockwise values. We tested this by asking whether the covariation illustrated in Figure 6 was explained entirely by blockwise value. For individual neurons, we fit the partial residuals for the value covariates with a linear model including blockwise values and dynamic values as covariates. If the blockwise values fully accounted for the covariation between PAN responses and value, then only the coefficient for blockwise values would be significant, there would be no additional explanatory power offered by the dynamic values. On the other hand, a significant coefficient for dynamic value would indicate that PAN responses covaried on a trial-by-trial basis with the dynamic value estimates. For the action-value and chosen-value neurons that were significantly correlated with dynamic action-values and chosen-values, respectively, we found that 58% (18/31) action-value neurons and 36% (8/22) chosen-value neurons were significantly correlated with dynamic value ($p<0.05$, F-test). Thus, some value-sensitive PANs were correlated with the trial-by-trial value estimates generated by our behavioral model.

Taken together, these results provide further evidence that a significant number of PANs encode the reward values of actions. By using a behavioral model to generate dynamic value estimates, we also found correlations that suggest that, at the population level, striatal PANs respond in a monotonic fashion to trial-by-trial variations in the reward value associated with specific actions.

Temporal evolution of value representations

We also examined the response profiles of action-value and chosen-value neurons to determine whether response time further differentiated action-value and chosen-value activity in the caudate nucleus. The average population responses are plotted in Figure 7A. Action-value

neurons were more active than chosen-value neurons prior to target acquisition, and this relationship reverses during reward delivery. At the individual neuron level (Figure 7B), a difference between action-value and chosen-value neurons is supported by the fact that the median peak response time was significantly earlier for action-value neurons (−58 ms vs. +393 ms relative to target acquisition; $p<0.05$, Mann-Whitney U-test).

We further examined neural responses across the population using smaller temporal windows. For each neuron, we fit the firing rate in 250 ms non-overlapping windows with the same model used above to categorize the neurons. Activity was categorized as action-value, chosen-value or non-value, and the results were tallied across the population (Figure 8). Action-values were represented primarily before a choice was made, peaking before saccade execution. Chosen-value activity peaks following saccade execution, with significantly more chosen-value activity than action-value activity late in the trial ($p<0.05$, z-test for differences in proportions from the same sample; Wild and Seber, 2000).

These data indicate that action-value and chosen-value representations in the caudate nucleus have different temporal profiles. The predominantly presaccadic representation of action-values is consistent with this activity biasing action selection, whereas the predominantly postsaccadic representation of chosen-values is consistent with this activity being related to evaluating the outcomes of particular actions.

Discussion

We found that roughly 60% of striatal PANs were modulated by the reward values associated with two different actions in a choice task based on Herrnstein's matching law (Herrnstein, 1961). The responses of individual PANs covaried with two distinct types of value: action-value (Samejima et al., 2005) and chosen-value (Padoa-Schioppa and Assad, 2006). Action-values represent the desirabilities of actions, and can be used to make choices (Luce, 1959; Sutton and Barto, 1998). Chosen-values depend on the action selected, and may be useful for both for executing actions and evaluating the consequences of those actions. PAN activity correlated with each of these value types emerged at different times within a trial. Action-values were more frequently correlated with PAN activity early in trials, before our subjects revealed their choices. In contrast, correlations with chosen-values tended to occur following saccade execution. These results suggest that the striatum participates in two different aspects of reinforcement learning; in promoting the selection of particular actions as well as in evaluating the outcomes associated with the chosen action.

Action-values

Our results compliment and extend existing studies of the caudate that have used forced-choice tasks. Hikosaka and colleagues manipulated which instructed saccade would be rewarded and found that some caudate PANs signal whether or not a reward can be expected for executing particular eye movements (Kawagoe et al., 1998). Further, they found that a subset of these PANs respond more when a reward is predicted for a particular saccade regardless of which movement is instructed (Lauwereyns et al., 2002a,b). Hikosaka and colleagues propose that these PANs signal the motivational context of the instructed movement, and that these neurons could bias the speed and latency of eye movements by disinhibiting the superior colliculus via the substantia nigra pars reticulata (Hikosaka et al., 2006). We found that action-value neurons encode the value of available movements in a choice context, suggesting that these neurons parametrically represent the value of all potential saccades. PANs may play a role in action selection as well, with projections to the thalamus and midbrain biasing choices in addition to modulating movement metrics.

Samejima et al. (2005) used a hand-movement choice task to show that PANs in the putamen, as well as in a portion of the caudate nucleus, are correlated with action-values. Our results suggest a similar representation: 1) the majority of oculomotor PANs that were significantly modulated by action-value were correlated with the action-value associated with only one of the two available movements, 2) contralateral action-value PANs as well as ipsilateral action-value PANs were found in the same hemisphere and 3) few PANs were correlated with the difference between the action-values associated with the two movements. One difference between these two sets of findings is that we found that the majority of action-value neurons (86%) also exhibited some degree of direction selectivity; neurons frequently combined action-value information with movement selectivity. In contrast, Samejima et al. (2005) found that only 22% of action-value neurons exhibit movement selectivity.

There are a number of possible explanations for this difference. First, Samejima et al. recorded in both the putamen and caudate nucleus while monkeys indicated their choices with hand movements. The caudate nucleus contains oculomotor neurons (Hikosaka et al., 1989a) but few neurons associated with skeletomotor movements. Perhaps they did not observe movement selectivity in the caudate because their monkeys did not indicate choice using eye movements. Second, we searched for neurons using an instructed saccade task. It is possible that we recorded from fewer action-value neurons without movement selectivity because those neurons may have been silent during the instructed saccade task. Despite this difference, both sets of results suggest that some striatal PANs encode action-values.

In many choice situations outcomes are linked to specific stimuli rather than actions. The reward values associated with specific stimuli have been referred to as *offer-values* (Padoa-Schioppa and Assad, 2006). Offer-values differ from action-values in that the former reflect the reward value of an alternative independent of action. Neurons encoding offer-value have been identified in the OFC (Padoa-Schioppa and Assad, 2006), and our data do not exclude the possibility that striatal PANs may encode offer-values when stimulus value and action are dissociated. Indeed, evidence suggests that some PANs are modulated by stimulus color when color rather than movement direction predicts reward (Lauwereyns et al., 2002a).

Chosen-values

We also found that roughly 25% of PANs convey information about chosen-values, a type of encoding not previously identified in the basal ganglia. This class of signal may be important for learning from the consequences of actions; Morris et al. (2006) showed that dopamine neurons in the primate midbrain encode a specific prediction error, the difference between obtained reward and chosen-value. Their data support reinforcement learning models that use the difference between obtained reward and chosen-value as a teaching signal (Niv et al., 2006; see also Roesch et al., 2007). Our observation that some PANs encode chosen-value suggests that these neurons may convey chosen-value information to dopamine neurons via projections to the midbrain. The striatum is chemically divided into regions known as striosomes surrounded by a more diffuse matrix (Graybiel and Penney, 1999); both contain PANs, but striosomal PANs project to the substantia nigra pars compacta (Gerfen et al., 1987; Joel and Weiner, 2000). This pathway may be specialized for computing prediction errors that promote learning about rewarded actions (Doya, 2000; Houk et al., 1995). We do not know whether our chosen-value PANs reside in the striosomes, but these signals are necessary for computing the prediction error that has been observed in primate dopamine neurons (Morris et al., 2006).

The chosen-value activity we observed might be inherited from the frontal cortex, although there are some differences between these two areas. Neurons in the OFC, which projects to the striatum (Selemon and Goldman-Rakic, 1985), encode chosen-value signals when monkeys are choosing between juice rewards (Padoa-Schioppa and Assad, 2006). The chosen-value

PANs we observed may reflect information passing from the OFC through one of the parallel basal ganglia pathways (Alexander et al., 1986). However, the OFC neurons represent chosen-value irrespective of the saccade direction used to indicate choice. We found that some caudate PANs reflected the chosen-value for only one saccade direction. This difference may reflect our task; the alternatives were associated with particular saccades. However, it may also reflect a feature of the chosen-value representation in the basal ganglia. It is possible that the basal ganglia is primarily involved in decisions between actions rather than between more abstract options, and future work dissociating action and outcomes will help to clarify this issue.

Finally, our findings of chosen-value signals in the caudate may also extend our understanding of activity observed in forced-choice tasks. Ding and Hikosaka (2006; see also Kobayashi et al., 2007) showed that some PANs exhibit response-dependent reward activity in an instructed saccade task. They found that some PANs respond for any movement that yielded a reward, while other PANs respond for movements in only one of the directions that yielded reward. It is possible that our chosen-value PANs are members of the same population, and that some of the neurons Hikosaka and colleagues recorded from may encode chosen-value in a continuous fashion.

Reinforcement learning

When rewards associated with stimuli or actions change, value estimates can be updated through experience. A number of theories describe how the values of actions may be learned, and decision-making models incorporating these algorithms are efficient in a variety of contexts (e.g., Sugrue et al., 2004). Circuits within the basal ganglia may instantiate components of these learning models (Daw and Doya, 2006; Houk et al, 1995). Midbrain dopamine neurons, for example, appear to encode a reward prediction error (RPE), a critical component of reinforcement learning models (Schultz et al., 1997). Functional magnetic resonance imaging (fMRI) studies show that blood oxygenation level dependent (BOLD) changes in the striatum are correlated with RPEs predicted by reinforcement learning models (McClure et al., 2003; O'Doherty et al., 2003). Interestingly, the site of BOLD changes depends on whether rewards are contingent on actions; correlations are observed in the ventral striatum during passive learning (McClure et al., 2003; O'Doherty et al., 2003), whereas correlations are observed in the dorsal striatum when rewards are contingent on actions (O'Doherty et al., 2004; Haruno and Kawato, 2006). This is consistent with the observation that the caudate nucleus is active only when there is a perceived contingency between actions and outcomes (Tricomi et al., 2004). Correlations with RPEs are thought to reflect inputs from dopamine neurons, consistent with the hypothesis that corticostriatal plasticity promotes the selection of rewarded actions (Houk et al., 1995; Reynolds et al., 2001). Our electrophysiological observations are consistent with this mechanism, but also point to the existence of a temporally distinct signal that reflects the value of the chosen action.

Conclusions

Our data support the hypothesis that striatal PANs encode the values of potential actions, reflecting what subjects learn from the outcomes of past actions. These neurons could promote the selection of rewarded actions through the outflow of the basal ganglia to the thalamus and midbrain. We also observed a novel type of striatal value representation; some PANs encode the reward values associated with the chosen action. This chosen-value activity occurred later in the trial, peaking after movement execution, which suggests that some striatal PANs may play an evaluative role in learning itself.

Experimental Procedures

Subjects & surgery

Two rhesus monkeys (Macaca mulatta) were used as subjects (Monkey B and Monkey H, 10.5 kg and 11.5 kg). All experimental procedures were approved by the New York University Institutional Animal Care and Use Committee and performed in compliance with the Public Health Service's Guide for the Care and Use of Animals.

Prior to training, each animal was implanted with a head-restraint prosthesis and a scleral eye coil. A second surgical procedure was performed to implant a recording chamber (2 cm diameter; Crist Instruments) centered over the body of the caudate nucleus (−3 mm behind the anterior commisure), 5 mm lateral to the midline, and oriented perpendicular to the stereotaxic horizontal plane. Surgical procedures were performed using aseptic techniques under general anaesthesia (Platt and Glimcher, 1997).

Experiments were conducted in a dimly lit sound-attenuated room. Eye movements were measured using a scleral coil (Fuchs and Robinson, 1966) and sampled at 500 Hz. Visual stimuli were generated using light-emitting diodes (LEDs) 145 cm from the monkeys' eyes.

Behavioral task

The monkeys performed a choice task while we varied the rewards associated with two alternatives (Lau and Glimcher, 2005). Each trial started with a 500 ms 500 Hz tone, after which the monkey was given 700 ms to align its gaze within 3° of a yellow LED in the center of the visual field. After maintaining fixation for 400 ms, two peripheral LEDs (one red and one green) were illuminated on either side of the centrally located fixation point. One second later, the central fixation point disappeared, cueing the monkey to choose one of the peripheral LEDs by shifting gaze to within 4° of its location. If a reward had been scheduled for the chosen target, it was delivered 300 ms after the eye movement was completed (defined as when the eye entered the eccentric target window). The timing and appearance of each trial was identical to the monkey whether or not reward was delivered, and the monkey was required to maintain fixation for the duration of the reward epoch (an additional 100–300 ms depending on the reward magnitude) in order for the trial to be considered correctly completed.

Rewards were scheduled using independent and equal arming probabilities for each alternative ($p=0.15$), meaning that an alternative that is not armed on the current trial has a probability of 0.15 of being armed on the next trial. On any trial both alternatives, neither alternative, or only one alternative might be armed to deliver a reward. Importantly, if a reward was scheduled for the alternative the monkey did not choose, it remained available until that alternative was next chosen (no information regarding scheduled rewards was given to the monkeys). This produces contingencies similar to those faced by animals performing under concurrent variable-interval schedules (Nevin, 1969). We did not impose a changeover delay or any other type of penalty for switching between the choice alternatives.

Water delivery was controlled by varying the amount of time a solenoid inline with the water spout was held open. Over the range of magnitudes used, solenoid time was linearly related to the volume of water dispensed, and found to be stable across sessions. In each session, the monkeys performed a series of trials under 4 different conditions in which the ratio of reward magnitudes took one of four values (3:1, 3:2, 2:3, 1:3). The reward magnitudes were constrained to sum to a constant that was the same (0.8 ml) for each ratio in order to minimize fluctuations in motivation from block to block. The monkeys performed blocks of trials at the different relative reward magnitudes. The number of trials in a block was 100 trials plus a random number of trials drawn from a geometric distribution with a mean of 30 trials. Transitions between blocks of trials with different reward ratios were unsignalled. When blocks

were switched, the larger reward always changed spatial location, but its magnitude was variable; the two possible ratios to switch to were chosen with equal probability.

We recorded a trial as aborted if the monkey failed to align its gaze within the required distance of the fixation or cue LEDs, if an eye movement was made prematurely, or if fixation at the peripheral LED was broken prematurely. When an abort was detected, any illuminated LEDs were extinguished immediately, and the next trial began after a 3000–6000 ms time-out.

Electrophysiological recording

For recording, an X-Y positioner (Crist Instruments) and a microdrive (Kopf Instruments) were mounted to the recording chamber. A 23-gauge sharpened guide tube housing a tungsten steel electrode (2–4 MΩ, FHC) was used to pierce the dura. The guide tube was lowered until its tip was above or just lateral to the cingulate sulcus as predetermined using MRI (3T; Siemens) in Monkey B and B-mode ultrasound imaging (General Electronics; Glimcher et al., 2001) in Monkey H. In the caudate, we distinguished PANs from tonically active neurons based on differences in spontaneous activity, spike waveform, and response to reward (Kimura et al., 1984; Hikosaka et al., 1989a; Aosaki et al., 1994). If we judged a PAN to be responsive in a delayed saccade task—by observing a phasic response during a trial—we collected data during the matching task. If the neuron was tuned to the location of targets placed in the visual field or saccadic eye movements, we placed one of the target LEDs at the approximate location that elicited the largest response and the other in the opposite hemifield, otherwise the target LEDs were positioned to the left and right of the fixation point at a eccentricity of 12–16°. Approximately 25% of PANs we encountered were not responsive during the delayed saccade task. This underestimates the number of non-responsive neurons since PANs have low baseline firing rates (0–3 spks/s; eg., Hikosaka et al., 1989a), and we likely overlooked many non-responsive PANs.

Recording sites were verified histologically in one monkey. Some of the neurons included in this report were also recorded during an instructed saccade task, and structural MRI and camera lucida drawings can be found in a paper focusing on that task (Lau and Glimcher, 2007).

Data analysis

The goal of our analysis was to determine whether the activity of individual PANs covaried with changes in action-value or chosen-value. These value representations are related, and we used multiple linear regression to differentiate between them. We used the programmed reward magnitudes to estimate the values the monkeys' associated with each action in our blockwise analysis. Value covariates for the regression were constructed in the following manner. The action-values of the contralateral ($\mathbf{AV}_C$) and ipsilateral ($\mathbf{AV}_I$) alternatives were defined as the programmed reward magnitudes associated with each of the alternatives. The only difference between the programmed reward magnitudes and these blockwise action-values was that at block transitions, the action-values changed only after the monkey had received the first reward at the new programmed reward magnitudes. This is to account for the fact that the monkey could not have known the reward magnitudes were changed until actually receiving a reward. The chosen-values of the contralateral ($\mathbf{CV}_C$) and ipsilateral ($\mathbf{CV}_I$) alternatives were defined as equal to $\mathbf{AV}_C$ and $\mathbf{AV}_I$ when the monkey chose the associated action, and zero otherwise (Figure 2A). That is, $\mathbf{CV}_C = \mathbf{AV}_C \times \mathbf{A}_C$ and $\mathbf{CV}_I = \mathbf{AV}_I \times \mathbf{A}_I$ where $\mathbf{A}_C$ and $\mathbf{A}_I$ are binary variables indicating contralateral and ipsilateral choices respectively. The responses of individual neurons were fit using the following multiple linear regression,

$$\mathbf{y} = \alpha_1 \mathbf{A}_C + \alpha_2 \mathbf{A}_I + \alpha_3(\mathbf{CV}_C - \mathbf{CV}_I) + \alpha_4(\mathbf{CV}_C + \mathbf{CV}_I) \quad (1)$$

where **y** is the firing rate. Bold-faced variables represent vectors where each element is the corresponding variable on a particular trial; for example, $\mathbf{y} = [y_1, y_2 \ldots y_N]$ where N is the number of trials. The constant term is implicit since the contralateral and ipsilateral choices sum to unity. Neurons were categorized as follows: 1) *non-value* if both α_3 and α_4 were not significant, 2) *action-value* if α_3 was significant but α_4 was not significant and 3) *chosen-value* if α_4 was significant. Statistical significance for this categorization was determined using incremental *F*-statistic with an alpha level of 0.05. Movement selectivity was assessed by testing whether α_1 and α_2 were equal (*F*-test).

To see how Equation 1 distinguishes between action-value and chosen value neurons, note that $\mathbf{CV}_C$ and $\mathbf{CV}_I$ do not overlap; one is positive whenever the other is zero and vice versa (Figure 2A). By design, the magnitudes of the rewards available from the two alternatives sum to a constant ($\mathbf{AV}_C + \mathbf{AV}_I = 0.8$ ml) within and across blocks, which means that $\mathbf{AV}_C = \mathbf{CV}_C + 0.8 \times \mathbf{A}_I - \mathbf{CV}_I$ and $\mathbf{AV}_I = \mathbf{CV}_I + 0.8 \times \mathbf{A}_C - \mathbf{CV}_C$. Thus, an action-value neuron ($\alpha_3 \neq 0$ and $\alpha_4 = 0$ in Equation 1) can be rewritten using substitution as

$$\begin{aligned}\mathbf{y} &= \alpha_1\mathbf{A}_C + \alpha_2\mathbf{A}_I + \alpha_3(\mathbf{CV}_C - \mathbf{CV}_I)\\ &= \alpha_1\mathbf{A}_C + \alpha_2^*\mathbf{A}_I + \alpha_3\mathbf{AV}_C,\end{aligned}$$

where $\alpha_2^* = (\alpha_2 - 0.8 \times \alpha_3)$, and a chosen-value neuron ($\alpha_4 \neq 0$ in Equation 1) can be rewritten as

$$\begin{aligned}\mathbf{y} &= \alpha_1\mathbf{A}_C + \alpha_2\mathbf{A}_I + \alpha_3(\mathbf{CV}_C - \mathbf{CV}_I) + \alpha_4(\mathbf{CV}_C + \mathbf{CV}_I)\\ &= \alpha_1\mathbf{A}_C + \alpha_2\mathbf{A}_I + \alpha_3^*\mathbf{CV}_C + \alpha_4^*\mathbf{CV}_I\end{aligned}$$

where $\alpha_3^* = \alpha_3 + \alpha_4$ and $\alpha_4^* = -\alpha_3 + \alpha_4$. Note that while Equation 1 can determine whether a neuron significantly covaries with action-value, it cannot distinguish covariation of firing rate with $\mathbf{AV}_C$ from covariation of firing rate with $\mathbf{AV}_I$. Since $\mathbf{AV}_C$ and $\mathbf{AV}_I$ sum to a constant, a model with α_3 for $\mathbf{AV}_C$ is equivalent to a model with $-\alpha_3$ for $\mathbf{AV}_I$. In order to determine which of the alternatives an action-value neuron encodes, we used a reinforcement learning model (see below) to generate behavioral value estimates that discriminated between contralateral and ipsilateral action-values.

We further assessed the value representations in single neurons using a second regression that incorporated dynamic value estimates derived from a model of choice behavior. Our behavioral model predicted trial-by-trial choices based on linear weightings of past rewards, currently available rewards, and past choices (Lau and Glimcher, 2005). We used the coefficients of this behavioral model—fit separately to the choice data pooled across behavioral sessions for each monkey—to estimate the expected value (to the monkey) of each alternative on trial-by-trial basis. From these trial-by-trial value estimates, we constructed another set of regression covariates. Just as for the regressions above, we have action-values ($\overline{\mathbf{AV}}_C$ and $\overline{\mathbf{AV}}_I$) and chosen-values ($\overline{\mathbf{CV}}_C$ and $\overline{\mathbf{CV}}_I$) for each alternative, although in this case, these values represent dynamic estimates of these properties. For chosen-value neurons, we fit the following model:

$$\mathbf{y} = \alpha_1\mathbf{A}_C + \alpha_2\mathbf{A}_I + \alpha_3\overline{\mathbf{CV}}_C + \alpha_4\overline{\mathbf{CV}}_I. \quad (2)$$

For action-value neurons, we fit the data using trial-by-trial estimates of action-value as covariates:

$$\mathbf{y} = \alpha_1\mathbf{A}_C + \alpha_2\mathbf{A}_I + \alpha_3\overline{\mathbf{AV}}_C + \alpha_4\overline{\mathbf{AV}}_I. \quad (3)$$

Since the trial-by-trial action-value estimates vary according to the stochastic delivery of rewards to each alternative, $\overline{\mathbf{AV}}_C$ and $\overline{\mathbf{AV}}_I$ are linearly independent and do not sum to a constant,

which allows us to determine whether individual neurons encode the action-value of the contralateral or ipsilateral alternative, or some mixture of both.

We included additional covariates to Equations 1–3 to protect against potential confounds due to variables correlated with value. To control for correlations with movement metrics, we included the reaction time (**RT**) and peak velocity (**VEL**) of the eye movement measured on each trial. We also included a covariate for the magnitude of the reward obtained (**R**) on each trial. This ensured that neurons that simply responded to obtained reward were not erroneously deemed value neurons. For the regression analyses using a tailored temporal window for each neuron we included **R** if the peak response of a neuron followed saccade completion. For the regression analyses using smaller fixed windows throughout the trial, we included **R** in all windows following the target acquisition. Coefficients for these additional variables were fit simultaneously with the other covariates to ensure that the occasional correlations between covariates were accounted for.

In order to examine in detail the effect of specific variables on firing rate, we used partial residuals, which are model residuals that are not adjusted for the effect of the particular covariate of interest (Larsen and McCleary, 1972). For example, partial residuals for contralateral action-value $(\overline{\mathbf{AV}}_C)$ were be computed as follows

$$\varepsilon_3 = \mathbf{y} - (\widehat{\alpha}_1\mathbf{A}_C + \widehat{\alpha}_2\mathbf{A}_I + \widehat{\alpha}_4\overline{\mathbf{AV}}_I + \widehat{\alpha}_5\mathbf{RT} + \widehat{\alpha}_6\mathbf{VEL} + \widehat{\alpha}_7\mathbf{R}) \quad (4)$$

Where $\hat{\alpha}_{1-7}$ are the coefficients estimated from fitting Equation 3 with the movement metrics and obtained reward. Plotting ε_3 against $\overline{\mathbf{AV}}_C$ is known as a partial residual plot, and the coefficient for action value ($\hat{\alpha}_3$), is equal to the slope of the best-fit line through the residuals in this plot. Partial residual plots directly reveal the relationship between the variable of interest and firing rate, after controlling for the influence of all other variables in the regression. We term these partial residuals the *effect* of a particular variable on firing rate, since they indicate the change in firing rate (in the same units of spks/s as the original response) due to that variable.

In the first portion of this report, we analyzed the value representations of individual neurons. To do this, we estimated statistics using a single temporal window tailored for each neuron (Lau and Glimcher, 2007). First, we estimated spike density functions for each movement direction using a Gaussian smoothing window. The degree of smoothing was chosen to maximize the information gain per spike (Paulin and Hoffman, 2001; see Supplementary materials) for each neuron. Next, the peak response for each neuron was estimated as the maximum of whichever spike density function (corresponding to contralateral or ipsilateral choices) had the largest response. Finally, we defined an analysis window using the first times to half-maximum preceding and following the peak response.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to Mehrdad Jazayeri, Kenway Louie and Marianna Yanike for helpful discussions. This work was supported by a NDSEG fellowship to BL and NEI EY010536 to PWG.

References

Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci 1986;9:357–81. [PubMed: 3085570]

Aosaki T, Tsubokawa H, Ishida A, Watanabe K, Graybiel AM, Kimura M. Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. J Neurosci 1994;14:3969–3684. [PubMed: 8207500]

Apicella P, Ljungberg T, Scarnati E, Schultz W. Responses to reward in monkey dorsal and ventral striatum. Exp Brain Res 1991;85:491–500. [PubMed: 1915708]

Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM. Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature 2005;437:1158–1161. [PubMed: 16237445]

Barraclough DJ, Conroy ML, Lee D. Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci 2004;7:404–410. [PubMed: 15004564]

Bayer H, Glimcher P. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 2005;47:129–141. [PubMed: 15996553]

Corrado GS, Sugrue LP, Seung HS, Newsome WT. Linear-nonlinear-poisson models of primate choice dynamics. J Exp Anal Behav 2005;84:581–617. [PubMed: 16596981]

Cromwell HC, Schultz W. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. J Neurophysiol 2003;89:2823–2838. [PubMed: 12611937]

Davison, M.; McCarthy, D. The matching law: A research review. Hillsdale, NJ: Erlbaum; 1988.

Daw ND, Doya K. The computational neurobiology of learning and reward. Curr Opin Neurobiol 2006;16:199–204. [PubMed: 16563737]

Ding L, Hikosaka O. Comparison of reward modulation in the frontal eye field and caudate of the macaque. J Neurosci 2006;26:6695–6703. [PubMed: 16793877]

Dorris MC, Glimcher PW. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron 2004;44:365–378. [PubMed: 15473973]

Doya K. Complementary roles of basal ganglia and cerebellum in learning and motor control. Curr Opin Neurobiol 2000;10:732–739. [PubMed: 11240282]

Fuchs AF, Robinson DA. A method for measuring horizontal and vertical eye movement chronically in the monkey. J Appl Physiol 1966;21:1068–1070. [PubMed: 4958032]

Gerfen CR, Herkenham M, Thibault J. The neostriatal mosaic: II. patch- and matrix-directed mesostriatal dopaminergic and non-dopaminergic systems. J Neurosci 1987;7:3915–3934. [PubMed: 2891799]

Glimcher PW, Ciaramitaro VM, Platt ML, Bayer HM, Brown MA, Handel A. Application of neurosonography to experimental physiology. J Neurosci Methods 2001;108:131–144. [PubMed: 11478972]

Graybiel, AM.; Penney, JB. Handbook of Chemical Neuroanatomy, Volume 15: The Primate Nervous System, Part I, (Elsevier). 1999. Chemical architecture of the basal ganglia; p. 227-284.

Haruno M, Kawato M. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. J Neurosci 2006;95:948–959.

Hassani OK, Cromwell HC, Schultz W. Influence of expectation of different rewards on behavior related neuronal activity in the striatum. J Neurophysiol 2003;85:2477–2489. [PubMed: 11387394]

Herrnstein RJ. Relative and absolute strength of response as a function of frequency of reinforcement. J Exp Anal Behav 1961;4:267–272. [PubMed: 13713775]

Herrnstein, RJ.; Vaughan, W. Melioration and behavioral allocation. In: Staddon, J., editor. Limits to Action: The Allocation of Individual Behavior. New York: Academic Press; 1980. p. 143-176.

Hikosaka O, Sakamoto M, Usui S. Functional properties of monkey caudate neurons. I activities related to saccadic eye movements. J Neurophysiol 1989a;61:780–798. [PubMed: 2723720]

Hikosaka O, Sakamoto M, Usui S. Functional properties of monkey caudate neurons. III activities related to expectation of target and reward. J Neurophysiol 1989b;61:814–832. [PubMed: 2723722]

Hikosaka O, Nakamura K, Nakahara H. Basal ganglia orient eyes to reward. J Neurophysiol 2006;95:567–584. [PubMed: 16424448]

Houk, JC.; Adams, JL.; Barto, AG. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Houk, JC.; Davis, JL.; Beiser, DG., editors. Models of information processing in the basal ganglia. Cambridge Mass: The MIT Press; 1995. p. 249-270.

Itoh H, Nakahara H, Hikosaka O, Kawagoe R, Takikawa Y, Aihara K. Correlation of primate caudate neural activity and saccade parameters in reward-oriented behavior. J Neurophysiol 2003;89:1774–1783. [PubMed: 12686566]

Joel D, Weiner I. The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. Neuroscience 2000;96:451–474. [PubMed: 10717427]

Kawagoe R, Takikawa Y, Hikosaka O. Expectation of reward modulates cognitive signals in the basal ganglia. Nat Neurosci 1998;1:411–6. [PubMed: 10196532]

Kimura M, Rajkowski J, Evarts E. Tonically discharging putamen neurons exhibit set-dependent responses. Proc Natl Acad Sci USA 1984;81:4998–5001. [PubMed: 6589643]

Kobayashi S, Kawagoe R, Takikawa Y, Koizumi M, Sakagami M, Hikosaka O. Functional differences between macaque prefrontal cortex and caudate nucleus during eye movements with and without reward. Exp Brain Res 2007;176:341–355. [PubMed: 16902776]

Larsen WA, McCleary SJ. The use of partial residual plots in regression analysis. Technometrics 1972;14:781–790.

Lau B, Glimcher PW. Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 2005;84:555–579. [PubMed: 16596980]

Lau B, Glimcher PW. Action and outcome encoding in the primate caudate nucleus. J Neurosci 2007;27:14502–14514. [PubMed: 18160658]

Lauwereyns J, Takikawa Y, Kawagoe R, Kobayashi S, Koizumi M, Coe B, Sakagami M, Hikosaka O. Feature-based anticipation of cues that predict reward in monkey caudate nucleus. Neuron 2002a; 33:463–473. [PubMed: 11832232]

Lauwereyns J, Watanabe K, Coe B, Hikosaka O. A neural correlate of response bias in monkey caudate nucleus. Nature 2002b;418:413–417. [PubMed: 12140557]

Luce, RD. Individual choice behavior; a theoretical analysis. New York: Wiley; 1959.

McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. Neuron 2003;38:339–346. [PubMed: 12718866]

Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 2006;9:1057–1063. [PubMed: 16862149]

Nevin J. Interval reinforcement of choice behavior in discrete trials. J Exp Anal Behav 1969;12:875–885. [PubMed: 16811416]

Niv Y, Daw ND, Dayan P. Choice values. Nat Neurosci 2006;9:987–988. [PubMed: 16871163]

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward related learning in the human brain. Neuron 2003;38:329–337. [PubMed: 12718865]

O'Doherty JP, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in intrumental conditioning. Science 2004;304:452–454. [PubMed: 15087550]

Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. Nature 2006;441:223–226. [PubMed: 16633341]

Pasupathy A, Miller EK. Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 2005;433:873–876. [PubMed: 15729344]

Paulin MG, Hoffman LF. Optimal firing rate estimation. Neural Netw 2001;14:877–881. [PubMed: 11665778]

Platt ML, Glimcher PW. Responses of intraparietal neurons to saccadic targets and visual distractors. J Neurophysiol 1997;78:1574–1589. [PubMed: 9310444]

Reynolds JN, Hyland BI, Wickens JR. A cellular mechanism of reward-related learning. Nature 2001;413:67–70. [PubMed: 11544526]

Reynolds JN, Hyland BI, Wickens JR. A cellular mechanism of reward-related learning. Nature 2001;413:67–70. [PubMed: 11544526]

Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differentially delayed or sized rewards. Nat Neurosci 2007;10:1615–1624. [PubMed: 18026098]

Schultz W. Multiple reward signals in the brain. Nat Rev Neurosci 2000;1:199–207. [PubMed: 11257908]

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997;275:1593–1599. [PubMed: 9054347]

Selemon LD, Goldman-Rakic PS. Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. J Neurosci 1985;5:776–794. [PubMed: 2983048]

Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. Science 2004;304:1782–1787. [PubMed: 15205529]

Sugrue LP, Corrado GS, Newsome WT. Choosing the greater of two goods: neural currencies for valuation and decision making. Nat Rev Neurosci 2005;6:363–375. [PubMed: 15832198]

Sutton, RS.; Barto, AG. Reinforcement learning: An introduction. Cambridge, MA: The MIT Press; 1998.

Tremblay L, Hollerman JR, Schultz W. Modifications of reward expectation-related neuronal activity during learning in primate striatum. J Neurophysiol 1998;80:964–977. [PubMed: 9705482]

Tricomi EM, Delgado MR, Fiez JA. Modulation of caudate activity by action contingency. Neuron 2004;41:281–292. [PubMed: 14741108]

Watanabe K, Lauwereyns J, Hikosaka O. Neural correlates of rewarded and unrewarded eye movements in the primate caudate nucleus. J Neurosci 2003;23:10052–10057. [PubMed: 14602819]

Wild, RJ.; Seber, GAF. Change encounters: A first course in data analysis and inference. New York, NY: Wiley; 2000.

Williams, B. Reinforcement, choice, and response strength. In: Atkinson, RC.; Herrnstein, RJ.; Lindzey, G.; Luce, RD., editors. Stevens's Handbook of Experimental Psychology. 2. New York, NY: Wiley; 1988. p. 167-244.

Williams ZM, Eskandar EN. Selective enhancement of associative learning by microstimulation of the anterior caudate. Nat Neurosci 2006;9:562–568. [PubMed: 16501567]

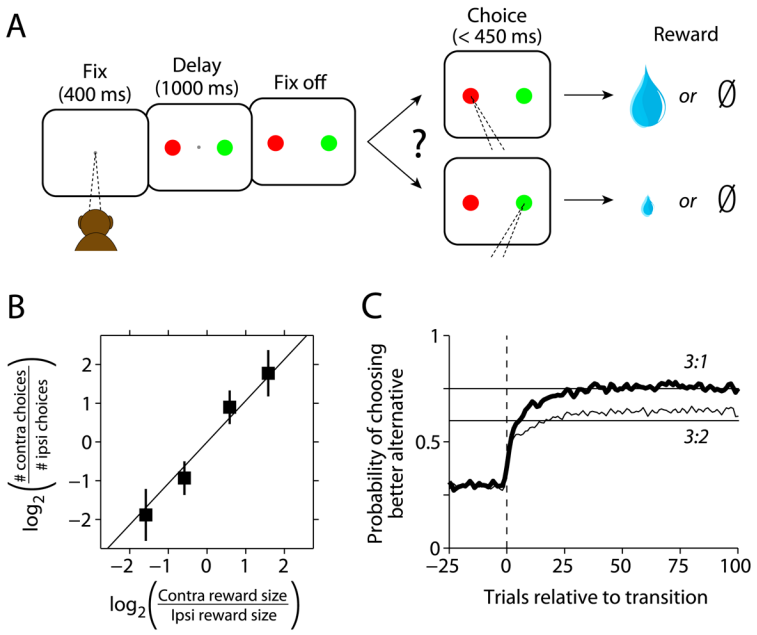

Figure 1.

Matching task and behavior. A) On each trial monkeys chose between two alternatives by shifting gaze to one of two peripheral targets. Rewards were delivered probabilistically, with one alternative yielding a larger reward than the other in each block. Reward contingencies were constant over blocks of trials and switched unpredictably. B) Monkeys allocate choices according to the relative magnitude of the available alternatives (mean ± 1 SD). The data points represent an average across sessions and monkeys. The four data points correspond to the 4 reward ratios used. C) Stable choice behavior emerges quickly following transitions to different ratios. Choice behavior from both monkeys aligned on the trial that the first reward after a block transition was obtained (smoothed with a 5-point moving average). The data were compiled across sessions with respect to the alternative associated with the larger reward following the transition, and averaged separately for the different post-transition ratios (3:1 is averaged together with 1:3 and 3:2 is averaged together with 2:3). The horizontal lines illustrate strict matching behavior (Herrnstein, 1961).

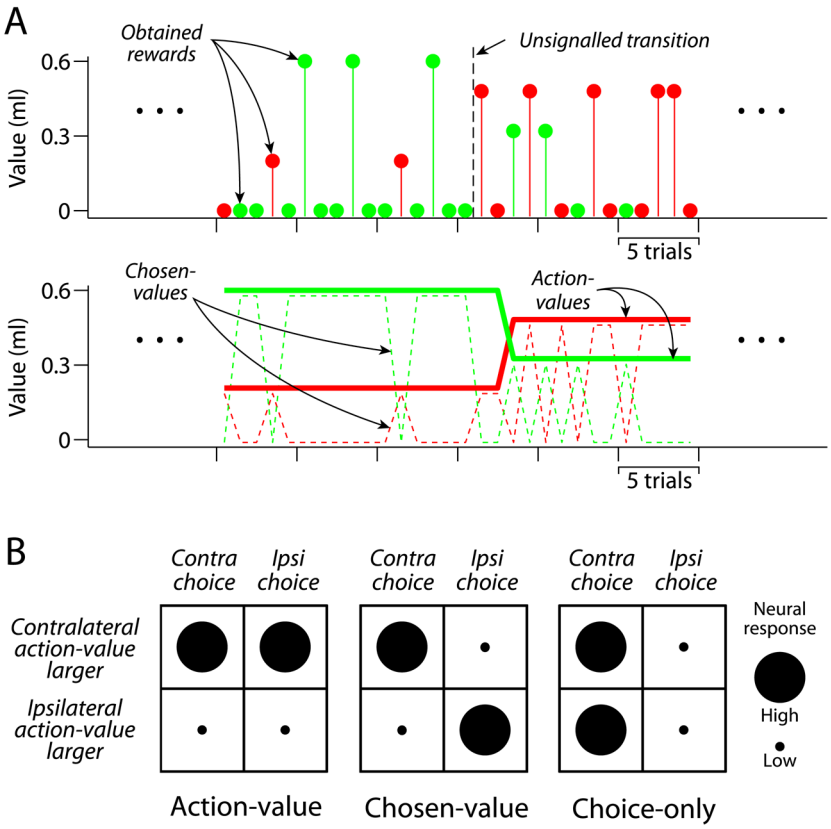

Figure 2.

Value representations and neuronal categories. A) The top panel shows a segment of actual behavior, including a block transition. The red and green points indicate individual choices to the contralateral and ipsilateral alternatives respectively. The value of each point on the ordinate indicates the magnitude of the obtained rewards. Since rewards were delivered probabilistically, there are frequently trials that are not rewarded. The middle panel illustrates value representations inferred from average behavior. Plotted are blockwise estimates of action-values and chosen-values corresponding to the choice behavior shown in the top panel. B) Exemplars of hypothesized neuron types. Note that the figure illustrates only one type of chosen-value neuron (see text for details).

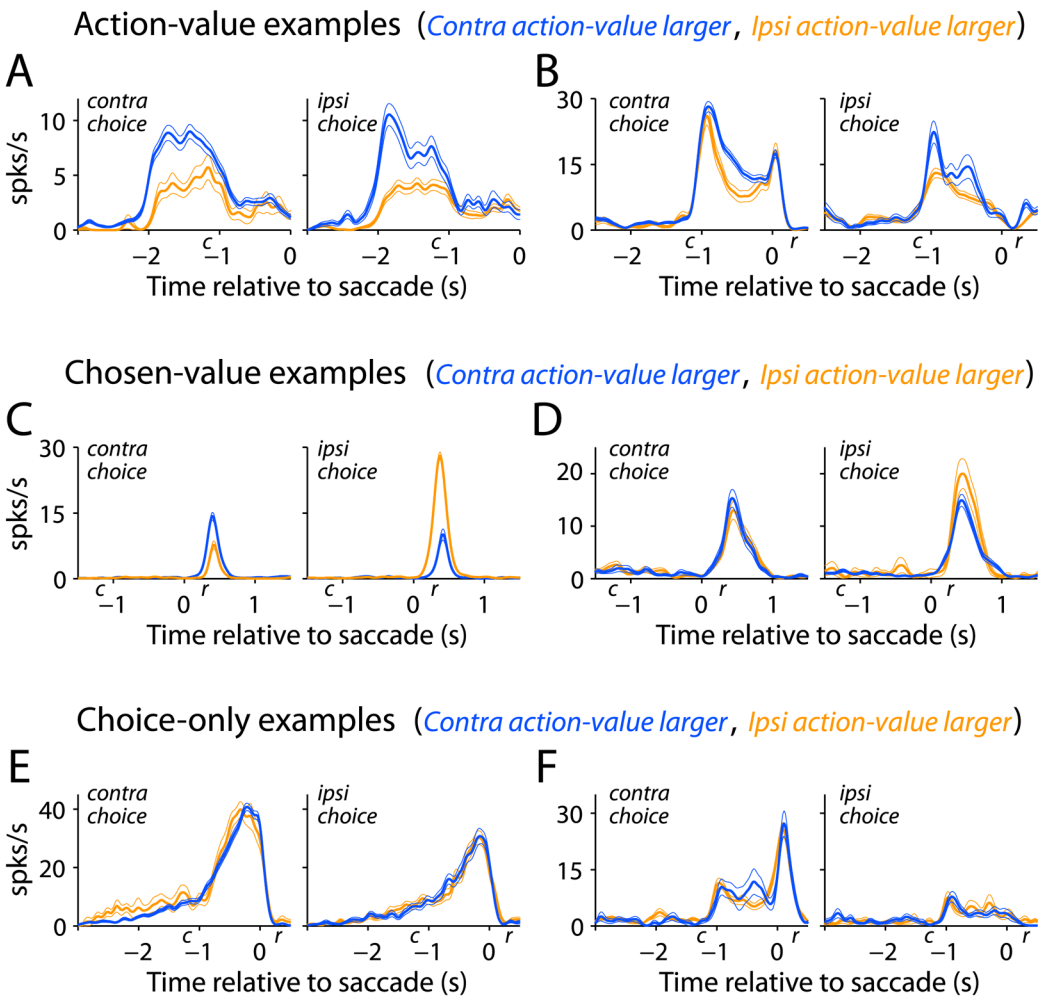

Figure 3.

Example neurons from the categories illustrated in Figure 2. A–B) Action-value neurons (Gaussian smoothing σ =62ms and 23 ms). C–D) Two types of chosen-value neuron, one that is sensitive to the reward associated with both actions, and the other sensitive to the reward associated with only one action when it is chosen (σ =42ms and 63 ms). E–F) Choice-only neurons (σ =47ms and 48 ms). The plots are spike density functions (mean ± 1 SEM) sorted according to two factors, choice (movement direction) and relative action-value. Reward onset is denoted by *r* and was a fixed time relative to saccade completion. Cue onset is denoted by *c* is the average cue onset time relative to saccade completion.

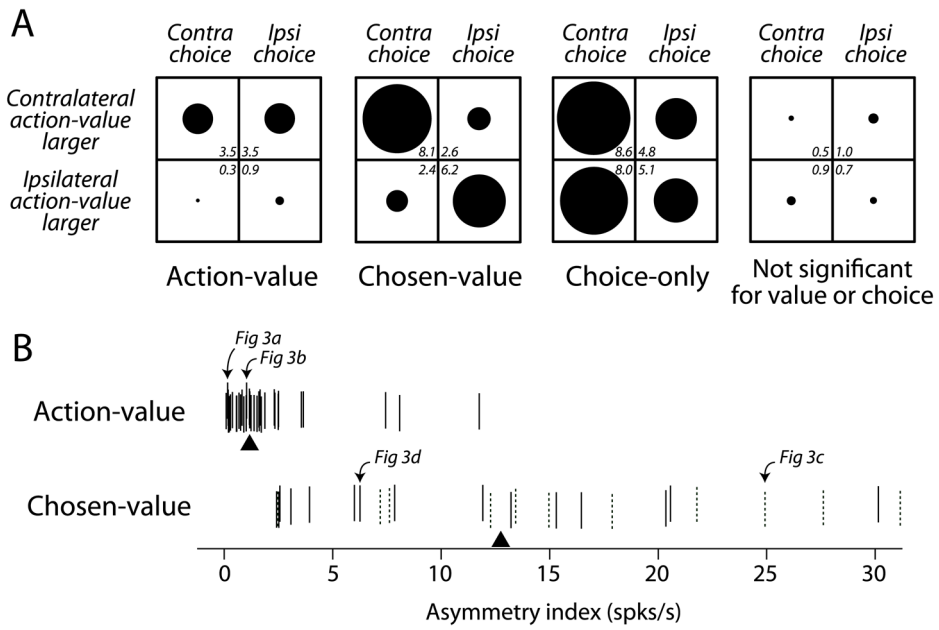

Figure 4.

Population summary of value sensitivity. A) Mean effect averaged for different neuronal categories. The radius of each symbol is proportional to the mean effect listed in the corner of each quadrant in each plot (in units of spks/s). B) Asymmetry index for individual action-value and chosen value neurons. This index is calculated by summing the absolute values of the row-wise differences of the quadrant plot for each neuron. Each tick represents one neuron (jittered vertically for visibility), and the triangles indicate the median for each distribution. The dashed ticks indicate those chosen-value neurons that reflected the chosen-values of both actions.

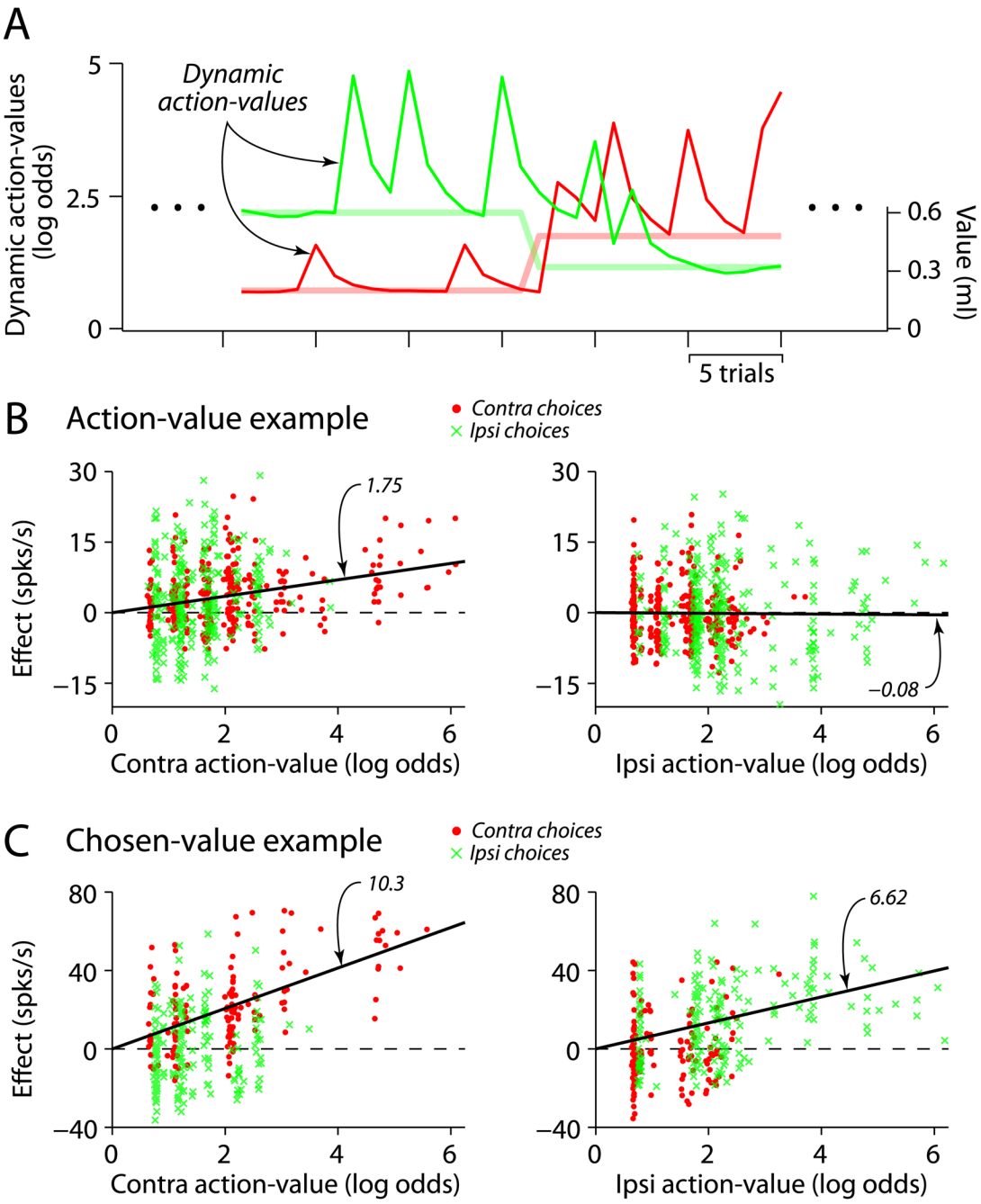

Figure 5.

Caudate neurons are correlated with dynamically estimated values. A) Trial-by-trial action-value estimates corresponding to the example behavior illustrated in Figure 2A. These estimates are plotted in units of log-odds (left axis), which is logarithm of the ratio formed by probability of choosing one action divided by the probability of choosing the alternative action. The programmed reward magnitudes are plotted as lighter lines (right axis). B) Example action-value neuron. The left and right panels plot the effect of changes in contralateral and ipsilateral action-values on firing rate respectively, after subtracting the effects due other regression variables (see Experimental Procedures). The symbols distinguish choices to each alternative; on the left, the red points extend further along the abscissa since the monkey more oftepn chose

the contalateral alternative when its value was high (vice versa for the green points on the right panel). The black lines illustrate the slope predicted from the regression analysis using dynamic values generated by a behavioral model. Slopes for the separate choices are 1.71 (contra) and 1.85 (ipsi) for the left panel and −0.14 (contra) and −0.05 (ipsi) for the right panel. C) Example chosen-value neuron, conventions as in C). Slopes for the separate choices are 10.3 (contra) and 0.79 (ipsi) for the left panel and 0.69 (contra) and 6.62 (ipsi) for the right panel.

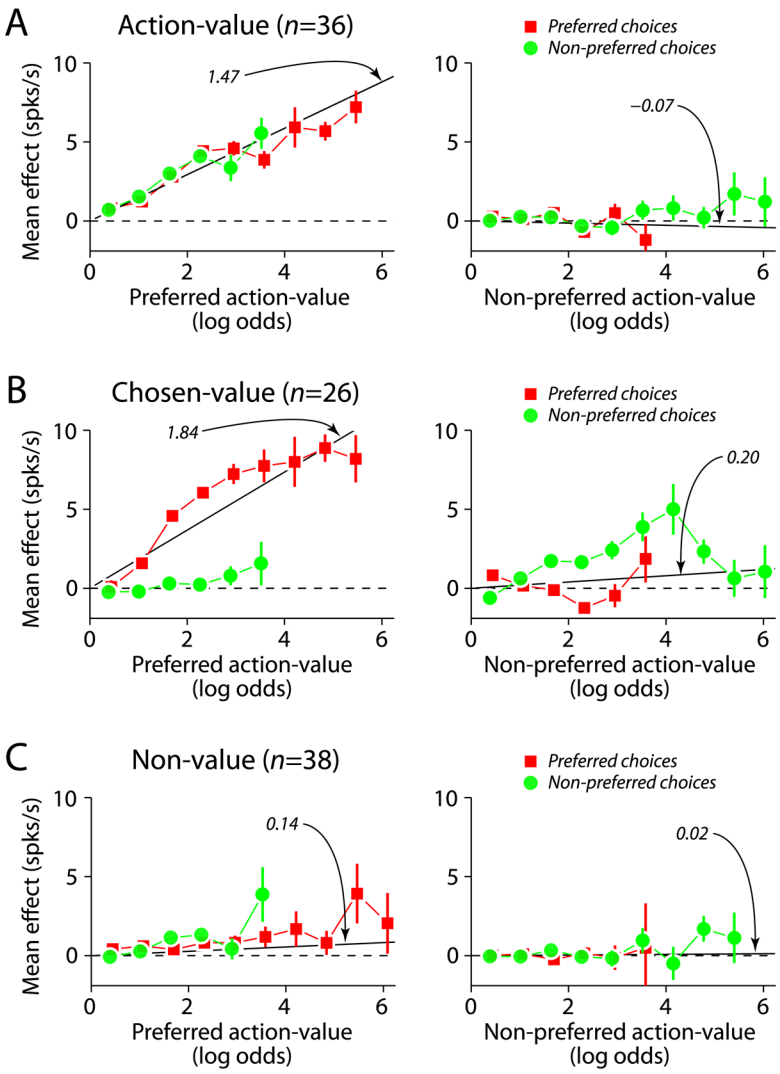

Figure 6.

Population summary of value encoding. A) Action-value. B) Chosen-value. C) Non-value. Each row represents the mean effects of changes in action-value on firing rates, binned and averaged over all trials and all neurons for each category. The averages were constructed with respect to the preferred action-value, which was defined as the alternative for which the absolute value coefficient was largest. The black line in each panel is the median value coefficient for the corresponding panel.

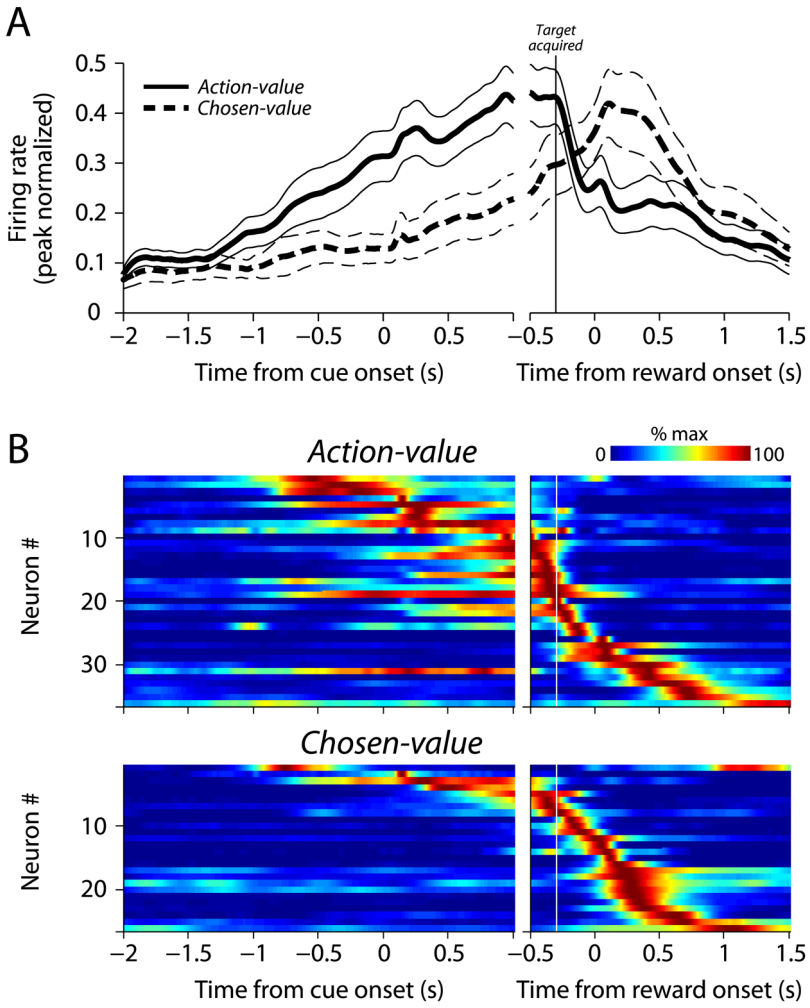

Figure 7.

The temporal profiles of action-value and chosen-value neurons are different. A) The average population response for action-value and chosen-value PANs (thick line = mean, thin lines = ±1 SEM). For each neuron, spike density functions were estimated from the trials in which saccades were made in the direction that elicited the largest response (averaged over blocks). The individual spike density functions were peak-normalized (divided by maximum activity) and then averaged to produce the population response. B) Individual spike density functions (peak-normalized) for action-value and chosen-value neurons, sorted by peak response time.

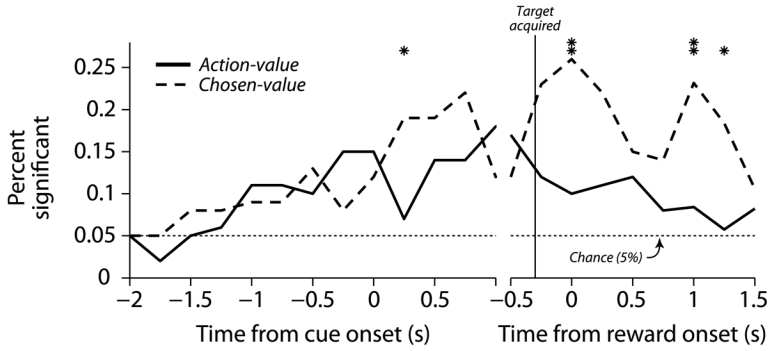

Figure 8.

Sliding-window regression summary. The percentage of significant neurons is aligned to cue onset and reward onset in the left and right panels, respectively. The non-overlapping 250 ms bins where the proportion of significant neurons differs for the two curves are indicated with an asterix (one asterix indicates $p<0.05$ and two indicates $p<0.01$, z-test).

Table 1

Regression summary broken down by value representation. The percentages listed in parentheses are relative to the row total. Significance level is 0.05.

	Total	Choice	Reaction time	Peak velocity	Obtained Reward
Action-value	36	31 (86%)	11 (31%)	13 (36%)	9 (25%)
Chosen-value	26	12 (46%)	7 (27%)	2 (8%)	13 (50%)
Non-value	38	29 (76%)	9 (24%)	6 (15%)	12 (31%)
	100	72	27	21	34