Behavioral/Systems/Cognitive

Ventral Striatum Encodes Past and Predicted Value Independent of Motor Contingencies

Brandon L. Goldstein,[1]* **Brian R. Barnett,**[1]* **Gloria Vasquez,**[1] **Steven C. Tobia,**[1] **Vadim Kashtelyan,**[1] **Amanda C. Burton,**[1] **Daniel W. Bryden,**[1,2] **and Matthew R. Roesch**[1,2]

[1]Department of Psychology and [2]Program in Neuroscience and Cognitive Science, University of Maryland, College Park, Maryland 20742

The ventral striatum (VS) is thought to signal the predicted value of expected outcomes. However, it is still unclear whether VS can encode value independently from variables often yoked to value such as response direction and latency. Expectations of high value reward are often associated with a particular action and faster latencies. To address this issue we trained rats to perform a task in which the size of the predicted reward was signaled before the instrumental response was instructed. Instrumental directional cues were presented briefly at a variable onset to reduce accuracy and increase reaction time. Rats were more accurate and slower when a large versus small reward was at stake. We found that activity in VS was high during odors that predicted large reward even though reaction times were slower under these conditions. In addition to these effects, we found that activity before the reward predicting cue reflected past and predicted reward. These results demonstrate that VS can encode value independent of motor contingencies and that the role of VS in goal-directed behavior is not just to increase vigor of specific actions when more is at stake.

Introduction

Traditionally, ventral striatum (VS) has been thought of as a "limbic–motor" interface (Mogenson et al., 1980), a hypothesis that was originally derived from its connectivity with limbic and motor output regions (Groenewegen and Russchen, 1984; Heimer et al., 1991; Brog et al., 1993; Wright and Groenewegen, 1995; Voorn et al., 2004; Gruber and O'Donnell, 2009). Through these connections, the ventral striatum is thought to integrate information about the value of expected outcomes with motor information to guide motivated behavior. Consistent with this proposal, lesions of VS impair changes in response latencies associated with different quantities of reward and impact other behavioral measures of vigor, salience and arousal that reflect the value of expected rewards (Berridge and Robinson, 1998; Hauber et al., 2000; Cardinal et al., 2002a,b; Di Chiara, 2002; Giertler et al., 2003).

More recently, it has been suggested that predicted value signals generated in VS might be used for functions other than energizing actions (van der Meer and Redish, 2011). In these models, downstream brain areas receive predictive value signals from VS so that reinforcement learning (actor-critic) and decision making (good-based economic choice) can occur (Barto, 1995; Houk et al., 1995; Sutton and Barto, 1998; Joel et al., 2002; Redish, 2004; Niv and Schoenbaum, 2008; Takahashi et al., 2008; Padoa-Schioppa, 2011). Unlike models that suggest that the function of VS is to interface value with motor output, these models require that value be represented independently from motor contingencies.

Unfortunately, it is still unclear whether VS can represent value in this way because studies examining activity in VS have either varied expected reward value or the instrumental response or have manipulated both simultaneously (Schultz et al., 1992; Carelli and Deadwyler, 1994; Bowman et al., 1996; Shidara et al., 1998; Hassani et al., 2001; Carelli, 2002; Cromwell and Schultz, 2003; Setlow et al., 2003; Janak et al., 2004; Nicola et al., 2004; Shidara and Richmond, 2004; Taha and Fields, 2006; German and Fields, 2007; Hollander and Carelli, 2007; Simmons et al., 2007; Takahashi et al., 2007; Robinson and Carelli, 2008; Ito and Doya, 2009; H. Kim et al., 2009; Minamimoto et al., 2009; van der Meer and Redish, 2009; van der Meer et al., 2010; Day et al., 2011). Further, in these studies, better rewards are almost always associated with faster reaction times. In fact, many studies use speeded reaction times as evidence that animals value one reward over another. This is true of single-unit recording studies and the majority of studies that examine behavior after VS inactivation or lesions. Thus, predicted reward and motor output signals have been intertwined in a way that makes it difficult to dissociate encoding of value from the direction and speed of action initiation.

To address this issue we designed a new task in which rats learned about expected outcomes before knowing the action necessary to acquire it. In addition, we designed the task so that rats reacted slower to stimuli that predicted larger rewards. We did this by instructing the behavior response with a temporally unpredictable short duration directional light cue. In general, we found that reducing the length and predictability of the directional cue reduced accuracy on the task and slowed reaction times. When a larger reward was at stake, rats were significantly
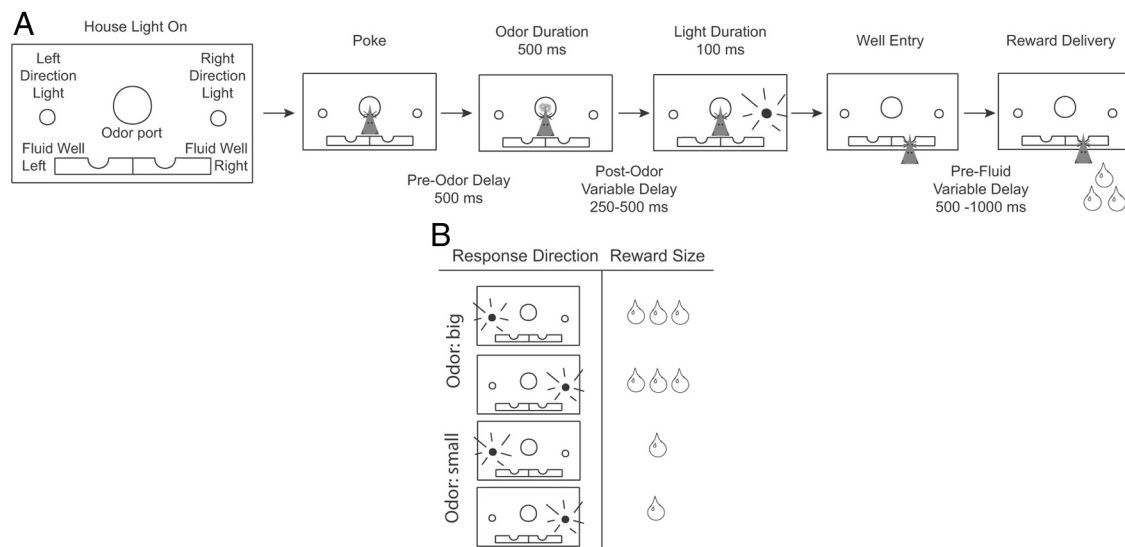
Figure 1. Task design. *A*, House lights signaled the rat to nose poke into the center odor port and wait 500 ms before odor delivery. The odor indicating the size (large or small) of the reward to be delivered at the end of the trial. Odor presentation lasted 500 ms and was followed by a 250–500 ms post-odor variable delay, which ended with the onset of directional cue lights. Directional lights flashed for 100 ms on either the left or right, instructing the rat to respond to the left or right fluid well, respectively. After the entering the correct fluid well rats were required to wait 500–1000 ms before reward delivery. *B*, There were four possible reward size and response direction combinations; large-left, large-right, small-left and small-right.

slower and more accurate than when a small reward was at stake. We found that activity in VS reflected the value of the expected reward before cuing of response direction and that activity was high even though reaction times were slower. Surprisingly, we also found that activity in VS did not just reflect predicted value on the upcoming behavioral trial, but was also modulated by the size of the reward on the previous trial.

Materials and Methods

Subjects. Male Long–Evans rats were obtained at 175–200 g from Charles River Labs. Rats were tested at the University of Maryland in accordance with NIH and Institutional Animal Care and Use Committee guidelines.

Surgical procedures and histology. Surgical procedures followed guidelines for aseptic technique. Electrodes were manufactured and implanted as in prior recording experiments. Rats had a drivable bundle of ten 25-μm-diameter FeNiCr wires (Stablohm 675, California Fine Wire) chronically implanted in the left hemisphere dorsal to VS ($n = 6$; 1.6 mm anterior to bregma, 1.5 mm laterally, and 4.5 mm ventral to the brain surface). Immediately before implantation, these wires were freshly cut with surgical scissors to extend ~1 mm beyond the cannula and electroplated with platinum (H_2PtCl_6, Aldrich) to an impedance of ~300 kΩ. Cephalexin (15 mg/kg, p.o.) was administered twice daily for 2 weeks postoperatively to prevent infection.

Behavioral task. Recording was conducted in aluminum chambers ~18 inches on each side with downward sloping walls narrowing to an area of 12 × 12 inches at the bottom. A central odor port was located above two adjacent fluid wells. Directional lights were located next to fluid wells. House lights were located above the panel. The odor port was connected to an air flow dilution olfactometer to allow the rapid delivery of olfactory cues. Task control was implemented via computer. Port entry and licking were monitored by disruption of photobeams.

The basic design of a trial is illustrated in Figure 1. Rats were trained to perform a value-based light detection task. The rats first learned to associate directional lights with reward locations. After the rats accurately responded to the lights 60% of the time, they were introduced to odors that preceded the direction light and indicated the size of the reward to be delivered at the end of the trial. Once the rats were able to maintain >60% correct performance with all these manipulations across 150–200 trials, we trained them for an additional month before surgeries were performed. Thus, rats had extended training on this task before recordings started.

Figure 1 illustrates the sequence of events during a trial. Each trial began by illumination of house lights that instructed the rat to nose poke into the central odor port. Nose poking began a 500 ms pre-odor delay period. Then, one of two possible odors, which cued upcoming reward size, was delivered for 500 ms. Odor offset was followed by a 250–500 ms variable odor delay. At the end of this delay, directional lights were flashed for 100 ms. The trial was aborted if a rat exited the odor port at any time before offset of a directional cue light. Left and right lights signaled which direction to make the response. Rats had to remain in the well 500–1000 ms (prefluid delay) before reward delivery for both large and small rewards.

Odors signaled that a large or small amount of 10% sucrose solution would be available if the rat correctly responded to the direction lights. Odor meanings never changed throughout the course of the experiment. Odors were presented in a pseudorandom sequence such that big/small odors and left/right directional lights were presented in equal numbers (±1 over 250 trials). In addition, the same odor could be presented on no more than 3 consecutive trials. Thus, after three correct trials of the same type, rats could predict what the next odor was going to be. This rule was not imposed on response direction. On average, rats performed >200 correct trials per session during collection of neural data.

Single-unit recording. Procedures were the same as those described previously (Bryden et al., 2011)Wires were screened for activity daily; if no activity was detected, the rat was removed, and the electrode assembly was advanced 40 or 80 μm. Otherwise, active wires were selected to be recorded, a session was conducted, and the electrode was advanced at the end of the session. Neural activity was recorded using two identical Plexon Multichannel Acquisition Processor systems, interfaced with odor discrimination training chambers. Signals from the electrode wires were amplified 20× by an op-amp headstage (Plexon Inc, HST/8o50-G20-GR), located on the electrode array. Immediately outside the training chamber, the signals were passed through a differential preamplifier (Plexon Inc, PBX2/16sp-r-G50/16fp-G50), where the single-unit signals were amplified 50× and filtered at 150–9000 Hz. The single-unit signals were then sent to the Multichannel Acquisition Processor box, where they were further filtered at 250–8000 Hz, digitized at 40 kHz and amplified at 1–32×. Waveforms (>2.5:1 signal-to-noise) were extracted from active channels and recorded to disk by an associated workstation with event timestamps from the behavior computer. Waveforms were not inverted before data analysis.
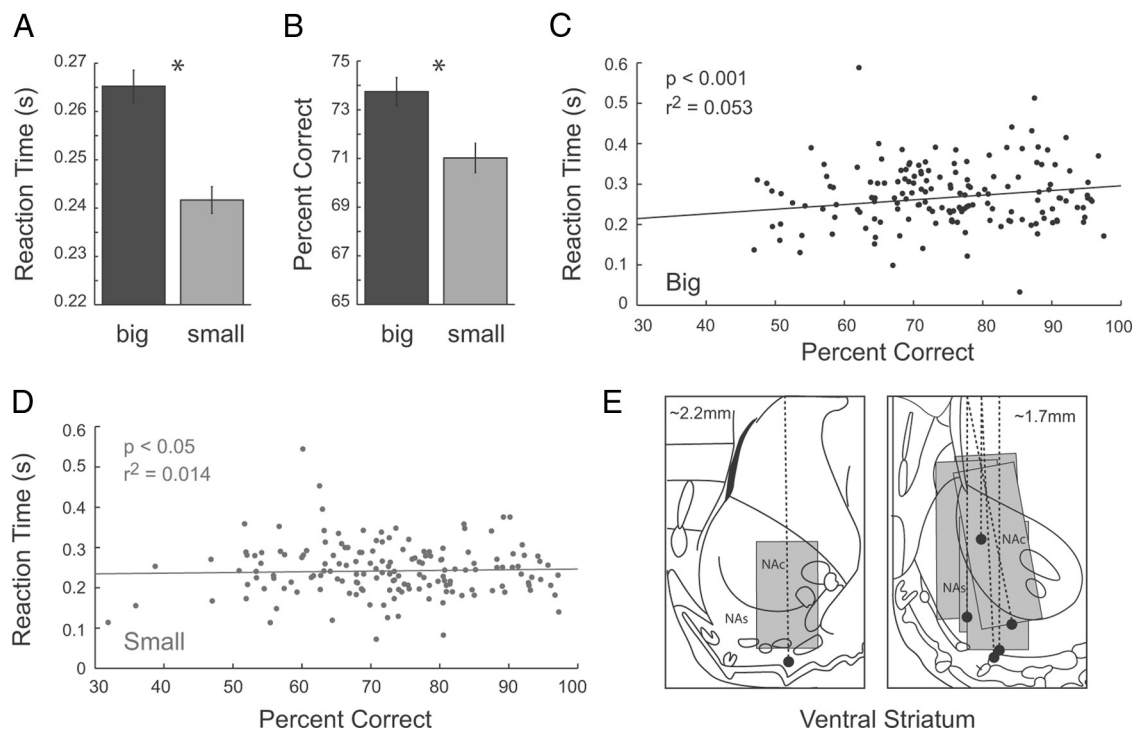
Figure 2. Rats were slower and more accurate on large reward trials. *A*, Latency to exit the odor port after directional cue lights had been extinguished. *B*, Percent correct scores as a function of all trials in which a choice was made to one of the fluid wells. *C*, *D*, Correlation between reaction time and percent correct scores for large and small reward trials, respectively. Asterisks, Planned comparisons revealing statistically significant differences (*t* test, *p* < 0.05). Error bars indicate SEM. *E*, Location of recording sites. Gray dots represent final electrode position. Gray box marks extent of recording sites. NAc, Nucleus accumbens core; NAs, nucleus accumbens, shell. *N* = 6 rats.

Data analysis. Units were sorted using Offline Sorter software from Plexon Inc, using a template matching algorithm. Sorted files were then processed in Neuroexplorer to extract unit timestamps and relevant event markers. These data were subsequently analyzed in Matlab (Math-Works). To examine activity related to odor sampling we examined activity 750 ms after odor onset (odor epoch). This activity precedes onset of direction light cues. We also examined activity 500 ms before odor presentation (pre-odor epoch) to quantify activity related to previous and predicted reward size before the reward predictive odor cue. Wilcoxon tests were used to measure significant shifts from zero in distribution plots ($p <$ 0.05). *t* tests or ANOVAs were used to measure within-cell differences in firing rate ($p < 0.05$). Pearson χ^2 tests ($p < 0.05$) were used to compare the proportions of neurons.

Results

Rats were trained on a task in which odor cues signaled the size of the expected reward (large or small). Subsequent directional cue lights then instructed the direction of the behavioral response necessary to obtain that reward. The sequence of events is illustrated in Figure 1*A*. House lights indicated the start of the trial. Rats began the trial by nose poking into the central odor port. After 500 ms one of two odors were presented for 500 ms. Odors signaled the size of the liquid sucrose reward to be delivered at the end of the trial; large (3 boli) or small (1 bolus). After a short post-odor variable delay, a light to the left or right of the odor port briefly flashed (100 ms), signaling which direction that the rat would have to respond to get reward. The rule was to simply detect the light and make a behavioral response in that direction. Rewards were delivered after a variable delay of 500–1000 ms. Essentially, there were a total of four trial-types: large-left, large-right, small-left, and small-right (Fig. 1*B*).

Rats were significantly slower and more accurate on large reward trials (Fig. 2*A*,*B*; *t* test; percent correct: $t_{(487)} = 9.08$, $p <$ 0.05; reaction time: $t_{(487)} = 11.8$, $p < 0.05$). Further, slower la-

tencies resulted in better task performance consistent with a speed accuracy trade off. This is illustrated in Figure 2, *C* and *D*, which plots reaction times (port exit minus light offset) versus accuracy for large and small reward trial types for each recording session. For both large and small reward trials, the slower the rat was, the better the performance. The correlation was weak but significant for both trials types (*p* values <0.05; large reward *r* = 0.23; small reward *r* = 0.12. Thus, in this reward task, high value reward was associated with slower not faster reaction times, which is atypical for studies that examine reward-related functions (Watanabe et al., 2001).

Activity in VS reflected the value independent of the instrumental response

We recorded 488 VS neurons in 6 rats during performance of the task. Recording locations are illustrated in Figure 2*E*. As has been reported previously (Carelli and Deadwyler, 1994; Nicola et al., 2004; Taha and Fields, 2006; Robinson and Carelli, 2008; Roesch et al., 2009), many VS neurons were excited (*n* = 229; 47%) during reward cue sampling (odor epoch = odor onset plus 750 ms) vs baseline (1 s before nose poke; *t* test comparing baseline to the odor epoch over all trials collapsed across direction; *p* < 0.05).

Activity of many of these neurons reflected the value of the predicted reward before directional cue lights. For example, the single neuron illustrated in Figure 3 fired more strongly during large reward trials compared with small reward trials after odor sampling and before the direction being cued. To quantify this effect we performed a *t* test on each of the cue-responsive cells during an epoch starting at odor onset and ending 750 ms later (Fig. 3; gray box in rasters). This time period preceded any knowledge of response direction. Of the 229 cue-responsive neurons, 33 cells fired more strongly for an expected big reward and 9 for

an expected small reward. The total number of significant cells ($n = 42$) exceeded the frequency expect by chance alone (type 1 error, 5%, $\chi^2 = 85.5$; $p < 0.0001$) and the counts of neurons that fired significantly more for larger reward were in the significant majority (33 vs 9; $\chi^2 = 13.6$; $p < 0.001$).

This effect is further illustrated in Figure 4, *A* and *B*, which plots the average activity across all cue-responsive neurons ($n = 229$). These plots were constructed by averaging over the mean firing rates obtained from each individual neuron. Curves were collapsed across each neuron's preferred direction and outcome. Preferred direction and outcome were designated according to the direction and outcome that elicited the highest firing during light illumination (100 ms) and odor sampling (odor onset to 750 ms after odor onset), respectively. In these plots, "preferred" refers to the direction and outcome that elicited the strongest neural response, not the outcome preferred by the rat. In the heat plot below, the average normalized firing for each neuron is illustrated by row for the four conditions (Fig. 4*B*). Clearly, activity was higher over

Figure 3. A single cell example. Lines represent the average firing rate over the course of the trial for all four conditions. Activity is aligned to odor onset. Each tick mark indicates one action potential. Gray bars represent the analysis epoch, which is referred to as the "odor epoch." The odor epoch encompasses activity (750 ms) after odor onset before any knowledge about response direction (i.e., cueing of directional light).

many VS neurons for one predicted reward over another during sampling of the odors before response instruction.

To further quantify these effects across the population we computed a size index for each neuron, defined by the difference between large and small reward divided by the sum of the two. Activity was taken during the odor epoch (odor onset plus 750; gray bar). The index was significantly shifted above zero, indicating higher firing rates when the reward cue predicted large reward (Fig. 4*C*; Wilcoxon; $p < 0.001$; $\mu = 0.045$).

Activity before odor onset reflects past and predicted reward
Also noticeable in the population histogram (Fig. 4*A*) is that activity at trial onset, just before odor presentation, appeared to reflect the predicted value of the upcoming trial. This was possible due to the pseudorandom nature of the task design. To ensure equal samples of each trial type within a given block of time, trial selection was randomized with the rule that if three of the same rewards were consecutively delivered the fourth would always be the opposite reward size. Thus, rats could predict a large reward trial after three smalls and a small reward trial after three large reward trials.

To test the hypothesis that activity in VS was representing the predicted value of the upcoming trial, we divided trials into conditions when the cell's preferred or nonpreferred outcome was predicted versus when it was not. This was done by examining large and small reward trials after 3 of the opposite type. The problem with this analysis is that any differences that might arise from this comparison might reflect what was delivered on the previous trial because predictable small and large reward trials were always preceded by a large and small reward, respectively. Thus, differences in firing when examining "predicted reward" effects might just reflect what the "previous reward" was. To rule this out, we also examined trials in which the previous trial was the same but there was no reward prediction. This was done by
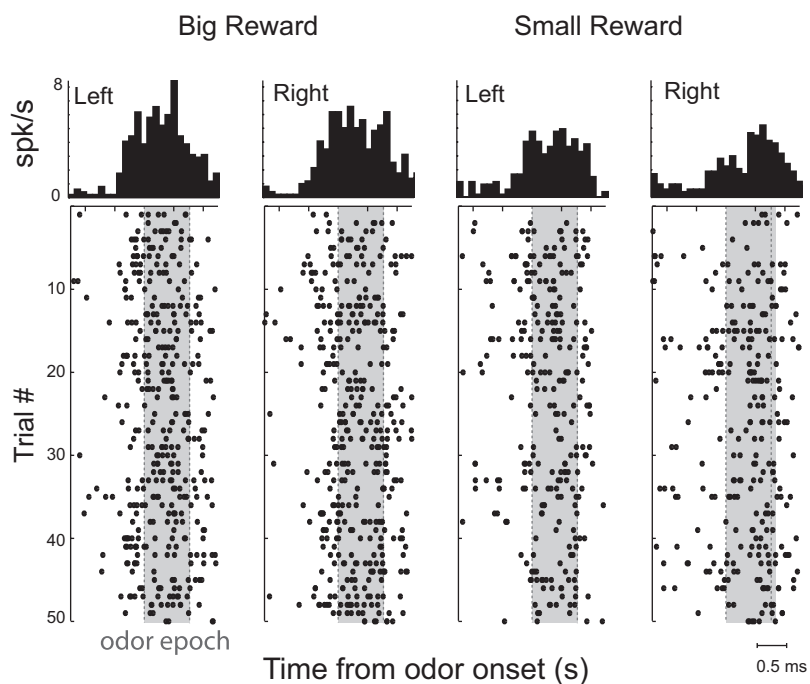
examining instances in which two of the same trial type occurred one right after the other. Since predictions were only possible after 3 trials of the same type, there was no way possible that the rats could guess what the current trial type was during these instances (50/50). We examined these trials to determine whether activity preceding the odor reflected the previous trial's reward.

The breakdown of these conditions is illustrated in the table in Figure 5. Black and gray lines indicate whether the reward on the current trial was the cell's preferred and nonpreferred reward, respectively (column 2; "current reward"). Thick and thin lines represent instances where the previous outcome was preferred and nonpreferred, respectively (column 3; previous reward). Finally, solid lines are trials in which no prediction was possible, and black and gray dashed lines represent trials when the reward was predicted to be preferred or nonpreferred, respectively (column 4; predicted reward). Note that as above, for population histograms, preferred and nonpreferred reflect the cell's not the rat's preference. Outcome preference was determined by firing during odor sampling (odor onset plus 750 ms), thus any differences that emerge before sampling cannot be due to how we defined preferred and nonpreferred outcomes.

The population histogram in Figure 5*A* illustrates that when the cell's preferred reward was predicted by the rat (after receipt of 3 nonpreferred rewards; thin black dashed) activity was high compared with when there was no prediction and the preceding reward was also nonpreferred (thin solid gray). This comparison is further illustrated in Figure 5*B* (left), which represents the same data, zoomed in, and isolated so that a better comparison can be made. These results indicate that, with previous reward held constant, activity was high when the predicted reward was preferred.

This effect is quantified in the right panel in Figure 5*B*, which plots the difference between predicted large reward trials and trials with no prediction divided by the sum of the two for activity during the 500 ms before odor onset (pre-odor epoch; gray bar in
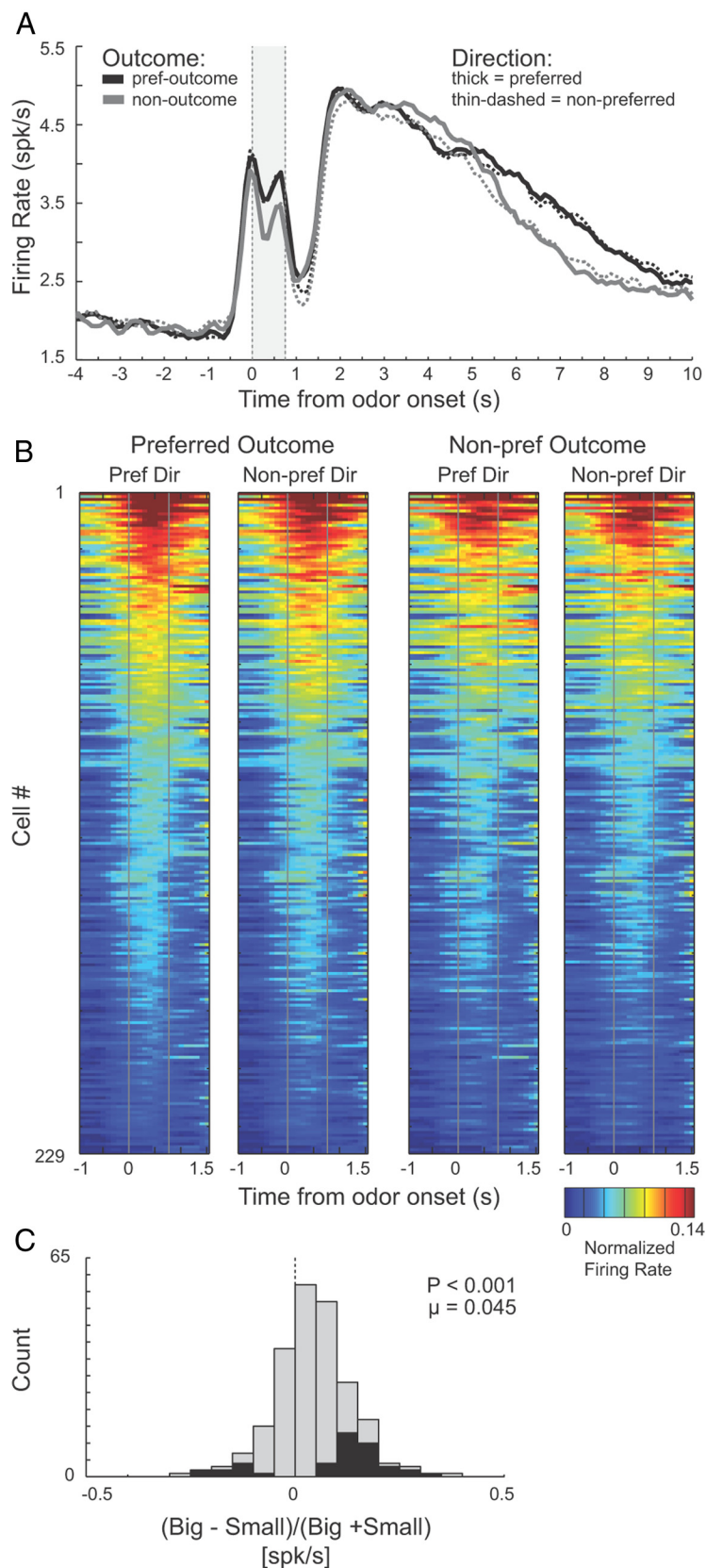
A

B

C

Fig. 5A). The distribution was significantly shifted in the positive direction (Wilcoxon; $p < 0.005$; $\mu = 0.039$) and the counts of neurons that fired significantly more strongly when a large reward was predicted (compared with when there was no prediction) were in the majority (18 vs 4; $\chi^2 = 8.79$; $p < 0.005$), demonstrating that activity was higher in VS when the larger reward trial was predicted.

Although these results are consistent with encoding of predicted value, activity was also high when there was no prediction, but the value of the preceding reward was preferred. This can be realized by examining activity on preferred and non-preferred trials in which the previous trial was of the same value (i.e., large followed by large or small followed by small; Fig. 5A,C; thick solid black vs thin solid gray). On these trials, there was a 50% chance that the current trial would be of the same value as the previous trial, thus they were unable to predict what the current trial might be. Activity was higher when the previous trial was preferred (thick solid black), even when no prediction was possible.

This effect is quantified in the right panel in Figure 5C, which plots the difference between firing during the pre-odor epoch when the previous reward was large versus small (divided by the sum of the two). Although the distribution was significantly shifted in the positive direction, the effect did not achieve significance (Wilcoxon; $p = 0.21$; $\mu = 0.015$), however the counts of neurons that fired significantly more strongly on trials following larger reward were in the significant majority (14 vs 1; $\chi^2 = 8.78$; $p < 0.005$).

Finally, we examined activity on trials in which the past reward was preferred but the value of the reward that was predicted on the next trial was nonpreferred (Fig. 5A,D; thick dashed gray). Again, these trials were compared with trials in which the previous reward was nonpreferred and the current trials was unpredictable (thin gray). In light of the other comparisons, firing under this condition could go either way. Activity might be low because the rats were predicting a nonpreferred trial (Fig. 5B), but activity might be

←

Figure 4. VS activity reflected value independent of direction. ***A***, Population firing based on preferred/nonpreferred reward outcomes and direction ($n = 229$). In this plot, for each neuron, direction and outcome were referenced to the max response before averaging. Black lines correspond to the preferred outcome; gray corresponds to the nonpreferred outcome. Thick lines represent the preferred direction; thin-dashed lines represent the nonpreferred direction. Zero is the time of odor onset. Directional light cues were presented 750–1000 ms after odor onset. Gray bar indicates odor analysis epoch. ***B***, Averaged normalized activity for each odor-responsive neuron. Hotter colors indicate higher firing. Each row represents activity of one neuron, which were sorted by

firing in the preferred outcome/direction. ***C***, Distribution of size indices determined by subtracting small reward from large reward trials and dividing by the sum of the two for activity during the odor epoch (gray bar in ***A***). Black bars represent the number of neurons that showed a significant difference between large and small reward trial types during the odor epoch (*t* test; $p < 0.05$).

high because the previous trial was preferred (Fig. 5C). We found that activity was high, reflecting the value of the reward on the pervious trial. Interestingly, after odor onset, activity quickly rectified itself reflecting the knowledge obtained by sampling the odor that predicted the non-preferred reward (Fig. 5A, inset, pre-odor vs odor epoch; Fig. 5D, left).

As above, this effect were quantified in the right panel of Figure 5D, which plots activity differences between trials in which the previous reward was large versus small. The distribution was significantly shifted in the positive direction (Wilcoxon; $p < 0.001$; $\mu = 0.060$) and the counts of neurons that fired significantly more strongly when the previous reward was large compared with when it was small were in the majority (21 vs 1; $\chi^2 = 18$; $p < 0.0001$). Together, these results suggest that activity was high whenever the past or predicted reward was of high value.

To determine whether past and predicted effects were correlated we plotted each of the two previous reward distributions (Fig. 5C,D) against the distribution quantifying predicted reward effects (Fig. 5B). That is, we asked whether effects related to past reward tended to occur in the same neurons that fired more strongly when the predicted reward was large. Both were significantly positively correlated indicating that activity in VS does not just predict expected reward but is also modulated by past reward delivery and that these effects tend to occur in the same neurons (Fig. 5E,F; p values <0.0001; $r > 0.40$).

Activity in VS was positively correlated with reaction time and accuracy

Activity in VS was high when reward value was high (Fig. 4). High value reward was associated with slower reaction times and slower reaction times were associated with better task performance (Fig. 2). This suggests that VS was involved in slowing down behavior so that fewer mistakes were made on large reward trials. If true, then one might expect that activity in VS would be positively correlated with both reaction time and performance.

To examine this issue we plotted reaction time and percent correct scores versus firing rate during the odor epoch for each VS neuron independently for large and small reward conditions (Fig. 6). The correlation with percent correct scores was significant and positive under both reward magnitudes (p values <0.01; $r > 0.12$). Thus, higher firing rate was associated with more accurate performance.

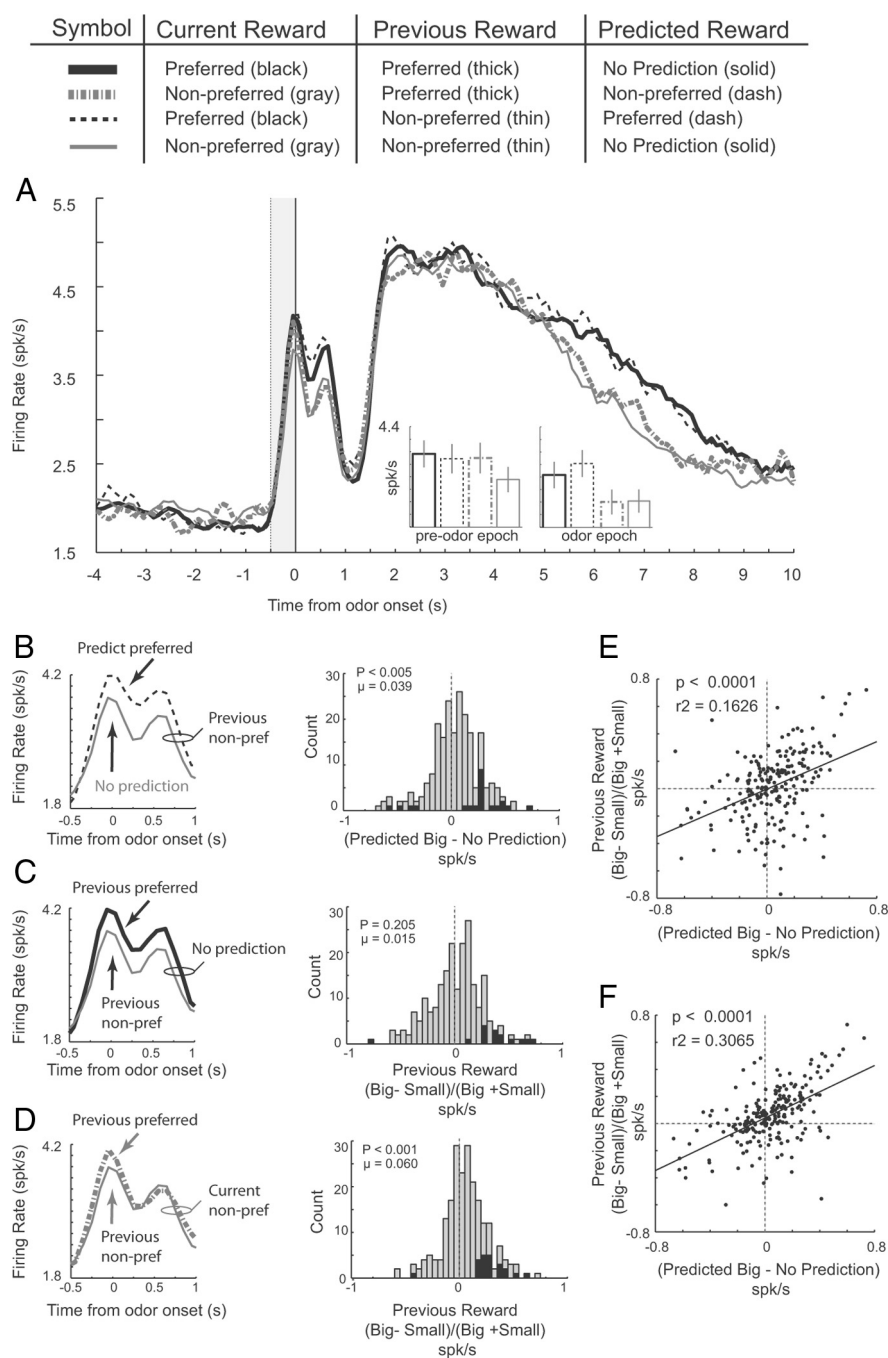| Symbol	Current Reward	Previous Reward	Predicted Reward
▬	Preferred (black)	Preferred (thick)	No Prediction (solid)
▮▮▮	Non-preferred (gray)	Preferred (thick)	Non-preferred (dash)
┄┄	Preferred (black)	Non-preferred (thin)	Preferred (dash)
──	Non-preferred (gray)	Non-preferred (thin)	No Prediction (solid)

Figure 5. Activity in VS represented past and predicted reward. Population firing under four different sequences of trials averaged across direction (n = 229). Table illustrates four trial types. Thick = previous outcome preferred; Thin = previous outcome nonpreferred; Black = current trial preferred; Gray = current trial nonpreferred; Solid = no prediction; Dashed = predictable trial (Single-dashed = predicted preferred; Multi-dashed = predicted nonpreferred). Insets represent average activity during the pre-odor (500 ms before odor) and the odor epoch (odor onset plus 750 ms) across these 4 conditions. Error bars, SEM. **B–D**, Left, Blown up versions of lines in A. Analyses comparing each pair are presented to right of the line histograms. **B**, Distribution reflecting the difference between pre-odor cue firing (500 ms; gray bar) when the predicted reward was to be large compared with when the reward was not predictable (predicted big minus no prediction; thin black dashed minus thin solid gray). Thus, positive values indicate higher firing when the predicted reward was large. Reward that preceded trials used in this analysis were small reward trials. **C**, Distribution reflecting the difference between precue firing when the preceding reward was large compared with when the preceding reward was small (big minus small divided by big + small; Thick solid black minus thin solid gray). Thus, positive values indicate higher firing when the previous reward was large. The value of the reward on the current trial for this analysis could not be predicted (50/50). **D**, Distribution reflecting the difference between precue firing when the preceding reward was large and the predicted reward was small compared with when the preceding reward was small and there was no prediction (Thick gray dashed minus thin solid gray). Black bars represent the number of neurons that showed a significant difference between respective comparisons (t test; p < 0.05). **E**, Correlation analysis between distributions in **B** and **C**. **F**, Correlation analysis between distributions in **B** and **D**.
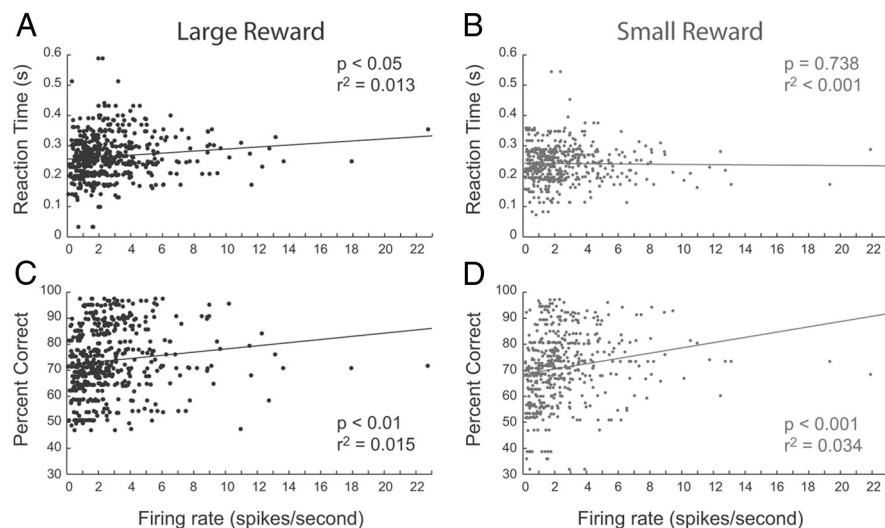
Figure 6. VS activity was positively correlated with accuracy and reaction time. *A*, *B*, Correlation between firing rate (odor epoch) and reaction time (light offset to odor port exit) for large and small reward trials during all recording sessions. *C*, *D*, Correlation between firing rate and percent correct scores for large and small reward trials during all recording sessions.

The correlation with reaction time was significant under big-reward conditions, demonstrating that increased activity was correlated with slower reaction times at least when more was at stake ($p < 0.04$; $r = 0.11$).

We also determined how many single neurons exhibited a significant trial by trial correlation between reaction time and firing rate. Again, this analysis was conducted independently for big and small reward trials to avoid any confound related to slower and faster responding on these trial types. As expected from the population analysis, significantly more VS neurons exhibited a positive correlation ($n = 38$) between firing rate and reaction time as opposed to a negative correlation ($n = 21$; $\chi^2 = 4.84$; $p < 0.05$).

Discussion

Here we show that single neurons in VS signal information regarding predicted value independent of response direction and speed of movement initiation. Cues predicting high value outcomes had a profound impact on behavior, increasing reaction time and accuracy. Slower reaction times and better performance were correlated with activity during cue-sampling at the population and single-cell level. The finding that activity in VS was high when the better reward was predicted is broadly consistent with other studies (Carelli and Deadwyler, 1994; Bowman et al., 1996; Shidara et al., 1998; Carelli, 2002; Setlow et al., 2003; Janak et al., 2004; Nicola et al., 2004; Shidara and Richmond, 2004; Taha and Fields, 2006; German and Fields, 2007; Hollander and Carelli, 2007; Y. B. Kim et al., 2007; Simmons et al., 2007; Takahashi et al., 2007; Robinson and Carelli, 2008; Ito and Doya, 2009; Kimchi and Laubach, 2009; van der Meer and Redish, 2009; van der Meer et al., 2010; Day et al., 2011). However, this is the first demonstration that single neurons in VS encode value in a task in which direction and predictive value cues were temporally separated. Additionally, this is the first experiment, that we are aware of, that examines value encoding in VS when high value reward is associated with slower, not faster latencies to respond. These results suggest that the role of VS is not to simply energize decisions toward valued goals, but instead, to signal value independent of motor contingencies, possibility in the

service of good-based decision-making and reinforcement learning as we will discuss below.

VS encodes value independent of motor contingencies

VS has long been thought to be a limbic–motor interface (Mogenson et al., 1980), a hypothesis that was originally derived from VS's connectivity with decision/motor-related areas including the prefrontal cortex, limbic-related areas including the hippocampus, amygdala, orbitofrontal cortex and midbrain dopamine neurons, along with its outputs to motor regions, such as ventral pallidum (Groenewegen and Russchen, 1984; Heimer et al., 1991; Brog et al., 1993; Wright and Groenewegen, 1995; Voorn et al., 2004; Gruber and O'Donnell, 2009). Through these connections, the ventral striatum is thought to integrate information about the value of expected outcomes with specific motor information to guide behavior. Consistent with this proposal, lesions of VS impact behavioral measures of motivation, vigor, salience and arousal, which are thought to reflect the value of reward expected (Wadenberg et al., 1990; Berridge and Robinson, 1998; Blokland, 1998; Ikemoto and Panksepp, 1999; Di Ciano et al., 2001; Cardinal et al., 2002a,b; Di Chiara, 2002; Salamone and Correa, 2002; Giertler et al., 2003; Wakabayashi et al., 2004; Yun et al., 2004; Floresco et al., 2008; Gruber et al., 2009; Ghods-Sharifi and Floresco, 2010; Stopper and Floresco, 2011). From these studies it has been suggested that VS is indeed critical for motivating behavior. However, there has been little direct single-unit recording data from VS in tasks designed to directly address this question and most studies have not varied both expected reward value and response direction (Hassani et al., 2001; Cromwell and Schultz, 2003).

We addressed this issue in a previous paper by recording from single neurons in VS while rats performed a choice task for two types of differently valued rewards (size and delay) (Roesch et al., 2009). On every trial, rats were instructed to choose between two wells (left or right) to receive reward. In different trial blocks, we manipulated the value of the expected reward associated with left and right movements. In that report we showed that cue-evoked activity in VS integrated the value of the expected reward and the direction of the upcoming movement, simultaneously. Furthermore, increases in firing rate were correlated with faster reaction times.

These results were entirely consistent with the notion that VS serves to integrate information about the value of an expected reward with motor output during decision-making, but as in so many studies before us, rewards were directly tied to the direction and latency of instrumental response. Further, value and response direction were cued together at the time when the animal was to make the choice. Thus, it was unclear whether activity was related to value encoding or just reflected enhanced motor output. It was also unclear whether or not VS could represent expected value when the instrumental response was unknown. Here, we clearly show that activity in VS can signal value of the expected reward before the direction is cued even when responding in that direction becomes slower as value increases. These data demonstrate that the sole purpose of predictive reward sig-

nals in VS is not just to energize specific actions but to signal value in a way that might be used to slow reaction times to improve task performance when more is at stake. More importantly, these results indicate that VS can encode expected value independent of motor contingencies.

The role of VS in actor-critic models

Many aspects of these data are consistent with theories suggesting that VS plays a critical role in actor-critic models, optimizing long term action selection through its connections with midbrain dopamine neurons (Barto, 1995; Houk et al., 1995; Sutton and Barto, 1998; Joel et al., 2002; Redish, 2004; Niv and Schoenbaum, 2008; Takahashi et al., 2008; van der Meer and Redish, 2011). In this model the Critic stores and learns values of states, which in turn are used to compute prediction errors necessary for learning and adaptive behavior. The Actor stores and forms a policy on which actions should be selected (Joel et al., 2002; Montague et al., 2004). The functions of Critic and Actor have been attributed to ventral and dorsal lateral striatum, respectively (Everitt et al., 1991; Cardinal et al., 2002a; O'Doherty et al., 2004; Voorn et al., 2004; Balleine, 2005; Pessiglione et al., 2006). Although encoding of predicted value independent of motor contingences is consistent with VS's role as the Critic in this model, the fact that activity in VS represented past not just the predicted reward is not entirely consistent.

The combination of past and predicted information at the start of behavioral trials is more in line with the rats' evaluation of the current state based on what was and what is to be (van der Meer and Redish, 2011). This is consistent with previous work suggesting VS inactivation results in the inability to incorporate past reward history with current behavior (Stopper and Floresco, 2011) and that activity in VS takes into account previous choices (Y. B. Kim et al., 2007; Ito and Doya, 2009; H. Kim et al., 2009). We suggest that activity in VS reflects the value the animal places on the current situation, which would reflect both past and predicted reward. That is, value is high when a good reward was just delivered and/or was predicted on the next trial. Regardless of what variables might alter this signal, these data clearly demonstrate that outcome-related activity in VS is not just predictive in nature.

The role of VS in good-based models of choice

Together, these findings suggest that the VS may play an important role in representing abstract value as described in good-based models of economic choice (Padoa-Schioppa, 2011). The good-based model suggests that the brain maintains abstract representation of a good's value and then makes choices by comparing the value of different goods. It has been proposed that two criteria must be satisfied for a region to possess an abstract representation of value. First, the encoding in this region should be domain general, meaning that the activity should incorporate all relevant determinants of a good's value (i.e., quantity, risk, cost). Second, the encoding should be independent of sensorimotor contingencies of choice. Single-unit studies in primates have suggested that activity in orbital frontal cortex (OFC) fits these criteria, but, unfortunately, few other areas have actually been tested in the same manner (Tremblay and Schultz, 1999; Roesch and Olson, 2004, 2005, 2007; Padoa-Schioppa, 2007, 2009, 2011; Wallis, 2007; Kennerley and Wallis, 2009a,b; Kobayashi et al., 2010; Wallis and Kennerley, 2010).

We suggest that VS serves the same function as OFC in this model (Padoa-Schioppa, 2011). We have previously shown that activity in VS is domain general, encoding reward size and delay

to reward. VS neurons fire more strongly when a rat expects a large reward compared with small reward and a short delay compared with a long delay, both of which were preferred by rats (Roesch et al., 2009). It has also been shown that VS encodes how much effort is required to obtain reward (Day et al., 2011). Last, the current dataset demonstrates that representations of reward in VS are influenced by past reward delivery. Thus, activity in VS fulfills the first criteria, incorporating relevant determinants of a good's value into its signal.

Previous work has also demonstrated that activity in VS encodes value independent from sensory cues that predict rewards and instruct responses (Cromwell and Schultz, 2003; Cromwell et al., 2005). For example, we have shown that activity in VS does not differ between two different odors that predict the same reward (Roesch et al., 2009). Here, we demonstrate that activity in VS reflects value independent of response direction and the latency of the response, demonstrating that activity in VS represents value independent of motor contingencies consistent with the second criteria described above. Thus, we conclude that activity in VS, like responses observed in primate OFC, fit the criteria of representing abstract value in the service of the good-based model of economic decision making.

References

Balleine BW (2005) Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. Physiol Behav 86:717–730.

Barto A (1995) Adaptive critics and the basal ganglia. Available at: http://works.bepress.com/andrew_barto/9.

Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res Brain Res Rev 28:309–369.

Blokland A (1998) Reaction time responding in rats. Neurosci Biobehav Rev 22:847–864.

Bowman EM, Aigner TG, Richmond BJ (1996) Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. J Neurophysiol 75:1061–1073.

Brog JS, Salyapongse A, Deutch AY, Zahm DS (1993) The patterns of afferent innervation of the core and shell in the "accumbens" part of the rat ventral striatum: immunohistochemical detection of retrogradely transported fluoro-gold. J Comp Neurol 338:255–278.

Bryden DW, Johnson EE, Diao X, Roesch MR (2011) Impact of expected value on neural activity in rat substantia nigra pars reticulata. Eur J Neurosci 33:2308–2317.

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002a) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321–352.

Cardinal RN, Parkinson JA, Lachenal G, Halkerston KM, Rudarakanchana N, Hall J, Morrison CH, Howes SR, Robbins TW, Everitt BJ (2002b) Effects of selective excitotoxic lesions of the nucleus accumbens core, anterior cingulate cortex, and central nucleus of the amygdala on autoshaping performance in rats. Behav Neurosci 116:553–567.

Carelli RM (2002) Nucleus accumbens cell firing during goal-directed behaviors for cocaine vs 'natural' reinforcement. Physiol Behav 76:379–387.

Carelli RM, Deadwyler SA (1994) A comparison of nucleus accumbens neuronal firing patterns during cocaine self-administration and water reinforcement in rats. J Neurosci 14:7735–7746.

Cromwell HC, Schultz W (2003) Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. J Neurophysiol 89:2823–2838.

Cromwell HC, Hassani OK, Schultz W (2005) Relative reward processing in primate striatum. Exp Brain Res 162:520–525.

Day JJ, Jones JL, Carelli RM (2011) Nucleus accumbens neurons encode predicted and ongoing reward costs in rats. Eur J Neurosci 33:308–321.

Di Chiara G (2002) Nucleus accumbens shell and core dopamine: differential role in behavior and addiction. Behav Brain Res 137:75–114.

Di Ciano P, Cardinal RN, Cowell RA, Little SJ, Everitt BJ (2001) Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. J Neurosci 21:9471–9477.

Everitt BJ, Morris KA, O'Brien A, Robbins TW (1991) The basolateral amygdala-ventral striatal system and conditioned place preference: further evidence of limbic-striatal interactions underlying reward-related processes. Neuroscience 42:1–18.

Floresco SB, St Onge JR, Ghods-Sharifi S, Winstanley CA (2008) Corticolimbic-striatal circuits subserving different forms of cost-benefit decision making. Cogn Affect Behav Neurosci 8:375–389.

German PW, Fields HL (2007) Rat nucleus accumbens neurons persistently encode locations associated with morphine reward. J Neurophysiol 97:2094–2106.

Ghods-Sharifi S, Floresco SB (2010) Differential effects on effort discounting induced by inactivations of the nucleus accumbens core or shell. Behav Neurosci 124:179–191.

Giertler C, Bohn I, Hauber W (2003) The rat nucleus accumbens is involved in guiding of instrumental responses by stimuli predicting reward magnitude. Eur J Neurosci 18:1993–1996.

Groenewegen HJ, Russchen FT (1984) Organization of the efferent projections of the nucleus accumbens to pallidal, hypothalamic, and mesencephalic structures: a tracing and immunohistochemical study in the cat. J Comp Neurol 223:347–367.

Gruber AJ, O'Donnell P (2009) Bursting activation of prefrontal cortex drives sustained up states in nucleus accumbens spiny neurons in vivo. Synapse 63:173–180.

Gruber AJ, Hussain RJ, O'Donnell P (2009) The nucleus accumbens: a switchboard for goal-directed behaviors. PLoS One 4:e5062.

Hassani OK, Cromwell HC, Schultz W (2001) Influence of expectation of different rewards on behavior-related neuronal activity in the striatum. J Neurophysiol 85:2477–2489.

Hauber W, Bohn I, Giertler C (2000) NMDA, but not dopamine D(2), receptors in the rat nucleus accumbens are involved in guidance of instrumental behavior by stimuli predicting reward magnitude. J Neurosci 20:6282–6288.

Heimer L, Zahm DS, Churchill L, Kalivas PW, Wohltmann C (1991) Specificity in the projection patterns of accumbal core and shell in the rat. Neuroscience 41:89–125.

Hollander JA, Carelli RM (2007) Cocaine-associated stimuli increase cocaine seeking and activate accumbens core neurons after abstinence. J Neurosci 27:3535–3539.

Houk J, Adams JL, Barto AG (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Models of information processing in the basal ganglia (Houk J, Davis JL, Beiser DG, eds), pp 249–270. Cambridge, MA: MIT.

Ikemoto S, Panksepp J (1999) The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. Brain Res Brain Res Rev 31:6–41.

Ito M, Doya K (2009) Validation of decision-making models and analysis of decision variables in the rat basal ganglia. J Neurosci 29:9861–9874.

Janak PH, Chen MT, Caulder T (2004) Dynamics of neural coding in the accumbens during extinction and reinstatement of rewarded behavior. Behav Brain Res 154:125–135.

Joel D, Niv Y, Ruppin E (2002) Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw 15:535–547.

Kennerley SW, Wallis JD (2009a) Encoding of reward and space during a working memory task in the orbitofrontal cortex and anterior cingulate sulcus. J Neurophysiol 102:3352–3364.

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

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

Kim YB, Huh N, Lee H, Baeg EH, Lee D, Jung MW (2007) Encoding of action history in the rat ventral striatum. J Neurophysiol 98:3548–3556.

Kimchi EY, Laubach M (2009) Dynamic encoding of action selection by the medial striatum. J Neurosci 29:3148–3159.

Kobayashi S, Pinto de Carvalho O, Schultz W (2010) Adaptation of reward sensitivity in orbitofrontal neurons. J Neurosci 30:534–544.

Minamimoto T, La Camera G, Richmond BJ (2009) Measuring and modeling the interaction among reward size, delay to reward, and satiation level on motivation in monkeys. J Neurophysiol 101:437–447.

Mogenson GJ, Jones DL, Yim CY (1980) From motivation to action: functional interface between the limbic system and the motor system. Prog Neurobiol 14:69–97.

Montague PR, Hyman SE, Cohen JD (2004) Computational roles for dopamine in behavioural control. Nature 431:760–767.

Nicola SM, Yun IA, Wakabayashi KT, Fields HL (2004) Cue-evoked firing of nucleus accumbens neurons encodes motivational significance during a discriminative stimulus task. J Neurophysiol 91:1840–1865.

Niv Y, Schoenbaum G (2008) Dialogues on prediction errors. Trends Cogn Sci 12:265–272.

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

Padoa-Schioppa C (2007) Orbitofrontal cortex and the computation of economic value. Ann N Y Acad Sci 1121:232–253.

Padoa-Schioppa C (2009) Range-adapting representation of economic value in the orbitofrontal cortex. J Neurosci 29:14004–14014.

Padoa-Schioppa C (2011) Neurobiology of economic choice: a good-based model. Annu Rev Neurosci 34:333–359.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442:1042–1045.

Redish AD (2004) Addiction as a computational process gone awry. Science 306:1944–1947.

Robinson DL, Carelli RM (2008) Distinct subsets of nucleus accumbens neurons encode operant responding for ethanol versus water. Eur J Neurosci 28:1887–1894.

Roesch MR, Olson CR (2004) Neuronal activity related to reward value and motivation in primate frontal cortex. Science 304:307–310.

Roesch MR, Olson CR (2005) Neuronal activity in primate orbitofrontal cortex reflects the value of time. J Neurophysiol 94:2457–2471.

Roesch MR, Olson CR (2007) Neuronal activity related to anticipated reward in frontal cortex: does it represent value or reflect motivation? Ann N Y Acad Sci 1121:431–446.

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

Salamone JD, Correa M (2002) Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. Behav Brain Res 137:3–25.

Schultz W, Apicella P, Scarnati E, Ljungberg T (1992) Neuronal activity in monkey ventral striatum related to the expectation of reward. J Neurosci 12:4595–4610.

Setlow B, Schoenbaum G, Gallagher M (2003) Neural encoding in ventral striatum during olfactory discrimination learning. Neuron 38:625–636.

Shidara M, Richmond BJ (2004) Differential encoding of information about progress through multi-trial reward schedules by three groups of ventral striatal neurons. Neurosci Res 49:307–314.

Shidara M, Aigner TG, Richmond BJ (1998) Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. J Neurosci 18:2613–2625.

Simmons JM, Ravel S, Shidara M, Richmond BJ (2007) A comparison of reward-contingent neuronal activity in monkey orbitofrontal cortex and ventral striatum: guiding actions toward rewards. Ann N Y Acad Sci 1121:376–394.

Stopper CM, Floresco SB (2011) Contributions of the nucleus accumbens and its subregions to different aspects of risk-based decision making. Cogn Affect Behav Neurosci 11:97–112.

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

Taha SA, Fields HL (2006) Inhibitions of nucleus accumbens neurons encode a gating signal for reward-directed behavior. J Neurosci 26:217–222.

Takahashi Y, Schoenbaum G, Niv Y (2008) Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. Front Neurosci 2:86–99.

Takahashi Y, Roesch MR, Stalnaker TA, Schoenbaum G (2007) Cocaine exposure shifts the balance of associative encoding from ventral to dorsolateral striatum. Front Integr Neurosci 1:11.

Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. Nature 398:704–708.

van der Meer MA, Redish AD (2009) Covert expectation-of-reward in rat ventral striatum at decision points. Front Integr Neurosci 3:1.

van der Meer MA, Redish AD (2011) Ventral striatum: a critical look at models of learning and evaluation. Curr Opin Neurobiol 21:387–392.

van der Meer MA, Johnson A, Schmitzer-Torbert NC, Redish AD (2010) Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. Neuron 67:25–32.

Voorn P, Vanderschuren LJ, Groenewegen HJ, Robbins TW, Pennartz CM (2004) Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci 27:468–474.

Wadenberg ML, Ericson E, Magnusson O, Ahlenius S (1990) Suppression of conditioned avoidance behavior by the local application of (−)sulpiride into the ventral, but not the dorsal, striatum of the rat. Biol Psychiatry 28:297–307.

Wakabayashi KT, Fields HL, Nicola SM (2004) Dissociation of the role of nucleus accumbens dopamine in responding to reward-predictive cues and waiting for reward. Behav Brain Res 154:19–30.

Wallis JD (2007) Orbitofrontal cortex and its contribution to decision-making. Annu Rev Neurosci 30:31–56.

Wallis JD, Kennerley SW (2010) Heterogeneous reward signals in prefrontal cortex. Curr Opin Neurobiol 20:191–198.

Watanabe M, Cromwell HC, Tremblay L, Hollerman JR, Hikosaka K, Schultz W (2001) Behavioral reactions reflecting differential reward expectations in monkeys. Exp Brain Res 140:511–518.

Wright CI, Groenewegen HJ (1995) Patterns of convergence and segregation in the medial nucleus accumbens of the rat: relationships of prefrontal cortical, midline thalamic, and basal amygdaloid afferents. J Comp Neurol 361:383–403.

Yun IA, Wakabayashi KT, Fields HL, Nicola SM (2004) The ventral tegmental area is required for the behavioral and nucleus accumbens neuronal firing responses to incentive cues. J Neurosci 24:2923–2933.