Behavioral/Systems/Cognitive

Ventral Striatum and Orbitofrontal Cortex Are Both Required for Model-Based, But Not Model-Free, Reinforcement Learning

Michael A. McDannald,[1] Federica Lucantonio,[2] Kathryn A. Burke,[2] Yael Niv,[4] and Geoffrey Schoenbaum[1,2,3]

[1]Department of Anatomy and Neurobiology, [2]Program in Neuroscience, and [3]Department of Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland 21201, and [4]Neuroscience Institute and Department of Psychology, Princeton University, Princeton, New Jersey 08540

In many cases, learning is thought to be driven by differences between the value of rewards we expect and rewards we actually receive. Yet learning can also occur when the identity of the reward we receive is not as expected, even if its value remains unchanged. Learning from changes in reward identity implies access to an internal model of the environment, from which information about the identity of the expected reward can be derived. As a result, such learning is not easily accounted for by model-free reinforcement learning theories such as temporal difference reinforcement learning (TDRL), which predicate learning on changes in reward value, but not identity. Here, we used unblocking procedures to assess learning driven by value- versus identity-based prediction errors. Rats were trained to associate distinct visual cues with different food quantities and identities. These cues were subsequently presented in compound with novel auditory cues and the reward quantity or identity was selectively changed. Unblocking was assessed by presenting the auditory cues alone in a probe test. Consistent with neural implementations of TDRL models, we found that the ventral striatum was necessary for learning in response to changes in reward value. However, this area, along with orbitofrontal cortex, was also required for learning driven by changes in reward identity. This observation requires that existing models of TDRL in the ventral striatum be modified to include information about the specific features of expected outcomes derived from model-based representations, and that the role of orbitofrontal cortex in these models be clearly delineated.

Introduction

For Americans, ordering fish and chips for the first time can be surprising: expecting to receive potato chips, we instead receive fries. Through the surprising receipt of fries, we quickly learn the meaning of "chips" when in an English pub. This is illustrative of a basic principle: learning occurs when there is a difference between what we expect and what we receive. Yet formal, albeit narrow, model-free accounts of reinforcement learning assume that learning is driven only by changes in reward value. Such learning is not affected by surprises due to changes in reward identity, assuming the reward value is as expected, because these accounts do not have access to information about the identity of the reward predicted by the cue. In reality, we do learn when the identity of a reward changes independent of value; even if we equally like potato chips and fries, we will still learn and update our predictions if we receive one when expecting the other.

Different brain circuits have been hypothesized to signal information about the general value of a cue versus the identity of reward it predicts (Cardinal et al., 2002). It has been proposed that the ventral striatum (VS) primarily signals information about value, contributing this information to the computation of value-based prediction errors in dopaminergic neurons (Montague et al., 1996; Joel et al., 2002) and functioning as a critic within a reinforcement-learning actor/critic learning system (Barto, 1994; O'Doherty et al., 2004). By contrast, the orbitofrontal cortex (OFC) has been shown to be involved in signaling more specific information about the identity of expected outcomes, and thus might contribute more selectively to identity-based learning (Schoenbaum et al., 2009).

Temporal-difference reinforcement learning models (and their counterpart in the critic in actor/critic models) have been developed to explain Pavlovian learning and responding, by assuming that conditioned responses are driven by learned values (Dayan et al., 2006). Here, we took advantage of the phenomenon of Pavlovian blocking to test the above hypothesis regarding the roles of VS and OFC in learning directly. In blocking (Kamin, 1969), a rat is trained that presentation of a cue predicts a food outcome. After this association is learned, a second cue is presented in compound with the first cue, followed by the same food. In this arrangement, learning to the second cue is blocked, as the reward is completely predicted by the first cue. In unblocking, the compound of the first and second cue is followed by a larger quantity or different identity of food than was predicted by the first cue—violating the original expectation. In contrast to blocking, changing the quantity (Holland, 1984) or identity (Rescorla,

Received Oct. 20, 2010; revised Dec. 10, 2010; accepted Dec. 15, 2010.

This work was supported by grants from the National Institute on Drug Abuse (to G.S. and K.A.B.) and National Institute of Mental Health (to M.A.M.).

The authors declare no conflict of interest.

Correspondence should be addressed to either Michael A. McDannald or Geoffrey Schoenbaum, Department of Anatomy and Neurobiology, University of Maryland, Baltimore, 20 Penn St, HSF-2 S251, Baltimore, MD 21201. E-mail: mmcda001@umaryland.edu or schoenbg@schoenbaumlab.org.

DOI:10.1523/JNEUROSCI.5499-10.2011

 0270-6474/11/312700-06$15.00/0

Table 1. Experimental outline

Lesion assignment	Cue conditioning	Identity unblocking	Identity probe	Value unblocking	Value probe
Control	A → ●●●	AX → ●●●	A, X	AX→●●●	A, X
OFC	B → ○○○	BY → ●●●	B, Y	CZ→●●●	C, Z
VS	C → ●				

Rats were first randomly assigned to the control, OFC lesion, or VS lesion condition. Following recovery, rats received cue conditioning in which three visual cues (A, B, and C) predicted different identities (grape- or banana-flavored, represented by solid and empty circles, respectively) and quantities of reward (one or three pellets, represented by the number of circles). In both identity and value unblocking, the visual cues were compounded with novel auditory cues and the quantity and identity of reward fixed. The compound AX predicted the same reward as cue A. BY predicted the similarly valued but differently flavored reward as B. CZ predicted the similarly flavored but differently valued reward as C. Responding to each cue alone (A, B, C, X, Y, and Z) was assessed in an extinction probe test.

1999) of food results in substantial learning to the added second cue. By using these procedures, we were able to manipulate learning about value and the identity separately, and to independently assess the circuits involved in learning in response to value-based errors, which can be supported by model-free reinforcement learning, versus identity-based prediction errors, which require model-based representations (Daw et al., 2005).

Materials and Methods

Subjects. Eighty male Long–Evans rats (Charles River Laboratories) weighing between 275 and 300 g on arrival, were housed individually and placed on a 12 h light/dark schedule. All rats were given *ad libitum* access to food except during testing periods. During behavioral testing, rats were food deprived to 85% of their baseline weight. All testing was conducted during the light period of their cycle. All testing was performed in accordance with the guidelines set forth by the University of Maryland School of Medicine Animal Care and Use Committee and the National Institutes of Health.

Surgical procedures. Orbitofrontal cortex lesions were made in stereotaxic surgery using intracerebral infusions of NMDA (12.5 μg/μl; Sigma) in saline vehicle. Infusion volumes and locations were as follows: 0.05 μl: AP, 3.0 mm; ML, ±3.2 mm; DV, 5.2 mm; 0.1 μl: AP, 3.0 mm; ML, ±4.2 mm; DV, 5.2 mm; 0.1 μl: AP, 4.0 mm; ML, ±2.2 mm; DV, 3.8 mm; and 0.1 μl: AP, 4.0 mm; ML, ±3.7 mm; DV, 3.8 mm. Surgical OFC controls received identical treatment but no infusion was given. Ventral striatum lesions were also made in stereotaxic surgery using infusions of quinolinic acid (QA; 20 μg/μl; Sigma) in Dulbecco's phosphate vehicle. Infusion volumes and locations were as follows: 0.4 μl: AP, 1.9; ML, ±1.9 mm; DV, −7.3 mm. Surgical VS controls received identical treatment but no infusion was given. After a 1 week recovery period, all rats were placed on food restriction. Testing began 2 weeks after surgery.

Apparatus. Testing was conducted in 16 standard-sized behavioral boxes (12 × 10 × 12 inches) and other equipment modules purchased from Coulbourn Instruments. A recessed food cup was located in the center of the right wall ~2 cm above the floor. The food cup was connected to a feeder mounted outside of the chamber to deliver 45 mg sucrose pellets (grape or banana flavored; Research Diets). Extensive pilot testing has found these pellets to be equally preferred but discriminable. A house light and cue light were placed on the wall to the left or right of the food cup ~10 cm from the floor. A third cue light was placed on the ceiling of the chamber, in line with the recessed food cup, two inches from the front wall. Additionally, white noise and tone (75 dB, 4 kHz) could be delivered through speakers mounted in the center of the wall. A clicker (1 Hz) was also attached to the front wall, providing a third auditory cue.

Cue conditioning. A summary of the behavioral procedures can be found in Table 1. Before training, rats were reduced to 85% of their baseline weights and exposed to ~50 grape- and banana-flavored sucrose pellets in their home cage on two consecutive days. Rats were then trained to retrieve pellets from the food cup during two sessions in which two 45 mg sucrose pellets (grape and banana) were delivered to the food cup 16 times over the course of an hour. After these training sessions, all rats received 10 d of conditioning in which three visual cues (a house light, a cue light, and a ceiling light, designated A, B, and C, respectively; counterbalanced) were paired with one of two distinctly flavored yet equally preferred sucrose pellets (45 mg grape and banana flavored sucrose pellets, designated O1 and O2, respectively; counterbalanced; Research Diets) in one of two quantities. Cue-outcome associations were as follows: A → O1 × 3, B → O2 × 3, and C → O1 × 1. Cue sessions consisted of 16 presentations of each cue and its associated outcome and quantity, with average intertrial intervals of 2.5 min. For cues A and B, during the 30 s presentation of each light cue, three food pellets were delivered with one food pellet being delivered every 8–10 s. For cue C, one food pellet was delivered at the end of the 30 s cue presentation. For all cues, conditioned responding was measured in the 10 s before first food pellet delivery. This was done so that the food cup rate would reflect responding in anticipation of food (not consumption) in the food cup.

Unblocking. All rats received two unblocking procedures in serial: value unblocking and identity unblocking (order counterbalanced). After conditioning, all rats received 1 d of preexposure to three auditory cues (white noise, tone, and clicker). This preexposure consisted of one session in which each auditory cue was delivered six times for 30 s, with an average intertrial interval of 2.5 min. The next day, the rats began 4 d of compound conditioning. In each identity unblocking session, compound cues, AX and BY, were presented for 30 s. Compound cue AX was paired with O1 × 3, which was the same flavor and quantity of sucrose pellets associated with A. Compound BY was also paired with O1 × 3, which was a different flavor but identical quantity of sucrose pellets as that associated with B. Each session consisted of eight presentations of each compound. The rats also received eight presentations of A → O1 × 3, B → O2 × 3, and C → O1 × 1 as reminder training; average intertrial intervals were 2.5 min. Following compound conditioning, all rats received a probe test, consisting of six unrewarded 30 s presentations of A, B, X, and Y; average intertrial intervals were 2.5 min.

Value unblocking proceeded in an identical manner. In each session, compound cues, AX and CZ, were presented for 30 s. Compound cue CZ was, like AX, paired with O1 × 3, which was the same flavor but greater quantity than that associated with C. Session length and reminder cue sessions were the same as those given in identity unblocking. Again, conditioned responding was measured in the 10 s before first food pellet delivery for both identity and value compound conditioning. The final probe test consisted of six unrewarded 30 s presentations of A, C, X, and Z; average intertrial intervals were 2.5 min. For the probe test data, the entire 30 s cue was analyzed because no rewards were given.

Statistical analysis. Data were acquired using Coulbourn GS2 software. Raw data were processed in Matlab to extract food cup percentage (the amount of time the rat was at the food cup relative to the total cue period sampled) and food cup rate (the number of entries to the food cup per minute during the cue period sampled). These data were analyzed using Statistica. To be included in the probe test analysis, rats must have shown both a food cup rate and food cup percentage greater than zero on more than one of the six X, Y, or Z trials. As noted above, our analysis of these measures of conditioned responding was restricted to the first 10 s of the cues during initial and compound conditioning to avoid contamination with behavior after food pellet delivery. However, during the probe test, when no food was delivered, we analyzed the full 30 s cue period. Previously, value unblocking has been reported using percentage of time spent at the food cup, whereas we and others have reported identity unblocking using rate of food cup responding (Burke et al., 2008). To fairly compare both forms of unblocking, primary analyses were performed with MANOVA in which food cup rate and food cup percentage were dependent variables. Given significant MANOVA results, to maintain consistency with previous reports, value unblocking data were plotted in food cup percentage and identity unblocking data were plotted in food cup rate. Finally, we note that the same pattern of X–Y and X–Z responding was seen in both food cup percentage and food cup rate. However, lesser variation and greater significance was seen in the traditional measure for each form of unblocking. (We are happy to send figures of the nontraditional measure of each form of unblocking upon request.)

Results

Histological results

One VS-lesioned rat failed to fully recover from surgery and did not go on to receive training. OFC lesions targeted the lateral areas on the dorsal bank of the rhinal sulcus, including the lateral and dorsolateral orbital regions and dorsal and ventral agranular regions. Lesions were estimated to have affected >50% of the area within these regions on average. Minimum/maximum lesion extent (Fig. 1*a*) and representative sham and neurotoxic lesions (Fig. 1*c*,*d*) are shown. VS lesions targeted the core subregion and were estimated to have affected >40% of the core, on average. Damage to the adjacent shell subregion was minimal and was estimated to affect <5% of the shell, on average. Minimum/maximum lesion extent (Fig. 1*b*) and representative sham and neurotoxic lesions (Fig. 1*e*,*f*) are shown.

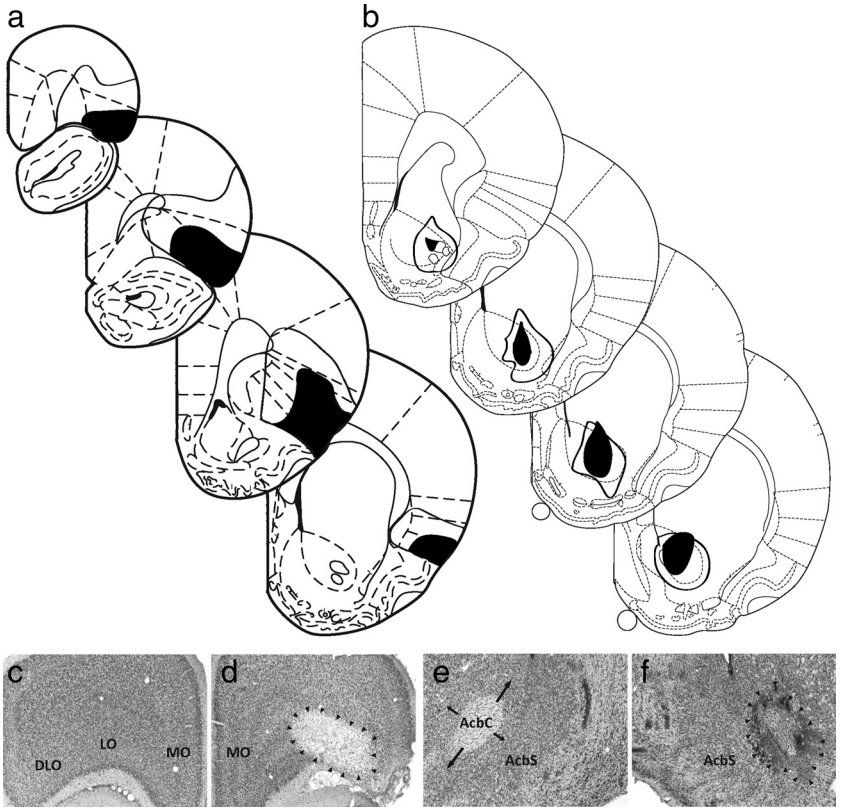

Figure 1. Histology. ***a***, Minimum (black) and maximum (white) OFC lesion extent are shown for bregma +4.7, +3.7, +2.7, and +1.7 (adapted from Paxinos and Watson, 1998). ***b***, Minimum (black) and maximum (white) VS lesion extent are shown for bregma +2.2, +1.7, +1.2, and +0.7 (adapted from Paxinos and Watson, 1998). ***c–f***, Representative intact orbitofrontal cortex (***c***), lesioned orbitofrontal cortex (***d***), intact ventral striatum (***e***), and lesioned ventral striatum (***f***) are shown. DLO, Dorsolateral OFC; LO, lateral OFC; MO, medial OFC; AcbC, accumbens core; AcbS, accumbens shell.

Cue and compound conditioning

During both cue and compound conditioning phases, OFC- and VS-lesioned rats performed similarly to controls, showing equivalent food cup rates to cues A–C as well as compound cues AX, BY, and CZ (Fig. 2). Moreover, all rats showed greater food cup rates to A and B than C, demonstrating sensitivity to value as a group. Importantly, differences reported between groups (see below) during the probe tests cannot be attributed to differences in cue or compound conditioning. This was most critical for the VS-lesioned rats, as previous studies have found VS lesions to produce general impairments in Pavlovian-conditioned responding (Parkinson et al., 1999).

Value sensitivity in cue conditioning

To provide another measure of value sensitivity, we first took the difference between the food cup rate to A and C on the final day of cue conditioning for all individuals. We then compared the distribution of individual value sensitivities between groups with Wilcoxon's rank sum test and found no difference between control and OFC ($p > 0.1$) but a trend toward significance for control and VS ($p = 0.09$). The difference on the final day was indicative of overall responding; day 10 strongly correlated with [A–C] responding over the previous 7 d ($F_{(1,74)} = 33.24$, $p < 0.0001$). Rats showing greater food cup rates for C than A were termed value-insensitive. The percentage of value-insensitive rats in each group was as follows: control, 27% (8/30); OFC, 21% (6/28); and VS, 43% (9/21).

Value and identity unblocking

Following each unblocking procedure, rats underwent the critical probe testing, in which X and Z or X and Y were presented separately to assess how much associative strength was acquired by each cue in the compound phase. The results are presented in Figures 3 and 4. As expected, controls showed learning when either the value or the identity of the expected reward was changed. This was evident in higher responding on the critical first presentation of either of the unblocked cues, Z or Y, compared with the blocked cue, X, in the probe tests (Figs. 3*a*, 4*d*). Interestingly, whereas value unblocking was evident across all the rats, significant identity unblocking was present only in rats that were insensitive to value in initial training. Indeed, there was an inverse correlation between the difference in responding to the high (A) versus the low (C) value cues during initial training, and the degree to which the identity shift unblocked learning to Y in the compound phase (Fig. 4*a*). This interaction between value and identity unblocking suggests that attention to, and representation of, the number of rewards predicted by the training cues normally come at the expense of representation (or signaling) of the different reward identities.

OFC performance in value and identity unblocking

Consistent with our two-pronged hypothesis regarding the circuitry mediating each form of learning, OFC-lesioned rats showed normal value unblocking, responding significantly more to Z than to X in the probe test (Fig. 3*b*). A direct comparison to responding in controls revealed no effect of lesion. At the same time, in the same rats, OFC lesions completely abolished unblocking in response to a shift in reward identity. OFC-lesioned rats responded similarly to Y and X in the probe test (Fig. 4*e*), and a direct comparison with controls revealed a significant interaction. Notably, OFC lesions abolished identity unblocking even though these lesions had no effect on the value sensitivity of these rats. Indeed, neither value-insensitive nor value-sensitive OFC-lesioned rats exhibited identity unblocking (Fig. 4*b*). This overall pattern of results suggests that the OFC is necessary for learning

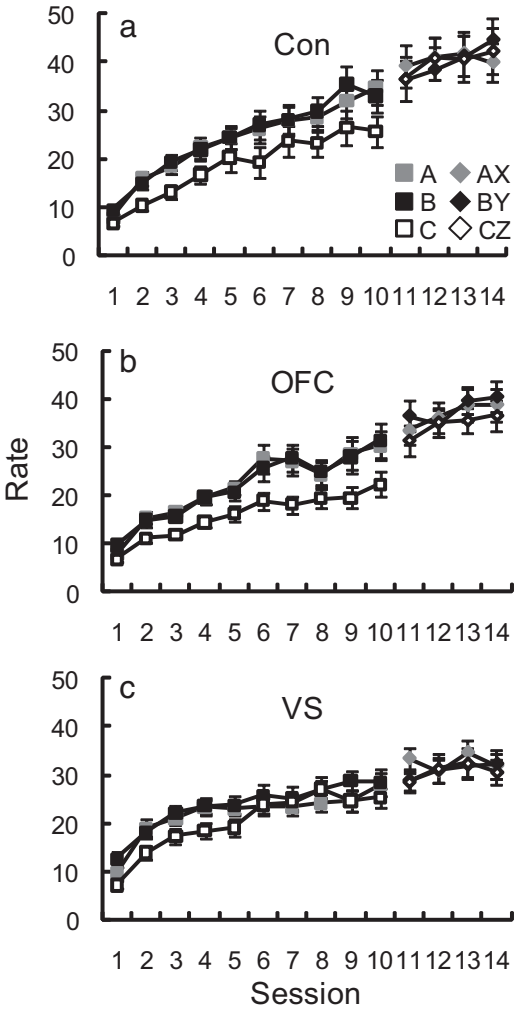

Figure 2. Performance in cue and compound conditioning. ***a–c***, Food cup rate over the initial 10 d of cue conditioning and 4 d of compound conditioning for control (Con; ***a***), OFC (***b***), and VS (***c***) rats are plotted above. ANOVA for control–OFC performance in cue conditioning (cue × day × lesion) revealed significant effects of cue ($F_{(2,100)} = 39.67$, $p < 0.01$), day ($F_{(9,450)} = 46.32$, $p < 0.01$), and cue × day ($F_{(18,900)} = 2.93$, $p < 0.01$). The effect of cue was driven by greater responding to A and B than to C ($p < 0.01$). The main effect of cue was also present when only the final day of conditioning was analyzed ($F_{(2,102)} = 13.50$, $p < 0.01$). ANOVA for performance in compound conditioning (cue × day × lesion) revealed only an effect of day ($F_{(3,153)} = 10.24$, $p < 0.01$). ANOVA for control–VS performance in cue conditioning (cue × day × lesion) revealed significant effects of cue ($F_{(2,82)} = 15.81$, $p < 0.01$), day ($F_{(9,369)} = 38.39$, $p < 0.01$), and cue × day ($F_{(18,738)} = 1.81$, $p < 0.05$). There was a trend toward a cue × lesion interaction but this failed to reach significance ($p = 0.077$). The effect of cue was driven by greater responding to A and B than to C ($p < 0.01$). The main effect of cue was also present when only the final day of conditioning was analyzed ($F_{(2,86)} = 7.67$, $p < 0.01$). ANOVA for control–OFC performance in compound conditioning (cue × day × lesion) revealed only an effect of day ($F_{(3,153)} = 10.24$, $p < 0.01$). Likewise, ANOVA for control–VS performance in compound conditioning (cue × day × lesion) revealed only an effect of day ($F_{(3,129)} = 4.56$, $p < 0.01$). Thus, differences observed between groups during the probe tests cannot be attributed to differences in cue or compound conditioning. Error bars represent mean ± SEM.

when predictions about the identity of the expected reward are violated, but not when prediction errors are based only on value differences.

VS performance in value and identity unblocking

By contrast, VS-lesioned rats were impaired in both value and identity unblocking; these rats responded at similar low levels to Z and X in the value probe test and Y and X in the identity probe test (Figs. 3*c*, 4*f*). Comparison with controls revealed significant main effects and significant interactions of lesion in both probe tests. Interestingly, identity unblocking was abolished despite the fact that VS lesions numerically increased the proportion of rats that were insensitive to value during initial conditioning. Despite this, neither the value-sensitive nor the value-insensitive VS-lesioned rats demonstrated identity unblocking, and there was no correlation between value sensitivity and identity unblocking in VS-lesioned rats (Fig. 4*c*). This pattern of results indicates that the VS is necessary for learning when errors are based on changes in either value and or identity. As discussed below, the latter result is not entirely consistent with our hypothesis or with neural implementations of temporal difference reinforcement learning (TDRL), in which the VS is proposed to serve as a value-based critic.

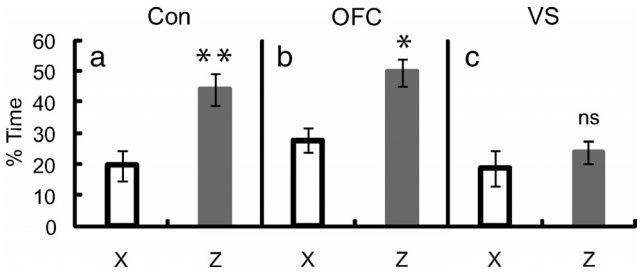

Figure 3. Performance in value unblocking probe. ***a–c***, Percentage time at food cup during the first presentation of X and Z in the value probe is shown for control (Con; ***a***), OFC (***b***), and VS (***c***) rats. MANOVA comparing control–OFC performance revealed a main effect of cue ($F_{(1,50)} = 13.03$, $p < 0.01$) but no effect nor interactions with lesion ($p > 0.1$). Planned comparisons of responding to Z and X found that both control ($p < 0.01$) and OFC ($p < 0.05$) rats demonstrated value unblocking. MANOVA comparing control–VS performance revealed a main effect of lesion ($F_{(1,42)} = 7.81$, $p < 0.01$) and a cue × lesion interaction ($F_{(1,42)} = 6.24$, $p < 0.05$). Planned comparisons of responding to Z and X found control ($p < 0.01$) but not VS ($p > 0.1$) rats demonstrated value unblocking. (**$p < 0.01$; *$p < 0.05$; ns, not significant). Error bars represent mean ± SEM.

Discussion

Generating prediction errors requires a comparison of the reward that is expected and the reward that is received. Here we used procedures that required rats to learn from errors in either reward identity (not accompanied by changes in value) or reward value (without changes in identity). We found that the OFC was necessary for learning driven by changes in reward identity, but not reward value. In contrast, the VS was necessary for learning driven by changes in either reward identity or value. Since the VS—particularly the core region affected by our lesions—receives dense input from the OFC (Voorn et al., 2004), this pattern of results suggests a model whereby information about reward identity signaled by the OFC converges with more general, value-based information in the VS before being sent to downstream areas to support both identity- and value-based error signaling. An obvious candidate to receive this input from VS would be the midbrain dopamine system, which is known to signal reward prediction errors (Montague et al., 1996; Schultz et al., 1997; Hollerman and Schultz, 1998; Waelti et al., 2001).

This model is consistent with data suggesting that the OFC is particularly critical for signaling information about the identity of expected outcomes (Gallagher et al., 1999; Tremblay and Schultz, 1999; Wallis and Miller, 2003; Izquierdo et al., 2004; Padoa-Schioppa and Assad, 2006; Ostlund and Balleine, 2007). Additionally, it is consistent with recent proposals that the VS

supplies predictions regarding expected reward value in models of TDRL (O'Doherty et al., 2004; but see Atallah et al., 2007). However, in these models, the VS has been hypothesized as the site of learning and representation of the general value of expected outcomes of different types, in units of a common currency (Joel et al., 2002). Our results suggest that the VS also incorporates information about the specific features of the expected outcome, signaled from OFC.

Importantly, the involvement of the VS in learning based on changes in the identity of the expected outcome, independent of its general value, is at odds with the fundamental basis of the model-free temporal difference reinforcement learning signal currently applied to understanding interactions between VS and midbrain dopamine neurons in learning (O'Doherty et al., 2004). This discrepancy arises because the expected value signaled by VS in these models, and the resultant error signals generated by downstream dopamine neurons, are calculated in a common currency (Sutton and Barto, 1990; Niv and Schoenbaum, 2008); by definition these representations do not incorporate information about the identity of the impending reward. Having access to information about the reward predicted by the cue requires a model-based representation unlike that used by so-called model-free reinforcement learning (Daw et al., 2005). As a result, a model-free temporal difference reinforcement learning framework cannot account for learning based only on violations of identity expectations, such as that induced here by switches between similarly valued rewards (Niv and Schoenbaum, 2008; Gläscher et al., 2010). The finding that the VS plays a pivotal role in both types of error signaling requires either that a separate system be used involving overlapping networks passing through the VS—one for value and one for identity-specific signals—or that TDRL models be modified to incorporate information beyond general value in their conceptualization of these expectancies and the resultant errors.

Although our findings are inconsistent with a narrow view of VS function in TDRL models, they are consistent with a long-held idea that the VS is a critical component of the goal system (Mogenson et al., 1980). Moreover, they are consistent with recent findings that VS neural activity tends to be coupled to hippocampal systems, linking it to the model-based system (Lansink et al., 2008, 2009) and VS neural activity reflects a search process in model-based decision making (van der Meer and Redish, 2009; van der Meer et al., 2010). Our findings extend these studies by showing that VS is critical to learning that requires access to these kinds of model-based representations.

Lastly, these results have implications for understanding the roles of the OFC and VS in signaling associative information. For the VS, our data support recent evidence that this critical relay station receives convergent input regarding different properties of Pavlovian cues. This is evident in the sources of input to VS (Voorn et al., 2004), which include regions involved in signaling

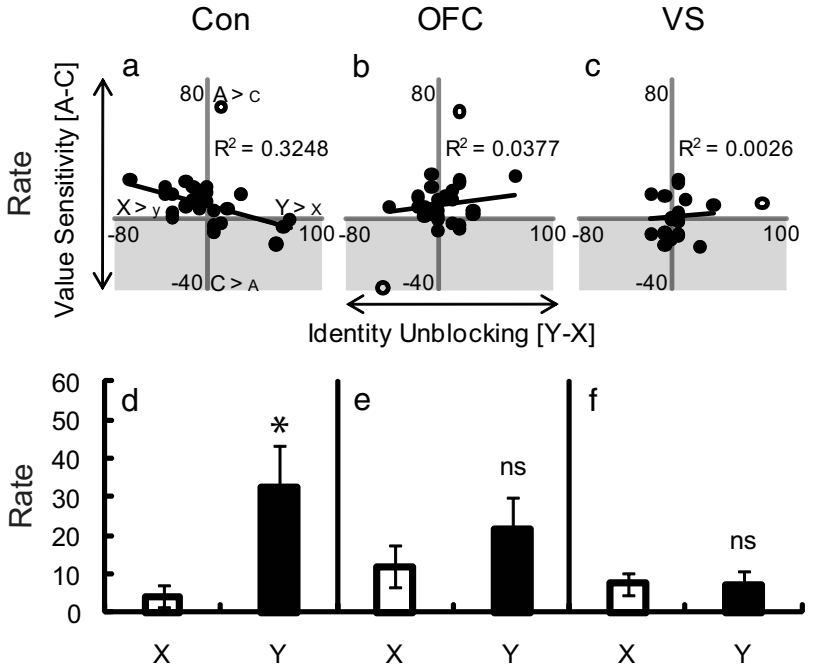

Figure 4. Performance in identity unblocking probe. ***a–c***, The correlation between value sensitivity in initial conditioning (food cup rate A–food cup rate C) and performance in identity unblocking (food cup rate Y–food cup rate X) is plotted for control (Con; ***a***), OFC (***b***), and VS (***c***) rats. The shaded gray region represents value insensitivity. Black circles indicate individual rats, white circles indicate outliers (individuals ±3 SDs from mean value sensitivity or mean identity unblocking). MANOVA (cue × lesion) comparing control–OFC and control–VS performance revealed no effects of cue nor cue × lesion interactions ($F < 0.2$, $p > 0.1$). Simple regression found value sensitivity predicted identity unblocking in controls ($R^2 = 0.3248$, $p < 0.01$) but not OFC ($R^2 = 0.0377$, $p > 0.1$) or VS ($R^2 = 0.0026$, $p > 0.1$) rats. MANCOVA (cue × lesion, covariate–value sensitivity) revealed significant cue × lesion × value interactions for both control–OFC ($F_{(1,48)} = 5.18$, $p < 0.01$) and control–VS ($F_{(1,39)} = 4.78$, $p < 0.05$) performance. *Post hoc* comparisons of responding to X and Y in value-insensitive rats found controls (***d***; $p < 0.05$), but not OFC (***e***) or VS (***f***) rats ($p > 0.1$) demonstrated identity unblocking. (*$p < 0.05$; ns, not significant). Error bars represent mean ± SEM.

both general affective information as well as more specific information about expected outcomes (Cardinal et al., 2002), and also in the contribution of this region to both outcome-specific and value-based behaviors (Parkinson et al., 1999; Corbit et al., 2001; McFarland and Kalivas, 2001; Setlow et al., 2002; Gan et al., 2010; Lex and Hauber, 2010; Singh et al., 2010). Our data indicate that VS is important for learning driven by both types of Pavlovian information.

For the OFC, our results further confirm the critical role this area plays in signaling information about specific outcomes (Pickens et al., 2003; McDannald et al., 2005) and show that such signals are particularly important for learning in situations in which the specifics of the outcome change. A role for the OFC in the generation of temporally specific error signals based on outcome identity is also consistent with recent reports that the OFC is necessary for proper credit assignment (Walton et al., 2010), because credit assignment relies on such teaching signals. Equally interesting, however, is that we failed to find a critical role for the OFC in learning driven by changes in the value of an expected outcome independent of its identity. This indicates that other brain regions are capable of driving learning in the absence of OFC when specific information about the outcome is not required.

References

Atallah HE, Lopez-Paniagua D, Rudy JW, O'Reilly RC (2007) Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. Nat Neurosci 10:126–131.

Barto AG (1994) Adaptive critics and the basal ganglia. In: Models of infor-

mation processing in the basal ganglia (Houk JC, Davis JL, eds), pp 215–232. Cambridge, MA: MIT Press.
Burke KA, Franz TM, Miller DN, Schoenbaum G (2008) The role of orbitofrontal cortex in the pursuit of happiness and more specific rewards. Nature 454:340–344.
Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321–352.
Corbit LH, Muir JL, Balleine BW (2001) The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. J Neurosci 21:3251–3260.
Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711.
Dayan P, Niv Y, Seymour B, Daw ND (2006) The misbehavior of value and the discipline of the will. Neural Netw 19:1153–1160.
Gallagher M, McMahan RW, Schoenbaum G (1999) Orbitofrontal cortex and representation of incentive value in associative learning. J Neurosci 19:6610–6614.
Gan JO, Walton ME, Phillips PE (2010) Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci 13:25–27.
Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) Prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595.
Holland PC (1984) Unblocking in Pavlovian appetitive conditioning. J Exp Psychol Anim Behav Process 10:476–497.
Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1:304–309.
Izquierdo A, Suda RK, Murray EA (2004) Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. J Neurosci 24:7540–7548.
Joel D, Niv Y, Ruppin E (2002) Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw 15:535–547.
Kamin LJ (1969) Predictability, suprise, attention, and conditioning. In: Punishment and aversive behavior (Campbell BA, Church RM, eds), pp 242–259. New York: Appleton-Century-Crofts.
Lansink CS, Goltstein PM, Lankelma JV, Joosten RN, McNaughton BL, Pennartz CM (2008) Preferential reactivation of motivationally relevant information in the ventral striatum. J Neurosci 28:6372–6382.
Lansink CS, Goltstein PM, Lankelma JV, McNaughton BL, Pennartz CM (2009) Hippocampus leads ventral striatum in replay of place-reward information. PLoS Biol 7:e1000173.
Lex B, Hauber W (2010) The role of nucleus accumbens dopamine in outcome encoding in instrumental and Pavlovian conditioning. Neurobiol Learn Mem 93:283–290.
McDannald MA, Saddoris MP, Gallagher M, Holland PC (2005) Lesions of orbitofrontal cortex impair rats' differential outcome expectancy learning but not conditioned stimulus-potentiated feeding. J Neurosci 25:4626–4632.
McFarland K, Kalivas PW (2001) The circuitry mediating cocaine-induced reinstatement of drug-seeking behavior. J Neurosci 21:8655–8663.
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, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. J Neurosci 16:1936–1947.
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.
Ostlund SB, Balleine BW (2007) Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental learning. J Neurosci 27:4819–4825.
Padoa-Schioppa C, Assad JA (2006) Neurons in orbitofrontal cortex encode economic value. Nature 441:223–226.
Parkinson JA, Olmstead MC, Burns LH, Robbins TW, Everitt BJ (1999) Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive Pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. J Neurosci 19:2401–2411.
Paxinos G, Watson C (1998) The rat brain in stereotaxic coordinates, Ed 4. San Diego: Academic Press.
Pickens CL, Saddoris MP, Setlow B, Gallagher M, Holland PC, Schoenbaum G (2003) Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. J Neurosci 23:11078–11084.
Rescorla RA (1999) Learning about qualitatively different outcomes during a blocking procedure. Anim Learn Behav 27:140–151.
Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK (2009) A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nat Rev Neurosci 10:885–892.
Schultz W, Dayan P, Montague PR (1997) A neural substrate for prediction and reward. Science 275:1593–1599.
Setlow B, Holland PC, Gallagher M (2002) Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive Pavlovian second-order conditioned responses. Behaval Neurosci 116:267–275.
Singh T, McDannald MA, Haney RZ, Cerri DH, Schoenbaum G (2010) Nucleus accumbens core and shell are necessary for reinforcer devaluation effects on Pavlovian conditioned responding. Front Integr Neurosci 4:126.
Sutton RS, Barto AG (1990) Time-derivative models of Pavlovian reinforcement. In: Learning and computational neuroscience: foundations of adaptive networks (Gabriel M, Moore J, eds), pp 497–537. Boston: MIT.
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, 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.
Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43–48.
Wallis JD, Miller EK (2003) Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur J Neurosci 18:2069–2081.
Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron 65:927–939.