Striatal intrinsic reinforcement signals during recognition memory: Relationship to response bias and dysregulation in schizophrenia

9 authors, including:

Striatal intrinsic reinforcement signals during recognition memory: relationship to response bias and dysregulation in schizophrenia

Daniel H. Wolf[1]*, **Raphael T. Gerraty**[1], **Theodore D. Satterthwaite**[1], **James Loughead**[1], **Timothy Campellone**[1], **Mark A. Elliott**[2], **Bruce I. Turetsky**[1], **Ruben C. Gur**[1,2,3] and **Raquel E. Gur**[1,2]

[1] Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

[2] Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA

[3] Philadelphia Veterans Administration Medical Center, Philadelphia, PA, USA

Ventral striatum (VS) is a critical brain region for reinforcement learning and motivation, and VS hypofunction is implicated in psychiatric disorders including schizophrenia. Providing rewards or performance feedback has been shown to activate VS. Intrinsically motivated subjects performing challenging cognitive tasks are likely to engage reinforcement circuitry even in the absence of external feedback or incentives. However, such intrinsic reinforcement responses have received little attention, have not been examined in relation to behavioral performance, and have not been evaluated for impairment in neuropsychiatric disorders such as schizophrenia. Here we used fMRI to examine a challenging "old" vs. "new" visual recognition task in healthy subjects and patients with schizophrenia. Targets were unique fractal stimuli previously presented as salient distractors in a visual oddball task, producing incidental memory encoding. Based on the prediction error theory of reinforcement learning, we hypothesized that correct target recognition would activate VS in controls, and that this activation would be greater in subjects with lower expectation of responding correctly as indexed by a more conservative response bias. We also predicted these effects would be reduced in patients with schizophrenia. Consistent with these predictions, controls activated VS and other reinforcement processing regions during correct recognition, with greater VS activation in those with a more conservative response bias. Patients did not show either effect, with significant group differences suggesting hyporesponsivity in patients to internally generated feedback. These findings highlight the importance of accounting for intrinsic motivation and reward when studying cognitive tasks, and add to growing evidence of reward circuit dysfunction in schizophrenia that may impact cognition and function.

Keywords: memory, reward, intrinsic motivation, ventral striatum, schizophrenia

INTRODUCTION

Any account of cognition is incomplete without integrating the influence of emotion and motivation. Interactions between cognition, emotion, and motivation have attracted increasing research, both in healthy individuals and patients with psychiatric disorders such as schizophrenia (Barch, 2005; Phelps, 2006; Satterthwaite et al., 2009, 2010; Duckworth et al., 2011; Murty et al., 2011).

Building upon an extensive animal literature, human fMRI research highlights the role of the ventral striatum (VS) in driving both motivation and learning in response to a wide range of rewards (Robbins and Everitt, 1996; McClure et al., 2004). While most fMRI studies focusing on VS have examined responses to explicit delivery of reinforcers such as money, VS activation has also been reported in response to purely cognitive feedback such as information reflecting performance accuracy (Rodriguez et al., 2006; Tricomi and Fiez, 2008). Rather than reflecting

reward value *per se*, VS fMRI responses have been shown to reflect positive prediction errors that occur when outcomes are better than expected (McClure et al., 2003; Schultz, 2010). VS thus may be a critical hub for motivation–cognition interactions, especially when cognitive performance can generate positive outcomes.

Memory is a critical aspect of cognition that is strongly modulated by inter-related emotional and motivational systems (Dere et al., 2010; Murty et al., 2011). Human fMRI studies have demonstrated enhancement of memory encoding by rewards that increase dopamine signaling and VS responses (Adcock et al., 2006; Wittmann et al., 2008; Bunzeck et al., 2011). Less attention has been paid to the role of reward and reinforcement in memory retrieval. However, successful retrieval is likely to itself produce reward signals as task success is psychologically rewarding. fMRI activation in VS/anterior caudate in response to successful retrieval has been observed in some studies including two meta-analyses

(von Zerssen et al., 2001; Iidaka et al., 2006; McDermott et al., 2009; Spaniol et al., 2009), but has generally not been interpreted in terms of reward processes given the cognitive focus of the research (see von Zerssen et al., 2001 for an exception). A single fMRI study has explicitly focused on striatal responses as a reflection of goal-directed reward processing during memory retrieval (Han et al., 2010). In that study, anterior caudate robustly activated to correctly recognized targets even in the absence of external rewards or trial-by-trial feedback. By manipulating task rewards, Han et al. (2010) further demonstrated that striatal activation during successful recognition was not due to memory retrieval *per se*, but to the reward significance of correct performance.

Schizophrenia is associated with a range of cognitive deficits, including prominent recognition memory deficits (Saykin et al., 1991; Heinrichs and Zakzanis, 1998; Ragland et al., 2004; Gur et al., 2007). Deficits in motivation are also often present (Wolf, 2006; Foussias and Remington, 2008). Hypofunction of VS has been found in schizophrenia in response to primary and monetary rewards as well as novelty (Juckel et al., 2006; Wolf et al., 2008; Waltz et al., 2009, 2010). VS hypofunction has also been observed in patients with schizophrenia during a cognitive inhibitory control task (Vink et al., 2006), and in dorsal striatum in a working memory task (Koch et al., 2008). However, striatum has not been a regional focus of interest in most recognition memory studies in schizophrenia. VS deficits in this context have not to our knowledge been reported, and are not identified in recent meta-analyses (Achim and Lepage, 2005; Ragland et al., 2009).

Here we examine data from controls and patients with schizophrenia performing a challenging old vs. new visual recognition memory task, focusing on VS responses during successful target recognition. Our hypothesis was that intrinsically motivated performance would generate VS reward prediction error signals in controls, an effect diminished in patients. Consistent with Han et al. (2010) and with the intuitive notion that being correct is intrinsically rewarding, we expected VS activation to successful retrieval even in the absence of explicit feedback or external rewards, as subjects internally evaluated their own performance relative to their expectations. We hypothesized that a difficult memory task would produce lower expectations of correct retrieval and hence generate robust prediction errors and VS activation for correctly identified targets. A previously untested prediction of this hypothesis is that subjects with a more conservative response bias, and hence a lower tendency to identify items as previously seen, would generate larger prediction errors manifesting as greater VS activation to correct retrieval. Given the association of schizophrenia with deficits in striatal function as well as impairments in memory, reward processing, and motivation, we predicted patients would show reduced activation of VS during successful retrieval, and a lack of VS modulation by response bias.

MATERIALS AND METHODS
PARTICIPANTS
The study included 26 clinically stable patients with schizophrenia (12 female), and 27 healthy comparison subjects. Two control subjects were excluded due to excessive motion, leaving a final sample of 25 healthy comparison subjects (12 female). The groups did not differ significantly with respect to age, sex, handedness, education level, or parental education level (**Table 1**).

All subjects were evaluated with a structured clinical interview (First et al., 2001) and symptom scales including SANS and SAPS (Andreasen, 1984a,b) by individuals trained to a criterion reliability of ICC = 0.90. Collateral medical and psychiatric histories were obtained from records, family members, and caregivers. Patients met criteria for a diagnosis of DSM-IV schizophrenia as established in a consensus conference based on all information obtained during study procedures. Except for one unmedicated individual, patients were on a stable antipsychotic regimen (4 on first-generation, 21 on second/third-generation). No participants were diagnosed with any other psychiatric disorders, recent substance use disorders (within past 6 months) or current use other than nicotine, or any other disease affecting brain function. All participants were volunteers at the Schizophrenia Research Center at the Hospital of the University of Pennsylvania and provided written consent after receiving a full description of study procedures. Study procedures were approved by the University of Pennsylvania IRB.

TASK
Participants were presented with a series of 60 images of fractals in a forced choice recognition memory paradigm (**Figure 1**). Half of the fractals had been previously displayed once each for 2 s as salient distractors during a visual oddball task performed ~10 min earlier in the scanning session, yielding 30 targets and

Table 1 | Demographics, clinical variables, and performance.

	Patients ($n = 26$)	Controls ($n = 25$)	p Value
Gender (% female)	46	48	0.99
Handedness (% right)	85	92	0.67
Age (year)	38.0 (10.7)	38.0 (10.3)	0.99
Education (year)	13.1 (2.6)	14.3 (1.9)	0.06
Parental education	13.31 (2.75)	13.69 (3.18)	0.63
SANS global	1.53 (0.82)		
SAPS global	1.40 (0.99)		
Overall accuracy	0.57 (0.09)	0.65 (0.07)	0.002
Hit rate	0.46 (0.20)	0.50 (0.14)	0.38
Miss rate	0.54 (0.21)	0.50 (0.14)	0.38
Correct rejection rate	0.69 (0.21)	0.80 (0.15)	0.049
False alarm rate	0.31 (0.21)	0.20 (0.15)	0.049
P_r	0.15 (0.17)	0.29 (0.13)	0.002
B_r	−0.13 (0.21)	−0.21 (0.17)	0.23
Hit RT (s)	1.17 (0.21)	1.07 (0.20)	0.11

Values are mean (SD) except where otherwise noted. p Values are two-tailed based on unpaired t-tests (except for gender and handedness which used Fisher's exact test). SAPS and SANS scores are the average of the global items (SANS attention subscale not included). Proportions for hits, misses, correct rejections, and false alarms are presented here as raw rates without the addition of 0.5 to the numerator and 1 to the denominator utilized by Snodgrass and Corwin (1988; see formulas in Materials and Methods).
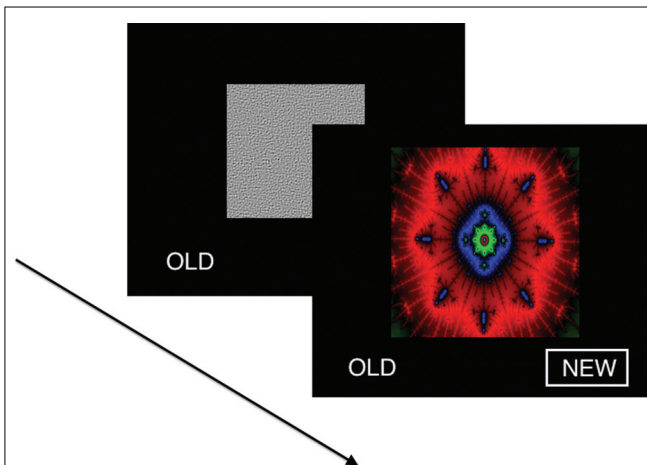
FIGURE 1 | Visual recognition task. In each of 60 trials, a unique fractal image was presented in a forced choice recognition memory paradigm. Half of the fractals had been previously displayed once each as salient distractors during a visual oddball task performed ~10 min earlier, yielding 30 targets and 30 foils. Stimuli were presented for 3 s in a fast event-related design, with a variable 0–18 s interstimulus interval during which a homogeneous visual noise gray background was displayed.

30 foils. Subjects were not specifically instructed to remember these items. The oddball task was adapted from an experiment previously described in detail (Gur et al., 2007). In the present recognition memory experiment, each stimulus was presented for 3 s in a fast event-related design, with a variable 0–18 s (mean 3.4 s) interstimulus interval during which a homogeneous visual noise gray background was displayed. Responses and reaction times were recorded with a two-button response pad (FORP™, Current Design Inc., Philadelphia, PA, USA). Total task duration was 6.3 min.

IMAGE ACQUISITION

Data were acquired on a 3-Tesla Siemens Magnetom TIM TRIO system (Erlangen, Germany) using an 8-channel head coil. A TI weighted whole-brain structural scan was acquired for use in co-registration to a standard brain atlas (MPRAGE, TR = 1630 ms, TE = 3.87 ms, FOV = 250 mm, matrix = 192 × 256, effective voxel resolution of 1 mm × 1 mm × 1 mm). Functional BOLD images were acquired using a 2-D echo-planar sequence (TR = 3000 ms, TE = 30 ms, FOV = 192, slice thickness = 3 mm, matrix = 64 × 64, effective resolution = 3 mm × 3 mm × 3 mm).

PERFORMANCE ANALYSIS

Subject performance during recognition trials was evaluated using a measure of discrimination index, P_r, and a measure of response bias, B_r (Snodgrass and Corwin, 1988). These variables were computed with the following formulas:

$$\text{Hit rate (HR)} = (0.5 + \text{number correct targets})/$$
$$(1 + \text{number total targets})$$
$$\text{False alarm rate (FR)} = (0.5 + \text{number incorrect foils})/$$
$$(1 + \text{number total foils})$$
$$P_r = \text{HR} - \text{FR}$$
$$B_r = [\text{FR}/(1 - P_r)] - 0.5$$

P_r provides a measure of discrimination accuracy that is unbiased by a subject's tendency toward a greater total number of responses; higher values reflect a greater degree of accuracy in discriminating between old and new items. B_r measures the overall tendency of a subject to respond that any given stimulus is new or old. Following Sergerie et al. (2007), we adapted the original B_r formula (where B_r $r = 0.5$ reflected neutrality or no bias), subtracting 0.5 so that 0 becomes the neutral point. Negative B_r values thus denote a tendency to respond that previously displayed images are new (conservative response bias), whereas positive B_r values indicate a tendency to call new stimuli old (liberal response bias). P_r, B_r and response times (RT, within-subject median for HITS) were compared between groups using unpaired t-tests ($\alpha = 0.05$, two-tailed).

IMAGE ANALYSIS

Images were preprocessed and analyzed using FEAT, part of FSL: FMRIB's Software Library (www.fmrib.ox.ac.uk/fsl). Images were slice-timing corrected, motion corrected to median image using a tri-linear interpolation with 6 degrees of freedom, high-pass filtered (100 s), spatially smoothed using a Gaussian kernel of 6 mm FWHM, and grand mean scaled. Subject-level time-series analysis was carried out using a general linear model with four event types: target correct (HIT), target incorrect (MISS), foil correct (correct rejection, CR), and foil incorrect (false alarm, FA). All event types were convolved with double-gamma canonical hemodynamic response functions. Six motion parameters were included in the model as confound regressors. Group level mixed-effects analyses (FLAME 1) were performed for controls, patients, controls > patients, and patients > controls.

Given our *a priori* hypotheses regarding reward signals in the VS and the low sensitivity of whole-brain analysis for small subcortical structures, we performed a region of interest (ROI) analysis of the HIT > baseline contrast within bilateral VS, defined as the combination of 14 mm radius spheres centered on MNI coordinates ±11, 9, −2 from Knutson et al. (2005). In order to evaluate effects outside of the VS, we followed this ROI analysis with an exploratory whole-brain analysis of the HIT > baseline contrast. We focused on the HIT > baseline contrast because it related most clearly to our hypotheses regarding successful task performance; the results of Han et al. (2010) suggested that correct target identification is more goal-relevant to subjects than foil performance when no explicit incentives are provided. In order to evaluate the specificity of the HIT > baseline response with a tighter contrast, we additionally evaluated the contrast of HIT > MISS on a voxelwise basis. We observed group differences in the number of correct and incorrect foil responses but not the number of HIT or MISS responses, adding an important *post hoc* rationale to our *a priori* hypothesis-driven decision to focus on the above contrasts. To investigate the relationship between task performance and brain responses, B_r, and P_r were included as covariates in both ROI and whole-brain analyses. Both whole-brain and ROI analyses were corrected for multiple comparisons using a Monte Carlo simulation with AFNI AlphaSim (R. W. Cox, National Institutes of Health) with 10,000 permutations and a probability of spatial extent $p < 0.01$. For the VS ROI, mask volume was 22,328 mm³, and the whole-brain mask volume was 1,415,672 mm³. Voxel height threshold for all analyses

was $Z > 2.33$ except for the exploratory whole-brain HIT > baseline analysis which used $Z > 3.10$ in order to separate anatomically meaningful clusters for this robust effect.

RESULTS

BEHAVIORAL RESULTS
Patients performed significantly worse than controls in overall accuracy and discrimination accuracy (P_r), but HIT rate and HIT reaction times did not significantly differ (**Table 1**). The reduction in discrimination accuracy in patients related to decreased foil performance with an increase in FAs. Both groups showed a conservative response bias as indicated by a negative B_r which did not differ significantly between groups. In both groups, a more negative (conservative) B_r correlated with lower hit rate (proportion of target *trials* which were hits; control $r = 0.75$, patient $r = 0.79$), a relationship implicit in the formula for calculating B_r. However, more negative B_r also correlated with *higher target response specificity* (proportion of target *responses* which were hits; control $r = -0.81$, patient $r = -0.57$). A *post hoc* simulation analysis (performed in MATLAB, code available upon request) demonstrated that this relationship was statistically significant (control, $p = 0.0001$; patient, $p = 0.01$) and did not result from a trivial mathematical relationship: the correlation observed across many "subjects" with random performance was essentially 0 ($r = -0.03$).

A PRIORI ROI RESULTS
Our *a priori* analysis of HIT > baseline in the VS revealed a robust response among healthy subjects in bilateral regions encompassing the anterior caudate, nucleus accumbens, and putamen (**Figure 2**). In contrast, patients did not show significant HIT > baseline activation in this region. Accordingly, the between-group contrast revealed greater activation for controls than patients in VS (significant on left, subthreshold on right).

WHOLE-BRAIN ANALYSIS
As expected, healthy subjects robustly activated regions associated with motivation and reinforcement processing in the HIT > baseline contrast, including the VS, midbrain, anterior insula/posterior orbital frontal cortex, and dorsal anterior cingulate/paracingulate

FIGURE 2 | Ventral striatum HIT > baseline ROI analysis. (A) Results in healthy controls. **(B)** Results from control > patient contrast. Images thresholded at $Z > 2.3$ and cluster corrected for multiple comparisons with spatial extent $p < 0.01$. Note: A similar cluster to that shown in **(B)** was the only significant result of the exploratory whole-brain controls > patients analysis. Left hemisphere shown on left side of images.

(**Figure 3**; **Table 2**). Patients activated a qualitatively similar set of regions. The voxelwise contrast of controls > patients contrast
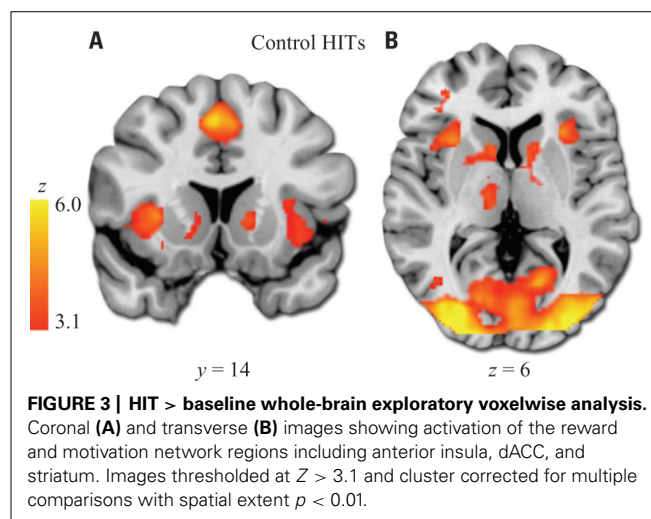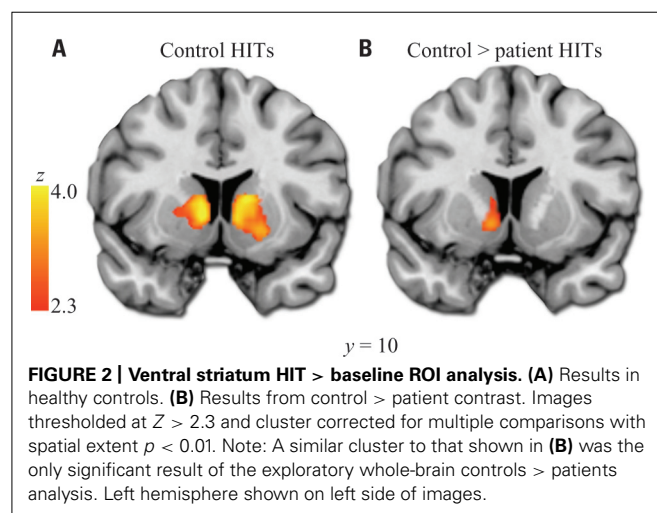

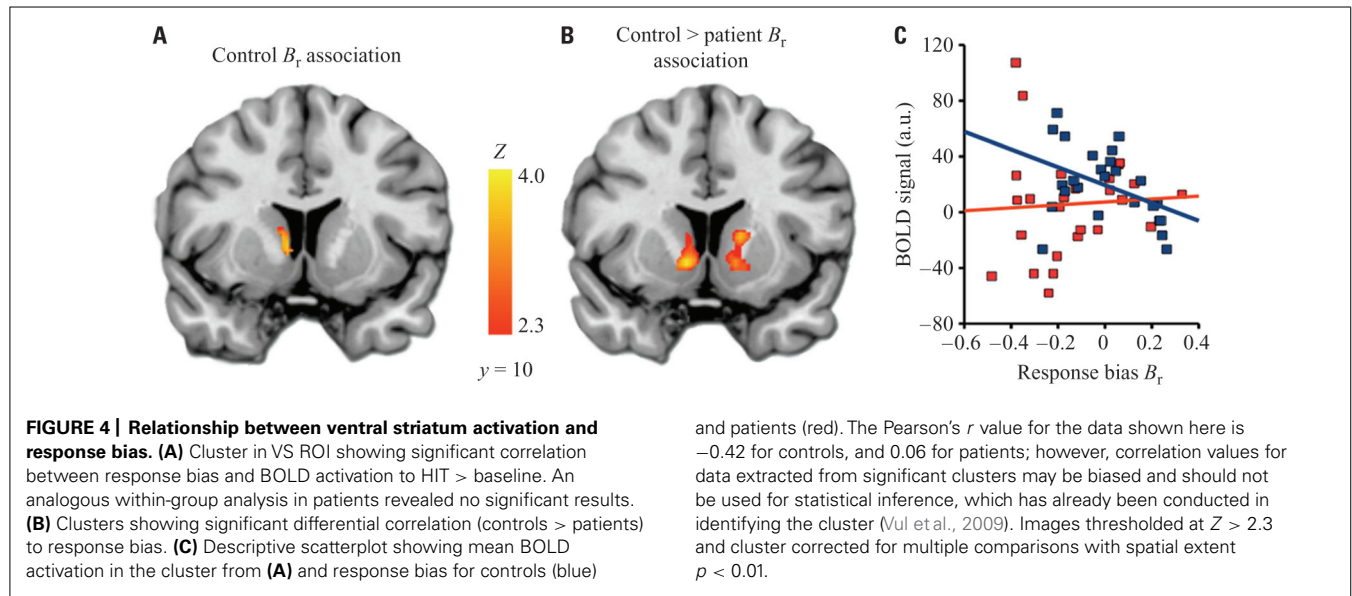**FIGURE 3 | HIT > baseline whole-brain exploratory voxelwise analysis.** Coronal **(A)** and transverse **(B)** images showing activation of the reward and motivation network regions including anterior insula, dACC, and striatum. Images thresholded at $Z > 3.1$ and cluster corrected for multiple comparisons with spatial extent $p < 0.01$.

Table 2 | HIT > baseline whole-brain significant clusters.

Voxels	Region	Max-Z	x	y	z
Controls					
24184	Visual cortex/fusiform/ cerebellum/midbrain*/ sup. parietal	7.55	28	−56	−20
1580	R insula/OFC/MFG	5.48	32	26	−2
1137	L insula/OFC/MFG	5.58	−32	20	−2
938	dACC/paracingulate	5.62	−4	14	48
862	L IFG	5.11	−44	4	30
483	R IFG	5.26	38	2	30
480	R caudate/thalamus	4.18	14	−12	10
327	L caudate	4.21	−10	8	0
239	L thalamus	4.54	−12	−16	8
Patients					
16629	Visual cortex/fusiform/ cerebellum	7.28	22	−92	6
4304	Superior parietal	5.64	−50	−24	50
610	dACC/paracingulate	4.8	0	6	48
292	L insula	4.64	−34	16	4
257	L thalamus	4.29	−14	−18	8
250	R insula	4.67	34	18	4
Controls > patients					
45	L ventral striatum	3.45	−14	0	−8

Data thresholded at $Z > 3.1$ and corrected for multiple comparisons using AlphaSim with a spatial extent $p < 0.01$.
OFC, orbitofrontal cortex; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; dACC, dorsal anterior cingulate cortex. Peak-Z x, y, z coordinates are in MNI space.
*The cluster listed is in the right posterior midbrain; however, an adjacent subthreshold cluster was present in the right substantia nigra.

FIGURE 4 | Relationship between ventral striatum activation and response bias. (A) Cluster in VS ROI showing significant correlation between response bias and BOLD activation to HIT > baseline. An analogous within-group analysis in patients revealed no significant results. **(B)** Clusters showing significant differential correlation (controls > patients) to response bias. **(C)** Descriptive scatterplot showing mean BOLD activation in the cluster from **(A)** and response bias for controls (blue) and patients (red). The Pearson's r value for the data shown here is -0.42 for controls, and 0.06 for patients; however, correlation values for data extracted from significant clusters may be biased and should not be used for statistical inference, which has already been conducted in identifying the cluster (Vul et al., 2009). Images thresholded at $Z > 2.3$ and cluster corrected for multiple comparisons with spatial extent $p < 0.01$.

revealed only one significant cluster located in the left VS, which was identical to that found in our ROI analysis. Patients did not activate any regions to a greater extent than controls.

BIAS CORRELATION

Within the *a priori* VS ROI among healthy controls, there was a significant negative relationship between B_r and left VS activation to HIT > baseline (**Figure 4**). This indicates that control subjects who were less likely to identify a given stimulus as a target had higher activation of the striatum when they correctly identified a target. Patients did not show a significant correlation with B_r in the striatum. When directly compared, this produced a significant differential association with B_r in bilateral VS between groups.

SPECIFICITY ANALYSIS

The contrast of HIT > MISS in healthy controls produced a pattern of activation similar to the results of the analysis of the HIT > baseline contrast, including activation of bilateral striatum, insula, and dorsal anterior cingulate, with subthreshold anterior midbrain activation. However, patients did not activate these reward-related regions. Direct comparison of controls and patients revealed a significant control > patient response in the right caudate and insula, among other regions (**Table 3**).

POTENTIAL CONFOUNDS

To assess whether the group difference in VS activation was driven by group differences in performance accuracy, we examined the relationship of P_r to VS. P_r did not correlate with target correct responses in VS in either the control or the patient group. Inclusion of P_r as an across-group covariate (thus statistically accounting for group differences in P_r) did not alter the significance of group differences in VS activation or B_r correlation. Thus, VS findings were not confounded by P_r effects. Including or excluding P_r and B_r from the models did not change the above within-group

or between-group findings for HIT > baseline or HIT > MISS. Gender, age, and smoking status did not relate significantly to VS activation nor did modeling their effects alter the significance of reported findings.

DISCUSSION

Our results confirm earlier reports that VS and anterior caudate respond preferentially to successful recognition (von Zerssen et al., 2001; Han et al., 2010). We extend this finding beyond prior literature by identifying a relationship of VS activation to response bias, and by demonstrating a deficit in these responses in patients with schizophrenia.

VS RESPONDS TO CORRECT RECOGNITION WITHOUT EXPLICIT FEEDBACK OR REWARD

Extensive literature documents activation of VS to external rewards. However, research examining VS responses reflecting intrinsic motivation is quite sparse. Prior studies have demonstrated that performance feedback activates VS (Rodriguez et al., 2006; Tricomi and Fiez, 2008; Daniel and Pollmann, 2010). Here we tested the hypothesis that striatal reward signals would be present even in the absence of explicit feedback, reflecting monitoring and evaluation of responses by intrinsically motivated subjects. The VS response we observe to correct targets is consistent with this hypothesis as well as the findings of Han et al. (2010). In that study, the anterior caudate response to correct targets occurred in the absence of external rewards or feedback, and was further enhanced when a monetary incentive was provided for correct target recognition. While Han et al. (2010) reported a peak of activation in anterior caudate dorsal to VS, the activated striatal regions overlapped with those observed in this study. Furthermore, Han et al. (2010) found that activation in VS itself correlated most strongly with personality measures of reward sensitivity. While our study is not a direct replication of Han et al. (2010), the differences in methods make the similarities in our findings all the more striking.

Table 3 | HIT > MISS whole-brain significant clusters.

Voxels	Region	Max-Z	x	y	z
Controls					
715	dACC	3.39	−6	18	20
634	L frontal pole	3.37	−46	42	−4
604	L lateral occipital cortex	3.61	−30	−92	16
459	R caudate/thalamus	3.89	10	12	−2
449	R insula	4.21	36	20	−12
396	L insula	3.63	−30	24	−6
358	L IFG	3.17	−46	−70	−6
317	L caudate/thalamus	3.55	−12	2	10
308	R MFG	2.94	52	24	20
261	R IFG	3.65	48	8	34
223	R intracalcarine	3.43	10	−86	−4
201	R superior parietal	3.4	34	−54	38
170	L MFG	3.18	−40	32	20
Patients					
598	L supramarginal gyrus	3.29	−52	−34	48
262	L superior parietal	3.54	−30	−52	54
210	L precentral gyrus	3.33	−28	−10	42
180	L IFG	3.61	−54	2	22
Controls > patients					
310	R insula	3.62	34	24	−14
280	R lingual gyrus	3.23	12	−74	−8
272	Cuneus	2.9	−4	−72	20
243	L lingual gyrus	3.26	−12	−60	4
242	R fusiform	3.68	30	−38	−18
219	R lateral occipital cortex	3.18	34	−76	46
181	R caudate/thalamus	3.24	14	14	−2
158	L fusiform	3.11	−24	−88	−4

Data thresholded at Z > 2.33 and corrected for multiple comparisons using AlphaSim with a spatial extent p < 0.01. Peak-Z x, y, z coordinates are in MNI space.
OFC, orbitofrontal cortex; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; dACC, dorsal anterior cingulate cortex.

VS RESPONSE IS CORRELATED WITH RESPONSE BIAS

By conceptualizing VS responses within the framework of reward prediction errors, we successfully predicted the novel finding that subjects with a more conservative response bias would activate VS more during correct retrieval. Importantly, this result was not driven by discrimination accuracy, which was accounted for in our model. Prediction errors reflect the difference between actual and expected outcome values. Expected outcome value depends on the subjective value of the outcome and its probability of occurring. In contrast, actual outcome value depends on subjective value and the probability that the outcome *has actually been obtained*. In typical reward tasks

feedback is given, making the actual outcome certain. However, in a task like the current one with no feedback, the probability that the desired outcome has occurred (here, a HIT) corresponds to subjective confidence in one's response. Therefore, VS prediction error signals should be enhanced by a higher subjective confidence in the correctness of a given response. A conservative response bias could therefore increase prediction errors in two ways. First, subjects with a conservative response bias will have lower expectations that any given stimulus will be a correctly identified target, thereby lowering the expected outcome value. Second, subjects with a conservative bias will therefore tend to report recognition only when they have high confidence they are correct (higher actual outcome value). Thus, subjects with a conservative response bias will have low expectations of correctly identifying a target on any given trial, but higher response confidence will generate a greater outcome value when a target is indeed correctly identified. Together, these two factors could produce greater positive prediction errors. Consistent with this account, conservative bias in the current study strongly correlated both with reduced target hit rate (fewer correct identifications) and with an increased proportion of target responses that were in fact correct. Thus, a subject with a very conservative response bias was unlikely to respond "old" to a previously displayed target (reducing hit rate), but had a high likelihood of being correct when she *did* respond "old."

One reason that such an effect has not been previously described may stem from the difficulty of the task. Greater task difficulty lowers pre-response confidence, and also may increase the subjective value of correct responses, producing greater prediction errors responses in VS. The task used here was quite challenging, as encoding occurred incidentally during brief presentations of abstract visual fractals as oddball distractors, without any expectation of a memory test. Old vs. new recognition tasks used in prior fMRI studies are typically designed to be easier, in order to reduce the number of errors (which are typically unanalyzed), and to ensure all subjects perform at greater than chance levels. We are currently testing the prediction that greater task difficulty enhances VS response to correct responses in a follow-up study.

VS IS PART OF AN INTRINSIC REINFORCEMENT NETWORK

Our ROI analysis focused on the VS based on its central importance in reward and motivation. An exploratory whole-brain analysis revealed a network of other regions previously associated with reinforcement learning that were also strongly activated to successful retrieval, including midbrain, insula/orbital frontal cortex, and dorsal anterior cingulate cortex (Kirsch et al., 2003; Sescousse et al., 2010; D'Cruz et al., 2011). Importantly, activation of these regions was not due to non-specific elements of the task, as these regions were also recruited in the contrast of HIT > MISS. The activation of this network, which is not typically associated with memory retrieval *per se*, further supports our contention that VS responses seen here reflect reward signaling. These findings underline the importance of reward-related circuits in understanding performance and brain responses in cognitive tasks.

PATIENTS WITH SCHIZOPHRENIA HAVE DIMINISHED VS INTRINSIC REINFORCEMENT RESPONSE

The hypothesized deficit in VS activation to correct targets in the schizophrenia group was confirmed. Notably, the result from the VS ROI analysis was also the single statistically significant group difference in the exploratory whole-brain analysis. Multiple fMRI studies have examined recognition memory tasks in schizophrenia patients, identifying abnormalities in various cortical and subcortical regions (Weiss and Heckers, 2001; Achim and Lepage, 2005; Leavitt and Goldberg, 2009; Ragland et al., 2009). To our knowledge prior recognition memory studies have not identified VS deficits. This may be a consequence of using easier tasks (to reduce group performance differences) that do not activate VS robustly even in controls. In addition, in studies where group differences in VS activation are less robust than ours, a VS ROI approach might have been necessary but has generally not been utilized. However, VS hypofunction in schizophrenia is commonly observed in tasks involving reward (Juckel et al., 2006; Wolf et al., 2008; Waltz et al., 2009, 2010). Two studies have reported striatal hypofunction in schizophrenia using different cognitive tasks without an explicit reward component. Koch et al. (2008) reported reduced dorsal striatum activation in schizophrenia during successful working memory retrieval, whereas Vink et al. (2006) reported diminished anterior caudate response in both schizophrenia patients and unaffected family members during a response inhibition task. Although these findings were not framed as deficits in intrinsic reinforcement, both these prior studies and the current results are consistent with such an interpretation. Taken together with the existing literature, the current study provides additional evidence of striatal dysfunction in schizophrenia, demonstrates its occurrence during a recognition memory task, and suggests that such dysfunction may be related to deficits in reward-related processing.

In addition to demonstrating reduced VS activation to correct recognition, schizophrenia patients also failed to demonstrate the increase in VS responses with more conservative bias that was found in controls. This occurred despite a similar average and range of response bias in the patient and control groups. This suggests a failure to generate prediction errors in response to changes in expectation and confidence, and is consistent with substantial evidence of blunted prediction error responses in schizophrenia (Corlett et al., 2007; Murray et al., 2008; Waltz et al., 2009; Romaniuk et al., 2010; Gradin et al., 2011; Morris et al., 2011).

Evidence of blunted VS reward responses in schizophrenia, particularly in those with prominent negative symptoms (Juckel et al., 2006; Wolf et al., 2008; Waltz et al., 2009, 2010), leads us to speculate that the abnormal VS responses seen here reflect reductions in intrinsic motivation and self-generated reward signals. Deficits in intrinsic motivation have been found in schizophrenia and related to clinically relevant outcomes. In particular, Nakagami et al. (2008) demonstrated that intrinsic motivation, neurocognitive performance, and psychosocial functioning are all inter-related in schizophrenia, and that intrinsic motivation mediated the relationship between neurocognition and psychosocial functioning. Similarly, Choi and Medalia (2009) have shown that an intervention aimed at increasing intrinsic motivation in schizophrenia patients improved performance in a difficult cognitive task. Reward and motivation impairments may therefore contribute to cognitive abnormalities in schizophrenia. Our results extend this literature, and suggest VS dysfunction may be a key neural mechanism.

ALTERNATIVE INTERPRETATIONS AND LIMITATIONS

While our findings are consistent with a prediction error framework, certain alternative interpretations and limitations should be acknowledged. First, anterior caudate responses have been observed during successful response inhibition and response switching (Vink et al., 2005; Li et al., 2008; Cameron et al., 2009). These responses have been interpreted as evidence that anterior caudate is important for overcoming prepotent responses. One can conceptualize "old" responses in the setting of a difficult memory task and a conservative response bias as requiring inhibition of the prepotent "new" response. This response inhibition theory would also predict stronger VS responses in those with greater conservative bias. However, it is unlikely that the response inhibition account fully explains VS activation to correct recognition responses, as we have observed stronger VS activation to correct than incorrect responses in the absence of feedback in a facial recognition memory paradigm (Wolf et al., 2011) that utilizes simultaneous target and foil presentations and therefore does not involve response inhibition (unpublished VS ROI analysis).

Second, while the presence of apparent reward prediction signals in the absence of external rewards or explicit feedback is one of the most interesting aspects of the study, this design is also a limitation as it does not directly vary reward outcomes. Future studies can bolster such interpretations by including additional task conditions that directly manipulate rewards, feedback, and task difficulty. In addition, future research should also assess relevant subjective states and traits not measured in the current study, including pre-response confidence, post-response confidence, subjective value of correct and incorrect responses, trait reward sensitivity, and intrinsic motivation.

Third, the number of correct targets available for fMRI analysis was relatively small (∼15 on average in each group), which may have decreased the reliability of within-subject activation estimates. However, this within-subject variance is accounted for in the imaging analysis, and while it may reduce the statistical significance of our results, it is unlikely to explain them. Regardless, future studies would benefit from the increased statistical power afforded by a greater number of trials, as well as a larger number of subjects.

Fourth, group differences in task performance present a potential confound. As in this study, schizophrenia patients' performance in cognitive tasks is typically impaired. If VS prediction error responses are more robust in control subjects when task difficulty increases, there is a tension between achieving adequate patient performance and capturing the phenomenon of interest. However, groups did not differ in accuracy or reaction time for the target recognition condition that was the focus of this study. Furthermore, discrimination accuracy did not correlate with VS responses within either group and inclusion of P_r as a confound covariate across groups did not attenuate the group

difference. Therefore, group differences in recognition performance are an unlikely explanation for observed VS differences.

Finally, our patient population was treated with antipsychotic medications, which are known to affect dopamine signaling in brain regions including VS, and which may affect reward processing, motivation, and cognition in complex ways. The observation of striatal hypofunction in unaffected family members as well as patients (Vink et al., 2006), suggests a vulnerability phenotype rather than an effect of frank illness or medication. Furthermore, given that most schizophrenia patients require long-term treatment with antipsychotics, a potential role of medication effects on VS hypofunction does not negate its importance. Nonetheless, clarifying the role of medications will be important and will require further studies including in drug-naïve populations.

CONCLUSION

Taking into account the above limitations, these findings add significantly to our understanding of normal VS responses during successful retrieval, and dysfunction in these responses in schizophrenia. Examining VS responses during cognitive tasks from the perspective of reward prediction errors can yield new approaches to understanding cognitive–emotional interactions that are critical to healthy functioning as well as psychopathological states. As cognitive deficits and motivational deficits are both disabling and treatment-resistant symptoms of schizophrenia, further investigation of VS dysfunction should prove illuminating and may yield novel approaches to pharmacological and psychological intervention.

REFERENCES

Achim, A. M., and Lepage, M. (2005). Episodic memory-related activation in schizophrenia: meta-analysis. *Br. J. Psychiatry* 187, 500–509.

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., and Gabrieli, J. D. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* 50, 507–517.

Andreasen, N. C. (1984a). *Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City: University of Iowa.

Andreasen, N. C. (1984b). *Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City: University of Iowa.

Barch, D. M. (2005). The relationships among cognition, motivation, and emotion in schizophrenia: how much and how little we know. *Schizophr. Bull.* 31, 875–881.

Bunzeck, N., Doeller, C. F., Dolan, R. J., and Duzel, E. (2011). Contextual interaction between novelty and reward processing within the mesolimbic system. *Hum. Brain Mapp.* doi: 10.1002/hbm.21288.

Cameron, I. G., Coe, B. C., Watanabe, M., Stroman, P. W., and Munoz, D. P. (2009). Role of the basal ganglia in switching a planned response. *Eur. J. Neurosci.* 29, 2413–2425.

Choi, J., and Medalia, A. (2009). Intrinsic motivation and learning in a schizophrenia spectrum sample. *Schizophr. Res.* 118, 12–19.

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R., Shanks, D. R., Robbins, T. W., Bullmore, E. T., Dickinson, A., and Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain* 130, 2387–2400.

Daniel, R., and Pollmann, S. (2010). Comparing the neural basis of monetary reward and cognitive feedback during information-integration category learning. *J. Neurosci.* 30, 47–55.

D'Cruz, A. M., Ragozzino, M. E., Mosconi, M. W., Pavuluri, M. N., and Sweeney, J. A. (2011). Human reversal learning under conditions of certain versus uncertain outcomes. *Neuroimage* 56, 315–322.

Dere, E., Pause, B. M., and Pietrowsky, R. (2010). Emotion and episodic memory in neuropsychiatric disorders. *Behav. Brain Res.* 215, 162–171.

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., and Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7716–7720.

First, M. B., Spitzer, R. L., Gibbon, M., and Williams, J. B. W. (2001). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders – Patient Edition*. New York: Biometrics Research, New York State Psychiatric Institute.

Foussias, G., and Remington, G. (2008). Negative symptoms in schizophrenia: avolition and Occam's razor. *Schizophr. Bull.* 36, 359–369.

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., Hall, J., and Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain* 134, 1751–1764.

Gur, R. E., Nimgaonkar, V. L., Almasy, L., Calkins, M. E., Ragland, J. D., Pogue-Geile, M. F., Kanes, S., Blangero, J., and Gur, R. C. (2007). Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. *Am. J. Psychiatry* 164, 813–819.

Han, S., Huettel, S. A., Raposo, A., Adcock, R. A., and Dobbins, I. G. (2010). Functional significance of striatal responses during episodic decisions: recovery or goal attainment? *J. Neurosci.* 30, 4767–4775.

Heinrichs, R. W., and Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* 12, 426–445.

Iidaka, T., Matsumoto, A., Nogawa, J., Yamamoto, Y., and Sadato, N. (2006). Frontoparietal network involved in successful retrieval from episodic memory. Spatial and temporal analyses using fMRI and ERP. *Cereb. Cortex* 16, 1349–1360.

Juckel, G., Schlagenhauf, F., Koslowski, M., Wustenberg, T., Villringer, A., Knutson, B., Wrase, J., and Heinz, A. (2006). Dysfunction of ventral striatal reward prediction in schizophrenia. *Neuroimage* 29, 409–416.

Kirsch, P., Schienle, A., Stark, R., Sammer, G., Blecker, C., Walter, B., Ott, U., Burkart, J., and Vaitl, D. (2003). Anticipation of reward in a nonaversive differential conditioning paradigm and the brain reward system: an event-related fMRI study. *Neuroimage* 20, 1086–1095.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., and Glover, G. (2005). Distributed neural representation of expected value. *J. Neurosci.* 25, 4806–4812.

Koch, K., Wagner, G., Nenadic, I., Schachtzabel, C., Schultz, C., Roebel, M., Reichenbach, J. R., Sauer, H., and Schlosser, R. G. (2008). Fronto-striatal hypoactivation during correct information retrieval in patients with schizophrenia: an fMRI study. *Neuroscience* 153, 54–62.

Leavitt, V. M., and Goldberg, T. E. (2009). Episodic memory in schizophrenia. *Neuropsychol. Rev.* 19, 312–323.

Li, C. S., Yan, P., Sinha, R., and Lee, T. W. (2008). Subcortical processes of motor response inhibition during a stop signal task. *Neuroimage* 41, 1352–1363.

McClure, S. M., Berns, G. S., and Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38, 339–346.

McClure, S. M., York, M. K., and Montague, P. R. (2004). The neural substrates of reward processing in humans: the modern role of fMRI. *Neuroscientist* 10, 260–268.

McDermott, K. B., Szpunar, K. K., and Christ, S. E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia* 47, 2290–2298.

Morris, R. W., Vercammen, A., Lenroot, R., Moore, L., Langton, J. M., Short, B., Kulkarni, J., Curtis, J., O'Donnell, M., Weickert, C. S., and Weickert,

T. W. (2011). Disambiguating ventral striatum fMRI-related bold signal during reward prediction in schizophrenia. *Mol. Psychiatry*. doi: 10.1038/mp.2011.75.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., Jones, P. B., Bullmore, E. T., Robbins, T. W., and Fletcher, P. C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol. Psychiatry* 13, 239, 267–276.

Murty, V. P., Labar, K. S., Hamilton, D. A., and Adcock, R. A. (2011). Is all motivation good for learning? Dissociable influences of approach and avoidance motivation in declarative memory. *Learn. Mem.* 18, 712–717.

Nakagami, E., Xie, B., Hoe, M., and Brekke, J. S. (2008). Intrinsic motivation, neurocognition and psychosocial functioning in schizophrenia: testing mediator and moderator effects. *Schizophr. Res.* 105, 95–104.

Phelps, E. A. (2006). Emotion and cognition: insights from studies of the human amygdala. *Annu. Rev. Psychol.* 57, 27–53.

Ragland, J. D., Gur, R. C., Valdez, J., Turetsky, B. I., Elliott, M., Kohler, C., Siegel, S., Kanes, S., and Gur, R. E. (2004). Event-related fMRI of frontotemporal activity during word encoding and recognition in schizophrenia. *Am. J. Psychiatry* 161, 1004–1015.

Ragland, J. D., Laird, A. R., Ranganath, C., Blumenfeld, R. S., Gonzales, S. M., and Glahn, D. C. (2009). Prefrontal activation deficits during episodic memory in schizophrenia. *Am. J. Psychiatry* 166, 863–874.

Robbins, T. W., and Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Curr. Opin. Neurobiol.* 6, 228–236.

Rodriguez, P. F., Aron, A. R., and Poldrack, R. A. (2006). Ventral-striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. *Hum. Brain Mapp.* 27, 306–313.

Romaniuk, L., Honey, G. D., King, J. R., Whalley, H. C., Mcintosh, A. M., Levita, L., Hughes, M., Johnstone, E. C., Day, M., Lawrie, S. M., and Hall, J. (2010). Midbrain activation during Pavlovian conditioning and delusional symptoms in schizophrenia. *Arch. Gen. Psychiatry* 67, 1246–1254.

Satterthwaite, T. D., Wolf, D. H., Gur, R. C., Ruparel, K., Valdez, J. N., Gur, R. E., and Loughead, J. (2009). Frontolimbic responses to emotional face memory: the neural correlates of first impressions. *Hum. Brain Mapp.* 30, 3748–3758.

Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Valdez, J. N., Siegel, S. J., Kohler, C. G., Gur, R. E., and Gur, R. C. (2010). Association of enhanced limbic response to threat with decreased cortical facial recognition memory response in schizophrenia. *Am. J. Psychiatry* 167, 418–426.

Saykin, A. J., Gur, R. C., Gur, R. E., Mozley, P. D., Mozley, L. H., Resnick, S. M., Kester, D. B., and Stafiniak, P. (1991). Neuropsychological function in schizophrenia. Selective impairment in memory and learning. *Arch. Gen. Psychiatry* 48, 618–624.

Schultz, W. (2010). Dopamine signals for reward and risk: basic and recent data. *Behav. Brain Funct.* 6, 24.

Sergerie, K., Lepage, M., and Armony, J. L. (2007). Influence of emotional expression on memory recognition bias: a functional magnetic resonance imaging study. *Biol. Psychiatry* 62, 1126–1133.

Sescousse, G., Redoute, J., and Dreher, J. C. (2010). The architecture of reward value coding in the human orbitofrontal cortex. *J. Neurosci.* 30, 13095–13104.

Snodgrass, J. G., and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J. Exp. Psychol. Gen.* 117, 34–50.

Spaniol, J., Davidson, P. S., Kim, A. S., Han, H., Moscovitch, M., and Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia* 47, 1765–1779.

Tricomi, E., and Fiez, J. A. (2008). Feedback signals in the caudate reflect goal achievement on a declarative memory task. *Neuroimage* 41, 1154–1167.

Vink, M., Kahn, R. S., Raemaekers, M., Van Den Heuvel, M., Boersma, M., and Ramsey, N. F. (2005). Function of striatum beyond inhibition and execution of motor responses. *Hum. Brain Mapp.* 25, 336–344.

Vink, M., Ramsey, N. F., Raemaekers, M., and Kahn, R. S. (2006). Striatal dysfunction in schizophrenia and unaffected relatives. *Biol. Psychiatry* 60, 32–39.

von Zerssen, G. C., Mecklinger, A., Opitz, B., and Von Cramon, D. Y. (2001). Conscious recollection and illusory recognition: an event-related fMRI study. *Eur. J. Neurosci.* 13, 2148–2156.

Vul, E., Harris, C., Winkielman, P., and Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect. Psychol. Sci.* 4, 247–290.

Waltz, J. A., Schweitzer, J. B., Gold, J. M., Kurup, P. K., Ross, T. J., Salmeron, B. J., Rose, E. J., McClure, S. M., and Stein, E. A. (2009). Patients with schizophrenia have a reduced neural response to both unpredictable and predictable primary reinforcers. *Neuropsychopharmacology* 34, 1567–1577.

Waltz, J. A., Schweitzer, J. B., Ross, T. J., Kurup, P. K., Salmeron, B. J., Rose, E. J., Gold, J. M., and Stein, E. A. (2010). Abnormal responses to monetary outcomes in cortex, but not in the basal ganglia, in schizophrenia. *Neuropsychopharmacology* 35, 2427–2439.

Weiss, A. P., and Heckers, S. (2001). Neuroimaging of declarative memory in schizophrenia. *Scand. J. Psychol.* 42, 239–250.

Wittmann, B. C., Schiltz, K., Boehler, C. N., and Duzel, E. (2008). Mesolimbic interaction of emotional valence and reward improves memory formation. *Neuropsychologia* 46, 1000–1008.

Wolf, D. H. (2006). Anhedonia in schizophrenia. *Curr. Psychiatry Rep.* 8, 322–328.

Wolf, D. H., Satterthwaite, T. D., Loughead, J., Pinkham, A., Overton, E., Elliott, M. A., Dent, G. W., Smith, M. A., Gur, R. C., and Gur, R. E. (2011). Amygdala abnormalities in first-degree relatives of individuals with schizophrenia unmasked by benzodiazepine challenge. *Psychopharmacology (Berl.)* 218, 503–512.

Wolf, D. H., Turetsky, B. I., Loughead, J., Elliott, M. A., Pratiwadi, R., Gur, R. E., and Gur, R. C. (2008). Auditory oddball fMRI in schizophrenia: association of negative symptoms with regional hypoactivation to novel distractors. *Brain Imaging Behav.* 2, 132–145.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.