Striatal Associative Learning Signals Are Tuned to In-groups

Katherine E. Powers[1,2], Leah H. Somerville[2], William M. Kelley[1], and Todd F. Heatherton[1]

Abstract

■ An important feature of adaptive social behavior is the ability to flexibly modify future actions based on the successes or failures of past experiences. The ventral striatum (VS) occupies a central role in shaping behavior by using feedback to evaluate actions and guide learning. The current studies tested whether feedback indicating the need to update social knowledge would engage the VS, thereby facilitating subsequent learning. We also examined the sensitivity of these striatal signals to the value associated with social group membership. Across two fMRI studies, participants answered questions testing their knowledge about the preferences of personally relevant social groups who were high (in-group) or low (out-group) in social value. Participants received feedback indicating whether their responses were correct or incorrect on a trial-by-trial basis. After scanning, participants were given a surprise memory test examining memory for the different types of feedback. VS activity in response to social feedback correlated with subsequent memory, specifying a role for the VS in encoding and updating social knowledge. This effect was more robust in response to in-group than out-group feedback, indicating that the VS tracks variations in social value. These results provide novel evidence of a neurobiological mechanism adaptively tuned to the motivational relevance of the surrounding social environment that focuses learning efforts on the most valuable social outcomes and triggers adjustments in behavior when necessary. ■

INTRODUCTION

Successfully navigating complex social environments and sustaining meaningful interpersonal relationships depends, in part, on our ability to remember and represent social knowledge according to the successes and failures of past experiences. From an evolutionary perspective, members of one's social in-group represent the greatest potential for affiliation and fulfillment of belongingness needs (Van Bavel, Swencionis, O'Connor, & Cunningham, 2012; Heatherton, 2011). Because of these important social and psychological benefits, understanding the intentions and beliefs of one's social in-group is especially valuable, and the ability to maintain accurate representations of in-group members is essential for harmonious social functioning (Van Bavel, Packer, & Cunningham, 2008). Here, we leverage knowledge of the neural systems involved in facilitating nonsocial associative learning to test the sensitivity of these feedback-based learning signals to the value associated with social groups.

The ability to learn contingencies between actions and positive or negative outcomes and flexibly adjust behavior is supported by a distributed, interconnected network of cortical and subcortical brain regions (Shohamy, 2011; Gläscher, Daw, Dayan, & O'Doherty, 2010; Doll, Jacobs, Sanfey, & Frank, 2009; Behrens, Hunt, Woolrich, & Rushworth, 2008; Rushworth, Behrens, Rudebeck, & Walton, 2007; Schultz, Dayan, & Montague, 1997). A wealth of neuroscientific findings from both human and animal models have shown that feedback-guided learning depends critically on the ventral striatum (VS), a subcortical region anatomically positioned to function as a key integration site for evaluating the motivational properties of external stimuli and processing feedback-related information to modify learning (Haber & Knutson, 2010; Delgado, 2007; Kelley, 2004). Neuroimaging studies in humans have revealed that the striatum responds robustly to learning tasks that provide explicit feedback about performance (Poldrack et al., 2001) and further that the BOLD response in the striatum is differentially sensitive to positive and negative feedback (Delgado, 2007; Hariri et al., 2006; King-Casas et al., 2005; Elliott, Frith, & Dolan, 1997), typically exhibiting preferentially greater activation following rewarding or positive feedback and activation decreasing below baseline following punishing or negative feedback (Delgado, Nystrom, Fissell, Noll, & Fiez, 2000).

Computational learning theories (e.g., Sutton & Barto, 1998) argue that the striatal response is particularly robust when feedback does not align with predictions (e.g., prediction errors) and that this sensitivity to errors underlies effective learning. Importantly, the magnitude of VS signaling tracks with perceived value, with larger responses generated for more valued outcomes (Knutson, Taylor, Kaufman, Peterson, & Glover, 2005; Nieuwenhuis et al., 2005). Thus, striatal signals facilitate learning from value-based feedback to help build and maintain updated representations of profitable actions (Delgado, 2007;

[1]Dartmouth College, [2]Harvard University

Schultz, 2007). Originally conceptualized in terms of non-social rewards (e.g., juice; Berns, McClure, Pagnoni, & Montague, 2001), the functionality of these striatal learning signals has recently been extended to some social contexts, specifically learning to associate peers with different probabilities of social acceptance (Lin, Adolphs, & Rangel, 2012; Jones et al., 2011) and learning who can be trusted (King-Casas et al., 2005).

As reward-seeking behaviors are shaped by changes in the motivational value of environmental stimuli (Schultz, 2002), demonstrating that the VS is sensitive to different levels of social value would provide important evidence of how the brain maintains accurate representations of knowledge about valued social groups and optimally focuses learning efforts on the most meaningful social outcomes. In Experiment 1, participants underwent fMRI scanning while answering questions testing their knowledge of the preferences and behaviors of a social in-group. Participants received feedback indicating whether their responses were correct or incorrect on a trial-by-trial basis. We hypothesized that feedback indicating the need to update social knowledge (i.e., incorrect feedback) would generate robust learning signals in the VS. In Experiment 2, we tested the sensitivity of this response to variations in social value by asking participants to make judgments about a personally relevant social in-group (i.e., high social value) and out-group (i.e., low social value). We hypothesized that feedback indicating the need to update in-group relevant knowledge would generate stronger learning signals relative to the out-group. We also tested whether individual differences in desire for social acceptance would modulate these neural responses. Finally, we administered a post-scan memory test to examine potential linkages between VS activity and learning outcomes.

METHODS

Participants

Experiment 1

Thirty-three participants (18 women, age range = 18–23 years) were recruited for Experiment 1. All participants were right-handed, had normal or corrected-to-normal visual acuity, and presented no fMRI contraindications. To ensure familiarity with the content of the study questions, all participants were Dartmouth undergraduate students of at least sophomore standing. They received course credit or were paid for their participation and gave informed consent in accordance with the guidelines set by the Committee for the Protection of Human Subjects at Dartmouth College.

Experiment 2

Twenty participants (*n* = 19 final sample; age range = 18–21) were recruited for Experiment 2. Because the social group manipulation capitalized on real-world sorority affiliations, all participants in this experiment belonged to a campus sorority and thus were female. As in Experiment 1, all participants were Dartmouth undergraduates of at least sophomore standing, were right-handed, had normal or corrected-to-normal visual acuity, and presented no fMRI contraindications. Participants were paid for their participation and gave informed consent in accordance with the guidelines set by the Committee for the Protection of Human Subjects at Dartmouth College.

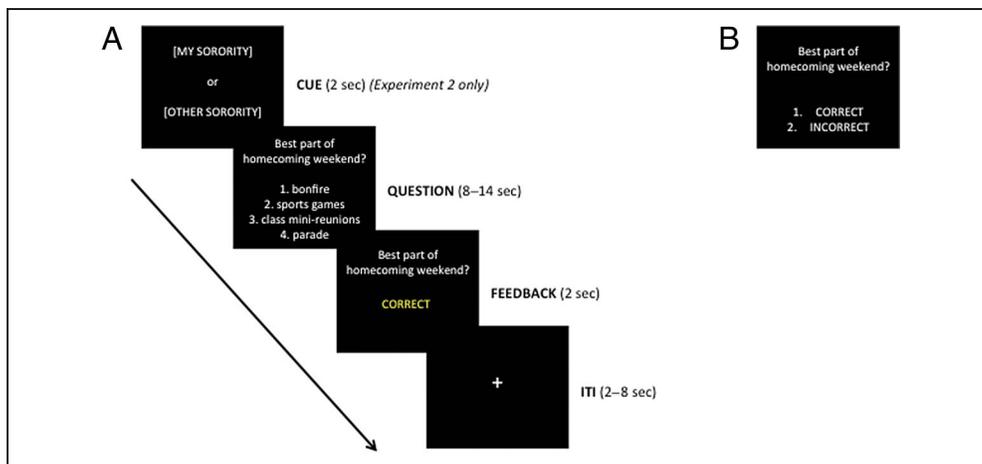### Experimental Design and Procedure

Overview

In both experiments, participants underwent fMRI scanning while answering questions probing their knowledge of the beliefs and preferences of various social groups at Dartmouth College. Participants were told that a survey containing questions about undergraduate student life had been administered to several social groups on campus. Once in the scanner, participants viewed each question and selected the answer they believed had been chosen by the majority of a specified social group; that is, they were instructed to choose the consensus answer, even if this differed from a personally held opinion. After making a response, participants were presented with feedback—purportedly based on responses collected from the mass survey—indicating whether they had accurately selected the consensus choice of the group (correct feedback) or not (incorrect feedback). Postexperiment debriefing confirmed that all participants believed this cover story with the exception of one participant from Experiment 2, who was subsequently dropped from all analyses.

Experiment 1

Each trial consisted of a multiple-choice question followed by feedback about the correctness of their response (see Figure 1A). Specifically, a question appeared on the screen (8–14 sec), accompanied by four possible answer choices. Participants were instructed to answer each question based on what they believed a representative sample of 100 other Dartmouth students would say. Because all participants were Dartmouth students, this task involved predicting the consensus beliefs and preferences of a salient social in-group to which participants had a powerful motivation to belong. Participants indicated their responses by pressing one of four buttons, and then the trial then shifted to the feedback phase (2 sec) in which participants received feedback indicating whether they had answered that question correctly or incorrectly. In this way, correct feedback confirmed participants' existing knowledge about their peers, whereas incorrect feedback signaled that representations were inaccurate. In actuality, the feedback was randomized and predetermined, such that participants received correct feedback on half of all trials and incorrect feedback on the remaining half. In Experiment 1, a total

Figure 1. Task parameters. (A) Schematic representation of a complete trial from Experiment 2. Time indicates the duration of each subcomponent. Experiment 1 did not include a CUE phase but otherwise employed the same trial structure and timing. (B) Example trial from the postscan surprise memory test, which was identical across both experiments.

of 88 questions (44 correct feedback, 44 incorrect feedback) were split evenly across four functional runs, and the order of the questions was pseudorandomized and counterbalanced across participants.

Experiment 2

In Experiment 2, we tested the sensitivity of striatal activity to variations in social value, specifically to the differential value associated with consensus knowledge about a social in-group compared with a social out-group. On the basis of prior work that has illustrated that the existing social group infrastructure on college campuses provides a natural context in which to examine social group attitudes and affiliations (Dunn, Wilson, & Gilbert, 2003), we recruited members of two campus sororities to participate in this experiment. To avoid introducing the potential confound of gender, a salient social group construct along which people readily identify, we chose to only test sorority members and not members of both sororities and fraternities. Participants were prescreened for MRI compatibility as well as levels of affiliation and identification with their respective sorority (i.e., their social in-group) and a selected sorority to which they did not belong (i.e., their social out-group). Participants reporting extremely low levels of identification with their own sorority (i.e., a response of 1 on a 1 [*not at all*] to 5 [*extremely*] scale in response to the question, "How much is membership in your sorority a part of your identity?") were excluded from participation.

The experimental protocol of Experiment 2 (see Figure 1A) was nearly identical to that of Experiment 1, except for the addition of a cue phase at the beginning of each trial (2 sec) specifying which social group a given question pertained to. For half of the questions, participants were instructed to select the response they believed a representative sample of members of their own sorority (i.e., social in-group; high social value) would choose. For the remaining questions, participants were instructed to consider the beliefs and preferences of members of the target out-group sorority (i.e., low social

value). The social group manipulation set up a 2 × 2 factorial design crossing social group (in/out-group) with feedback (correct/incorrect), resulting in 28 trials per category that were split evenly across seven functional runs. Half of the participants were presented with 29 incorrect in-group trials and 27 incorrect out-group trials because of a coding error. Association of social group and question/ feedback pairings was counterbalanced, and the presentation order of the questions was pseudorandomized and counterbalanced across participants.

In both experiments, all trials were randomly intermixed with periods of fixation during which participants passively viewed a white cross-hair at the center of the screen (2–8 sec). The temporal jitter introduced by varying the duration of the question phase (8–14 sec), as well as the variable fixation periods, allowed unique estimates of the hemodynamic response function to be computed for each subcomponent of the trial.

Postscan Behavioral Assessments

Following the fMRI session, participants in both experiments underwent a surprise memory test to determine how well they remembered the feedback they were given during scanning. Each question from the scanning session was presented individually, and participants were instructed to recall whether they had previously answered that question correctly or incorrectly (see Figure 1B). The memory test was self-paced, and no feedback was provided about the accuracy of their responses.

Participants in Experiment 2 also completed an exit questionnaire to assess levels of affiliation with each social group, which served as a manipulation check for the validity of the selected out-group. This measure consisted of an adaptation of Aron, Aron, and Smollan's (1992) Inclusion of Other in Self Scale for intergroup relations (see also Schubert & Otten, 2002). Specifically, participants were presented with a series of increasingly overlapping circles, ranging from zero overlap to almost completely overlapping, and asked to indicate which graphical representation most accurately represented their level of identification

with their in-group and out-group. Finally, Experiment 2 participants also filled out the Need to Belong Scale (Leary, Kelly, Cottrell, & Schreindorfer, 2013), a 10-item questionnaire designed to assess the strength of an individual's motivation to be accepted by others.

fMRI Acquisition and Parameters

All participants were scanned on a Phillips Intera Achieve 3-T scanner (Amsterdam, The Netherlands) at Dartmouth College using a 32-channel phase arrayed coil. Structural images were acquired using a T1-weighted MPRAGE protocol (160 sagittal slices, repetition time = 9.9 msec, echo time = 4.6 msec, 8° flip angle, 1 × 1 × 1 mm voxels). Functional images were acquired using a T2*-weighted echo-planar sequence (repetition time = 2000 msec, echo time = 35 msec, 90° flip angle, field of view = 24 cm). In Experiment 1, data were collected in four functional runs (35 axial slices per volume, 3 mm thick, 0.5 mm gap, 3 × 3 mm in-plane resolution), each consisting of 206 whole-brain volumes. In Experiment 2, seven functional runs, each consisting of 170 whole-brain volumes, were collected using the same parameters. A Panasonic PT-4000 DLP projector (Osaka, Japan) was used to project stimuli onto a screen at the end of the magnet bore that participants viewed via an angled mirror mounted on the head coil.

Neuroimaging data were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK) in conjunction with a suite of neuroimaging analysis tools (https://github.com/ddwagner/SPM8w). Functional data were preprocessed to remove sources of noise and artifact and corrected for differences in slice acquisition time. Images were realigned within and across functional runs to correct for head movement and then unwarped to reduce residual movement-related image distortions not corrected by realignment. Functional data were then normalized into standard space (3 mm isotropic voxels) based on the SPM8 EPI template that conforms to the ICBM 152 brain template (Montreal Neurological Institute) and spatially smoothed using a 6-mm FWHM Gaussian kernel. In Experiment 1, one functional run was discarded from one participant because of excessive movement (>2 mm) and one functional run was discarded from another participant because of excessive artifact and noise in the imaging data. In Experiment 2, three functional runs were discarded from one participant because of excessive movement (>2 mm) and one functional run was discarded from another participant because of scanner malfunction.

fMRI Analysis

Experiment 1

To estimate neural responses specific to type of feedback, a general linear model (GLM) incorporating task effects (question, correct feedback, incorrect feedback) and covariates of noninterest (a session mean, a linear trend, and six

movement parameters derived from realignment corrections) was specified for each participant. The extensive temporal jitter included in the task design served to reduce colinearity between trial components and allowed for distinct phases of each trial to be uniquely estimated. Additional nuisance regressors were included for one participant who exhibited an isolated movement of more than 3.5 mm (one regressor per affected volume and additional regressors for the two volumes preceding and following the movement). This procedure serves to isolate the affected time points from estimates of task effects. Each task effect regressor was convolved with a canonical hemodynamic response function and used to generate contrast images (weighted parameter estimates) for each participant. Given our focus on neural responses to feedback that would prompt the need to update social knowledge (i.e., incorrect feedback), these contrast images were then entered into a second-level random effects analysis comparing brain regions displaying greater activity in response to incorrect feedback than correct feedback. Monte Carlo simulations using AFNI's AlphaSim were used to calculate the minimum cluster size at an uncorrected threshold of $p < .005$ required for whole-brain correction of $p < .05$. These simulations ($n = 1000$) were performed using smoothness estimated from the residuals obtained from the GLM and resulted in a critical cluster size of 82 contiguous voxels.

Experiment 2

A GLM incorporating seven task regressors (cue, question [in-group], question [out-group], and four types of feedback [correct in-group, correct out-group, incorrect in-group, and incorrect out-group] and covariates of noninterest (a session mean, a linear trend, and six movement parameters derived from realignment corrections) was specified for each participant. Additional nuisance regressors were included for one participant who exhibited an isolated movement of more than 3 mm (one regressor per affected volume and additional regressors for the two volumes preceding and following the movement). Again, each regressor was convolved with a canonical hemodynamic response function and used to generate contrast images for each participant, and these contrast images were then entered into a second-level random effects analysis comparing brain regions sensitive to incorrect feedback relative to correct feedback. Alphasim Monte Carlo simulations (performed as described in Experiment 1, at an uncorrected threshold of $p < .005$) estimated the minimum cluster size necessary to achieve whole-brain correction of $p < .05$ at 59 contiguous voxels. The slight difference in contiguity thresholds across studies can be attributed to small variations in acquisition and analysis parameters as described above.

To examine how VS activity varied as a function of social group affiliation, an ROI analysis was performed on parameter estimates extracted by centering bilateral 4-mm spheres on the voxels of peak VS deactivation

identified by the incorrect > correct feedback contrast. These parameter estimates were submitted to an offline repeated-measures ANOVA, with Feedback (correct, incorrect) and Social group (in-group, out-group) both entered as within-subject factors. ROIs were thus defined in an unbiased manner with respect to Social group.

Brain-subsequent Memory Correlations (Experiments 1 and 2)

To investigate whether VS activity facilitated encoding of socially relevant feedback, we correlated performance on the postscan memory test with VS parameter estimates extracted during the ROI analysis described above. Although we initially planned to conduct an item-specific subsequent memory analysis to assess this relationship, the high degree of accurate recall present in both experiments resulted in too few forgotten trials per category to permit this type of analysis. Thus, for each participant, the percentage of correctly remembered items was calculated from the postscan memory test, collapsed across feedback type (i.e., items associated with both correct and incorrect feedback). These scores were then corre-

lated offline with parameter estimates of VS activity. As the VS ROIs were defined based on feedback-related signal, the correlation with postscan memory test performance is independent of the ROI selection criteria.

Brain–Individual Difference Correlations (Experiment 2)

To test whether this striatal response is magnified in individuals with a heightened motivation for social acceptance, we performed another ROI analysis (4-mm sphere), this time isolating VS responses to incorrect feedback during in-group trials (relative to out-group trials) and submitting these parameter estimates to an offline correlation with individual scores on the Need to Belong Scale.

RESULTS
Experiment 1
BOLD Response in the VS
Given our a priori hypotheses that the need to update socially relevant knowledge would generate robust learning
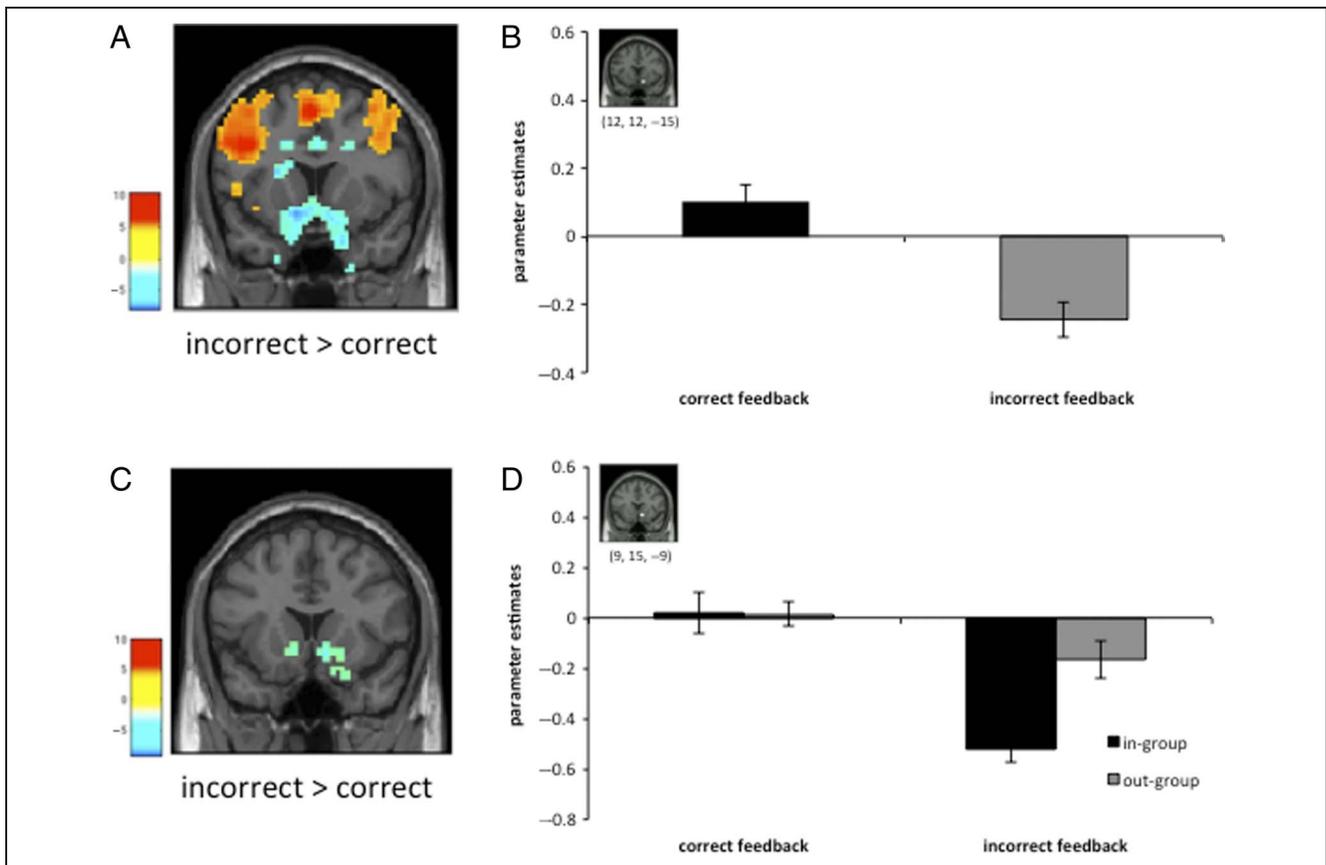

Figure 2. fMRI results revealing feedback-based learning signals in the VS and modulation by social group. A whole-brain, random-effects analysis ($p < .05$, corrected) comparing brain regions displaying greater activity when receiving incorrect feedback relative to correct feedback revealed robust deactivation of the VS in both Experiment 1 (A) and Experiment 2 (C). In Experiment 1 (B), incorrect feedback produced a robust striatal deactivation relative to fixation baseline. This bar plot is for display purposes only and is not intended to represent effect sizes. In Experiment 2 (D), feedback indicating judgments about in-group members were incorrect elicited a greater striatal deactivation relative to fixation baseline in the right VS compared with out-group members (interaction $p = .018$). Insets display locations of VS ROIs. Error bars depict SE calculated for within-subject designs.
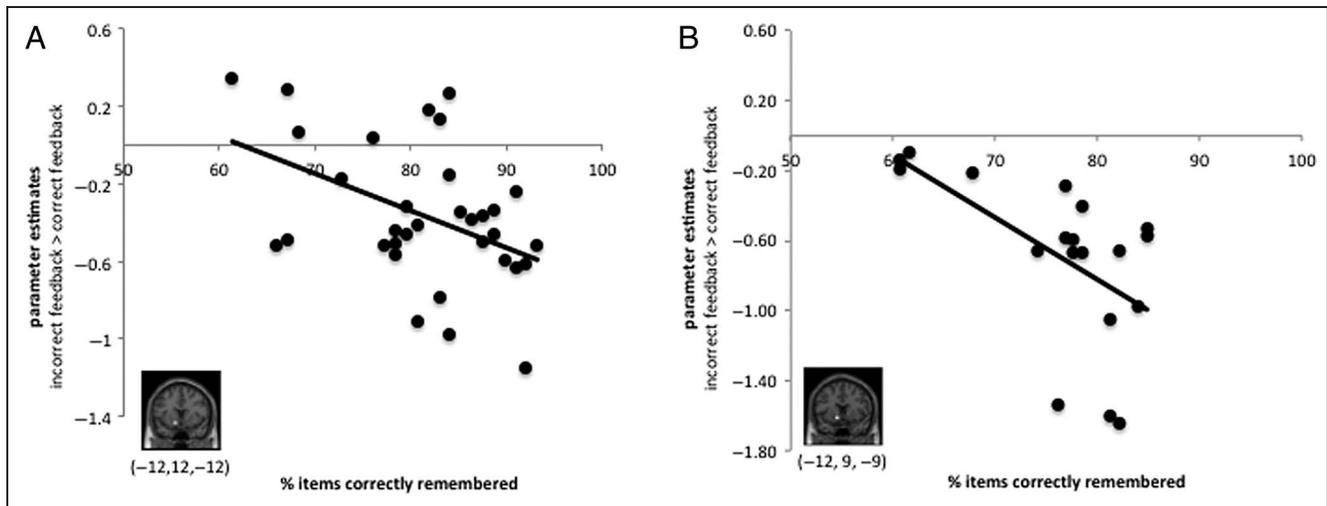
Figure 3. Scatterplots depicting the relationship between VS response and subsequent memory for feedback. The magnitude of the VS response during feedback predicted subsequent memory in both Experiment 1 (A; left VS, shown: $r = -0.45, p = .009$) and in Experiment 2 (B; left VS, shown: $r = -0.58, p = .009$).

signals in the VS, our primary analysis examined regional hemodynamic responses in the VS on trials in which participants were incorrect in their judgments about the choices of in-group members compared with those in which they made correct judgments. Incorrect judgments about in-group members elicited a robust deactivation in the VS bilaterally relative to correct feedback (peak of left VS cluster: −12, 12, −12; peak of right VS cluster 12, 12, −15; see Figure 2A). In an effort to visualize the specific pattern of responses in the VS during these trials relative to fixation baseline, parameter estimates of VS activity were extracted from a 4-mm sphere centered on the peak activation in both the left and right VS (see Figure 2B). Inspection of this response reveals a substantial deactivation in striatal activity relative to baseline produced by receiving incorrect feedback about social in-group members, which is consistent with prior work linking negative feedback with a suppression in the striatal BOLD response in nonsocial contexts (Delgado et al., 2000).

Postscan Memory Test and Correlations with VS Response

Participants were able to successfully remember which questions they had gotten correct and incorrect during the scanning session, accurately recalling feedback they had received for 81.3% (*SD* = 8.4) of the total trials presented. On average participants were more accurate for correct feedback trials (*M* = 86.4%, *SD* = 8.0) compared with incorrect feedback trials (*M* = 76.2%, *SD* = 11.5), *t*(32) = 5.51, *p* < .001. Participants were also faster to recall trials on which they had received correct feedback (*M* = 2112.7 msec, *SD* = 298.3) compared with incorrect feedback (*M* = 2237.9 msec, *SD* = 308.6), *t*(32) = 3.9, *p* < .001.

A composite score was calculated for each participant to capture memory for socially relevant feedback, consist-

ing of the percentage of correctly remembered items (both correct and incorrect feedback) from the postscan memory test. These scores were then correlated with parameter estimates of VS activity extracted from the incorrect > correct feedback contrast. This correlation analysis revealed that the magnitude of VS signaling during feedback, which is primarily characterized by deactivation to incorrect feedback, was associated with better memory for the feedback in both the left (*r* = −0.45, *p* = .009) and right VS (*r* = −0.52, *p* = .002; see Figure 3A).

Supplemental Whole-brain Analyses

To supplement these targeted analyses of our a priori VS ROIs and facilitate cross-study comparison, we also report results from whole-brain, random effects analyses (*p* < .05, corrected) comparing responses to the different types of feedback. Brain regions demonstrating greater activity for incorrect feedback relative to correct feedback include the dorsal ACC (−3, 36, 42; *t* = 10.36; 1242 voxels), the left (−33, 24, −9; *t* = 9.12; 1854 voxels) and right (33, 24, −9; *t* = 6.69; 322 voxels) inferior frontal gyrus, two regions of the middle frontal gyrus (54, 30, 24; *t* = 6.50; 641 voxels and 33, 57, 6; *t* = 6.07; 120 voxels), and two regions of the middle temporal gyrus (−54, −24, −12; *t* = 5.22; 107 voxels and 66, −33, −18; *t* = 4.71; 111 voxels). Brain regions displaying greater activity to correct feedback compared with incorrect feedback include the medial pFC (3, 54, −9; *t* = 4.43; 117 voxels) and a large, confluent cluster encompassing the parahippocampal gyrus, paracentral lobules, and superior occipital gyrus (peak: 36, −87, 24; *t* = 7.95; 16,164 voxels). As described in the ROI analysis above, incorrect feedback deactivated the VS bilaterally (i.e., more activity for correct feedback relative to incorrect feedback); at this threshold, the local peaks of VS activity (left peak: −12,

12, −12; $t = 6.19$ and right peak: 12, 12, −15; $t = 4.83$) were subsumed under the larger cluster involving the parahippocampal gyrus. We note that similar patterns of activity in these regions have been shown in visual working memory tasks comparing BOLD responses to correct and incorrect trials (Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002) and error detection and behavioral monitoring (e.g., Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004).

Experiment 2

Ratings of Self-group Overlap

Analysis of the exit questionnaire measuring the degree to which participants identified with each social group revealed that self in-group overlap ($M = 5.2$, $SD = 1.2$) was significantly higher than self out-group overlap ($M = 1.9$, $SD = 1.1$), $t(18) = 9.9$, $p < .001$, confirming validity of the selected out-group.[1]

BOLD Response in the VS

Replicating Experiment 1, trials in which participants were incorrect about their judgments of other Dartmouth students, regardless of social group, were accompanied by reduced activity in the VS bilaterally relative to correct judgments (peak of left VS cluster: −12, 9, −9; peak of right VS cluster: 9, 15, −9). Again, incorrect feedback produced a robust deactivation relative to baseline (see Figure 2C).

Critically, this deactivation was more pronounced for in-group trials (group by feedback interaction, $F(1, 18) = 6.8$, $p = .018$; right VS). Specifically, feedback indicating that judgments about in-group members were incorrect produced a significantly greater deactivation in the VS than feedback indicating incorrect judgments about out-group members (pairwise $p = .001$; see Figure 2D). Social group did not modulate the VS response to correct feedback (pairwise $p = .97$). The greater deactivation in response to incorrect in-group feedback was also reflected in a main effect of social group, such that the VS exhibited greater deactivation for in-group trials relative to out-group trials, $F(1, 18) = 7.7$, $p = .012$. In the left VS, these effects did not reach significance (ME social group: $F(1, 18) = 1.0$, $p = .33$; interaction: $F(1, 18) = 2.2$, $p = .15$).

Postscan Memory Test and Correlations with VS Response

Participants accurately recalled the feedback they had received for 76.2% ($SD = 7.9$) of the total trials presented. Replicating results from Experiment 1, participants were more accurate for correct feedback trials ($M = 81.6\%$, $SD = 9.4$) compared with incorrect feedback trials ($M =$ 70.8%, $SD = 8.6$), $F(1, 18) = 28.4$, $p < .001$. Participants also recalled more items relating to their in-group ($M = 78.5\%$, $SD = 8.8$) than those relating to the out-group ($M = 73.8\%$, $SD = 8.7$), $F(1, 18) = 8.0$, $p = .011$, replicating the well-established finding that people preferentially encode cues relevant to their in-group (e.g., Van Bavel et al., 2012; Bernstein, Young, & Hugenberg, 2007; Meissner & Brigham, 2001). The Feedback type × Social group interaction was not significant ($p = .23$). Also replicating results from Experiment 1, participants were faster to recognize trials on which they had received correct feedback ($M = 2220.9$ msec, $SD = 554.6$) compared with incorrect feedback ($M = 2386.4$ msec, $SD = 532.6$), $F(1, 18) = 4.9$, $p = .04$. There was no difference in RTs as a function of Social group ($p = .56$) or interaction ($p = .48$).

We performed the same analysis correlating individual estimates of VS activity with subsequent memory for the feedback described in Experiment 1. VS activity tracked with memory performance in the left VS, $r(19) = −0.58$, $p = .009$ (see Figure 3B), providing converging evidence of the role of the VS in encoding feedback related to social group knowledge. Although observed bilaterally in Experiment 1, we note that this relationship was not significant in the right VS in Experiment 2 ($p = .87$).

To test whether this relationship varied as a function of social group, we conducted the analysis correlating VS activity and subsequent memory for in-group relevant trials and out-group relevant trials. In the left VS, there was a significant correlation between VS activity specific to in-group trials and memory for in-group relevant feedback, $r = −0.46$, $p = .05$, whereas for out-group trials, there was no relationship between VS activity and memory for feedback ($p = .36$). Direct comparison of these correlation coefficients did not reach significance (Pearson–Filon test of independent correlations, $p = .49$). In the right VS, there was no VS–memory relationship observed for either the in-group trials ($p = .45$) or the out-group trials ($p = .74$). The fact that these in-group effects were specific to the left VS may be attributable to the fact that, in Experiment 2, the effects of the broader memory analysis containing all trials were also left-lateralized.

Individual Differences in Social Belongingness Needs and Correlations with VS Activity

Scores on the Need to Belong Scale ranged from 24 to 48 (possible range: 10–50, $M = 36.5$), with higher scores signaling greater motivation for social acceptance. For this analysis, we compared differences in the magnitude of VS deactivation in response to incorrect feedback about in-group members relative to incorrect feedback about out-group members. Individuals higher in the need to belong demonstrated more robust deactivations in the left VS when wrong about their in-group, $r = −0.46$, $p = .045$ (see Figure 4). The same inverse relationship
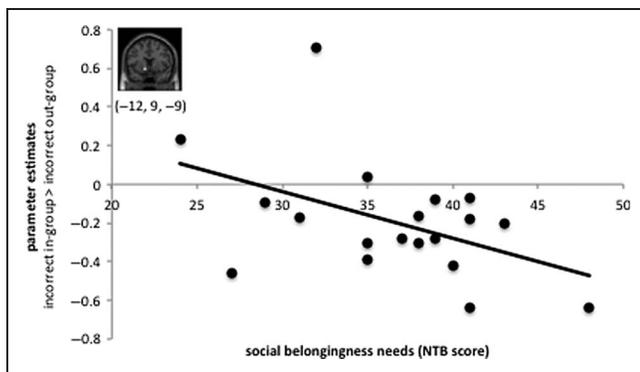
Figure 4. Scatterplot depicting the relationship between VS response and individual differences in social belongingness needs. In Experiment 2, individuals with high belongingness needs exhibit the largest decreases in VS activity when wrong about their in-group (left VS, shown: $r = -0.46$, $p = .045$).

was also exhibited by the right VS, although it did not reach significance ($r = -0.23$, $p = .34$).

Supplemental Whole-brain Analyses

Brain regions displaying greater activity for incorrect compared with correct feedback trials at this threshold include the caudate ($-6, 0, 24$; $t = 6.24$; 107 voxels). Examination of trials that elicited greater activity for correct feedback compared with incorrect feedback included the medial pFC ($12, 48, -9$; $t = 7.55$; 65 voxels), visual cortex ($-6, -105, 12$; $t = 9.04$; 3004 voxels), fusiform extending into the parahippocampus ($-30, -6, -33$; $t = 6.78$; 80 voxels), and a region at the intersection of the posterior cingulate cortex and precuneus ($0, -36, 42$; $t = 4.50$; 74 voxels). As described in the ROI analysis above, incorrect feedback deactivated (i.e., more activity for correct feedback relative to incorrect feedback) the VS bilaterally (left peak: $-12, 9, -9$; $t = 6.3$; 93 voxels; right peak: $9, 15, -9$, $t = 5.51$; 95 voxels).

DISCUSSION

The ability to represent accurate knowledge about valued social groups is essential for sustaining harmonious intergroup relations. The current experiments leveraged knowledge of the specific neural mechanisms involved in associative learning and updating value representations in nonsocial domains to determine how the brain maintains accurate representations of knowledge about social groups. We found that feedback-driven learning signals in the VS were modulated by social value, with stronger signals generated to promote learning about more highly valued social groups.

Research has shown that the VS is critical for successful feedback-based learning, playing an important role in evaluating current rewards and guiding future reward-maximizing behaviors by attaching value to specific actions (Glimcher, 2011; Delgado, 2007; Schultz, 2000). Here, we show that the same depression in striatal activity observed during negative performance-related feedback in nonsocial contexts (e.g., Delgado et al., 2000) occurs when people realize their assumptions about what valued peers think are incorrect. These findings provide evidence that the brain regions that subserve reward-related learning are sensitive to feedback that calls for social group knowledge to be revised (Schultz & Dickinson, 2000).

Critically, we found that social group membership modulated the strength of these striatal learning signals, such that incorrect judgments about in-group relative to out-group members produced a greater striatal deactivation. Moreover, those individuals especially concerned with being accepted by others and gaining social approval (i.e., those high in the need to belong) exhibited increased striatal responsivity in the face of errors about their social in-group. From an evolutionary perspective, there is clear adaptive value in being able to accurately predict the behaviors and judgments of one's in-group, as these individuals present the greatest potential for social affiliation and have the potential to most drastically affect our well-being (Van Bavel et al., 2012; Heatherton, 2011). Indeed, behavioral research has shown that enhanced importance of in-group members is reflected in increased accuracy in judgments concerning their beliefs and preferences (Mitchell, Macrae, & Banaji, 2006) as well as a boost in memory for the faces of in-group members (Van Bavel et al., 2012). Here we show that feedback indicating the need to update knowledge about a social in-group generated stronger learning signals relative to the out-group, providing evidence that the dopaminergic system tracks the value associated with different social groups. Taken together, these results provide evidence that social value biases recruitment of brain regions that support reward learning. This sensitivity to social value may help sustain important social relationships by signaling situations that are most in need of social adjustment and focusing learning efforts accordingly.

Consistent with prior work documenting contributions of the dopaminergic system to memory formation (Wimmer & Shohamy, 2011; Wittmann, Schiltz, Boehler, & Düzel, 2008; Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Wittmann et al., 2005), the strength of the striatal response in the current experiments correlated with individual differences in performance on the postscan memory test. Specifically, those individuals who demonstrated the most robust striatal response, primarily characterized by deactivations provoked by incorrect feedback, also demonstrated the best memory for the feedback. Furthermore, participants generally displayed better memory for feedback about their in-group, and the relationship between striatal activity and subsequent memory was more pronounced when the feedback pertained to a social in-group. This bias for in-group relevant cues has been well documented

in social psychological research (e.g., Van Bavel et al., 2012; Bernstein et al., 2007; Meissner & Brigham, 2001) and is thought to promote cohesion within valued groups. Results from the memory test also showed that participants were slightly better at remembering correct feedback, although incorrect feedback was remembered quite successfully overall. Prior work has indicated that people fail to process self-threatening feedback as deeply as self-affirming feedback (see Green, Sedikides, & Gregg, 2008), raising the possibility that incorrect feedback may constitute a threat to the self in that it directly questions the accuracy of one's understanding of their social environment and thus is somewhat less deeply encoded than correct feedback. Altogether, these findings identify a role for striatally mediated learning processes in intergroup contexts, extending the current understanding of how striatal signals build and update memories (Wimmer & Shohamy, 2012; Adcock et al., 2006) and drive evolutionarily adaptive behavior (Wise, 2004).

A growing number of prior studies have noted evidence of striatal involvement in social learning contexts, including when attractiveness ratings differ from consensus judgments (Zaki, Schirmer, & Mitchell, 2011; Klucharev et al., 2009), when assessments of an individual's competence conflict with group stereotypes (Harris & Fiske, 2010), and associating probabilities of receiving positive social feedback from peers (Lin et al., 2012; Jones et al., 2011; see also Ruff & Fehr, 2014). The present findings distinctly contribute to prior work by demonstrating the capacity of the VS to code discrete variations in social value and to integrate these signals into lasting memory traces that may shape future behavior. More broadly, these findings align nicely with reports of nonsocial contextual modulation of striatal activity (De Martino, Kumaran, Holt, & Dolan, 2009) as well as with recent work by Bhanji and Delgado (2014), showing the importance of feedback-related striatal signals in driving task persistence in the face of setbacks. Future work can investigate how these motivational signals overlap and differ across social and nonsocial domains.

The patterns and localization of VS activity reported here are consistent with neurocomputational accounts of striatal function in animal and human models quantifying VS activity in response to events that do not align with predictions (e.g., reward prediction errors; Floresco, 2015; Glimcher, 2011; O'Doherty, Deichmann, Critchley, & Dolan, 2002; Pagnoni, Zink, Montague, & Berns, 2002; Berns et al., 2001). Recent neuroimaging work has revealed a close correspondence between reward prediction error signaling and brain activity in the VS, specifically showing that the BOLD signal increases and decreases to events that are better or worse than expected, respectively (Niv, Edlund, Dayan, & O'Doherty, 2012). Thus, one interpretation of our findings is that the VS occupies a critical role in registering errors in the predictions of expected social feedback and uses this information to optimally focus learning. This interpretation suggests that individuals may use prediction error learning to dynamically update representations about valued social groups. Such a conceptualization aligns nicely with recent computational neuroimaging studies, showing that people utilize prediction error learning mechanisms to update social impressions of relationship partners (Fareri, Chang, & Delgado, 2012; Chang, Doll, van't Wout, Frank, & Sanfey, 2010).

In contrast to the robust striatal response to incorrect feedback, correct feedback generally yielded minimal differences in these experiments. One possible explanation for this result, which is consistent with prediction error learning accounts of striatal function (Ruff & Fehr, 2014; Delgado, 2007), is that, because correct feedback was consistent with participants' existing knowledge, no learning signals were generated because no updating was required. Although some other studies have reported increased VS activity in response to positive task outcomes (Fareri, Chang, & Delgado, 2015; Fareri, Niznikiewics, Lee, & Delgado 2012; Izuma, Saito, & Sadato, 2008), one key difference between these studies and the present task is the degree to which positive outcomes are expected by the participants. In the experiments reported here, the receipt of correct feedback should not have surprised participants, as they were instructed to make judgments reflecting what they thought to be true. On the other hand, the positive outcomes delivered in these other tasks (e.g., correctly guessing a card's value; Fareri et al., 2012) were not necessarily expected by the participants and sometimes even occurred randomly. Positive outcomes of an unexpected nature do typically elicit VS activity, just as these studies report.

We also observed no difference in the striatal response to correct feedback as a function of social group in Experiment 2. It is possible that the expectation of being correct was powerful enough to minimize potential differences produced by this manipulation. Another possibility is that, to some degree, participants still retained commonalities with members of the target social out-group (i.e., all are Dartmouth students), which produced enhanced confidence in predictions for both groups than might be observed with a more distant social out-group. Even so, that we did observe a striatal response to incorrect feedback yoked to social group membership in spite of this overarching group affiliation further underscores the sensitivity of this system to the value associated with different social outcomes.

By providing a mechanistic account of VS function during learning about social groups and tracking associated value, our findings further emphasize the critical role of this brain region in motivated learning across different contexts (see Shohamy, 2011). Despite this centralized focus, it should be noted that the VS contributes to a distributed neural network that learns to evaluate rewards and selects actions based on their predicted value. Interactions between reward learning circuits and brain regions that mediate other aspects of cognition, affect, and motor action are thought to underlie the formation

of adaptive behaviors, and the anatomical position of the VS facilitates interfacing between these circuits (see Haber & Knutson, 2010). Consistent with this broader framework, results from our whole-brain analyses demonstrate increased signaling during feedback delivery in other brain regions known to participate in associative learning, including the caudate (Li, Schiller, Schoenbaum, Phelps, & Daw, 2011; Shohamy, 2011) and the ACC (Behrens et al., 2008; Rushworth & Behrens, 2008; Passingham, Toni, & Rushworth, 2000). These results extend our primary striatal findings to implicate the distributed mechanisms that support learning cue–outcome associations in nonsocial domains in learning about the social world.

Limitations

One limitation of these experiments is that the task design is not optimized for computational modeling of neural prediction error signals on a trial-by-trial basis. Prediction error-based learning models are designed to track the updating of discrete representations of stimuli/outcome contingencies in a trial-by-trial manner over the course of a task, whereas each trial in the present task contained unique information that did not bear directly on subsequent trials. Nevertheless, the pattern and localization of results closely resembles patterns of fundamental prediction error signaling observed in neuroimaging studies adopting this type of computational approach in both social (Jones et al., 2011) and nonsocial (Niv et al., 2012) contexts. It will be important for future work to adopt computational modeling approaches to investigate how these social value signals guide real-world behaviors aimed at promoting group cohesion. A second methodological limitation is that participants were not provided with the "correct" answer on incorrect feedback trials—they were just told that the answer was wrong. In this way, incorrect feedback indicated which representations needed revision and provided participants with some degree of information to refine their existing representations (e.g., by revising the answer they thought was correct). Although these findings demonstrate that the receipt of incorrect feedback engages fundamental social learning processes, this aspect of the design precludes the ability to test how the content of representations was revised at a more item-specific level. Future work will benefit from developing experimental paradigms that track the learning process for specific social contingencies and testing whether certain representations are more resistant to updating than others.

Conclusions

Converging theories from social and evolutionary psychology underscore the physiological and psychological benefits of belonging to social groups. These findings suggest a key functional role for the VS in signaling when revisions to social group knowledge are necessary and demonstrate that associative learning brain systems are adaptively tuned to the motivational relevance of the surrounding social environment. Taken together with prior work, these patterns of striatal activity and memorial biases point to a common neural mechanism underlying feedback-based learning across multiple domains of rewards that subserves the ability to flexibly modify future behavioral decisions based on the successes or failures of past experiences. From an evolutionary perspective, the sensitivity of this system to the fundamental drive for social belonging is adaptively beneficial, as it enables individuals to effectively navigate complex social environments and maximizes power to predict socially profitable outcomes.

Note

1. Because of potential concerns about the level of measurement of this scale, a nonparametric analysis was also performed, which revealed similar levels of significance (Wilcoxon signed-rank test, $p < .001$).

REFERENCES

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

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63,* 596–612.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature, 456,* 245–249.

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *Journal of Neuroscience, 21,* 2793–2798.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18,* 706–712.

Bhanji, J. P., & Delgado, M. R. (2014). Perceived control influences neural responses to setbacks and promotes persistence. *Neuron, 83,* 1369–1375.

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology, 61,* 87–105.

De Martino, B., Kumaran, D., Holt, B., & Dolan, R. J. (2009). The neurobiology of reference-dependent value computation. *Journal of Neuroscience, 29,* 3833–3842.

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences, 1104,* 70–88.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology, 84,* 3072–3077.

Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Research, 1299,* 74–94.

Dunn, E. W., Wilson, T. D., & Gilbert, D. T. (2003). Location, location, location: The misprediction of satisfaction in housing lotteries. *Personality and Social Psychology Bulletin, 29,* 1421–1432.

Elliott, R., Frith, C. D., & Dolan, R. J. (1997). Differential neural response to positive and negative feedback in planning and guessing tasks. *Neuropsychologia, 35,* 1395–1404.

Fareri, D. S., Chang, L. J., & Delgado, M. R. (2012). Effects of direct social experience on trust decisions and neural reward circuitry. *Frontiers in Neuroscience, 6,* 1–17.

Fareri, D. S., Chang, L. J., & Delgado, M. R. (2015). Computational substrates of social value in interpersonal collaboration. *Journal of Neuroscience, 35,* 8170–8180.

Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., & Delgado, M. R. (2012). Social network modulation of reward-related signals. *Journal of Neuroscience, 32,* 9045–9052.

Floresco, S. B. (2015). The nucleus accumbens: An interface between cognition, emotion, and action. *Annual Review of Psychology, 66,* 25–52.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66,* 585–595.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences, U.S.A., 108*(Suppl. 3), 15647–15654.

Green, J. D., Sedikides, C., & Gregg, A. P. (2008). Forgotten but not gone: The recall and recognition of self-threatening memories. *Journal of Experimental Social Psychology, 44,* 547–561.

Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology, 35,* 4–26.

Hariri, A. R., Brown, S. M., Williamson, D. E., Flory, J. D., de Wit, H., & Manuck, S. B. (2006). Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *Journal of Neuroscience, 26,* 13213–13217.

Harris, L. T., & Fiske, S. T. (2010). Neural regions that underlie reinforcement learning are also active for social expectancy violations. *Social Neuroscience, 5,* 76–91.

Heatherton, T. F. (2011). Neuroscience of self and self-regulation. *Annual Review of Psychology, 62,* 363–390.

Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron, 58,* 284–294.

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., et al. (2011). Behavioral and neural properties of social reinforcement learning. *Journal of Neuroscience, 31,* 13039–13045.

Kelley, A. E. (2004). Ventral striatal control of appetitive motivation: Role in ingestive behavior and reward-related learning. *Neuroscience & Biobehavioral Reviews, 27,* 765–776.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308,* 78–83.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron, 61,* 140–151.

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

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct validity of the need to belong scale: Mapping the nomological network. *Journal of Personality Assessment, 95,* 610–624.

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience, 14,* 1250–1252.

Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience, 7,* 274–281.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7,* 3–35.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50,* 655–663.

Nieuwenhuis, S., Heslenfeld, D. J., von Geusau, N. J. A., Mars, R. B., Holroyd, C. B., & Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage, 25,* 1302–1309.

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience, 32,* 551–562.

O'Doherty, J. P., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron, 33,* 815–826.

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience, 5,* 97–98.

Passingham, R. E., Toni, I., Rushworth, M. F. (2000). Specialisation within the prefrontal cortex: The ventral prefrontal cortex and associative learning. In W. X. Schneider, A. M. Owen, & J. Duncan (Eds.), *Executive control and the frontal lobe: Current issues* (pp. 103–113). Berlin, Heidelberg: Springer.

Pessoa, L., Gutierrez, E., Bandettini, P. A., & Ungerleider, L. G. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron, 35,* 975–987.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature, 414,* 546–550.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306,* 443–447.

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience, 15,* 549–562.

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience, 11,* 389–397.

Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and

orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences, 11,* 168–176.

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity, 1,* 353–376.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience, 1,* 199–207.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron, 36,* 241–263.

Schultz, W. (2007). Behavioral dopamine signals. *Trends in Neurosciences, 30,* 203–210.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience, 23,* 473–500.

Shohamy, D. (2011). Learning and motivation in the human striatum. *Current Opinion in Neurobiology, 21,* 408–414.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1*,* No. 1). Cambridge, MA: MIT Press.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. *Psychological Science, 19,* 1131–1139.

Van Bavel, J. J., Swencionis, J. K., O'Connor, R. C., & Cunningham, W. A. (2012). Motivated social memory:

Belonging needs moderate the own-group bias in face recognition. *Journal of Experimental Social Psychology, 48,* 707–713.

Wimmer, G. E., & Shohamy, D. (2011). The striatum and beyond: Contributions of the hippocampus to decision making. In M. R. Delgado, E. A. Phelps, & T. W. Robbins (Eds.), *Decision making, affect and learning: Attention and performance XXII* (pp. 281–309). New York: Oxford University Press.

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science, 338,* 270–273.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience, 5,* 483–494.

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

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J., Heinze, H., & Duzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45,* 459–467.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science, 22,* 894–900.