

Characterizing the Associative Content of Brain Structures Involved in Habitual and Goal-Directed Actions in Humans: A Multivariate fMRI Study

Daniel McNamee,^{1,2*} Mimi Liljeholm,^{1,2,3*} Ondrej Zika,^{1,2} and John P. O'Doherty^{1,2}

¹Computation and Neural Systems Program and ²Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125 and ³Department of Cognitive Sciences, University of California, Irvine, Irvine, California 92697

While there is accumulating evidence for the existence of distinct neural systems supporting goal-directed and habitual action selection in the mammalian brain, much less is known about the nature of the information being processed in these different brain regions. Associative learning theory predicts that brain systems involved in habitual control, such as the dorsolateral striatum, should contain stimulus and response information only, but not outcome information, while regions involved in goal-directed action, such as ventromedial and dorsolateral prefrontal cortex and dorsomedial striatum, should be involved in processing information about outcomes as well as stimuli and responses. To test this prediction, human participants underwent fMRI while engaging in a binary choice task designed to enable the separate identification of these different representations with a multivariate classification analysis approach. Consistent with our predictions, the dorsolateral striatum contained information about responses but not outcomes at the time of an initial stimulus, while the regions implicated in goal-directed action selection contained information about both responses and outcomes. These findings suggest that differential contributions of these regions to habitual and goal-directed behavioral control may depend in part on basic differences in the type of information that these regions have access to at the time of decision making.

Key words: decision making; fMRI; MVPA; goal-directed; habitual

Introduction

Two distinct strategies support behavioral control: a goal-directed strategy that flexibly generates decisions based on deliberate evaluation of the consequences of actions and a habitual strategy relying on a reflexive, automatic elicitation of actions (Dickinson, 1985; Balleine and Dickinson, 1998; Balleine et al., 2008). These distinct mechanisms depend on at least partly dissociable brain systems, with the posterior dorsolateral striatum (DLS) implicated in habits (Yin et al., 2004; Tricomi et al., 2009) and the ventromedial prefrontal cortex (vmPFC) or its rodent homolog, and the dorsomedial striatum (DMS), contributing to goal-directed control (Balleine and Dickinson, 1998; Killcross and Coutureau, 2003; Yin et al., 2005; Tanaka et al., 2008). However, the nature of the information encoding in these regions is much less understood.

According to associative learning theory, in habits, associations are formed between stimuli (S) and responses (R), without

any encoding of the goal or outcome (O). In contrast, in goal-directed learning, associations are formed between stimuli, responses, and outcomes (Balleine and Dickinson, 1998). Specifically, goal representations are suggested to be elicited via S-O associations, which in turn retrieve response representations via an O-R association (Balleine and Ostlund, 2007; de Wit and Dickinson, 2009).

Several neurophysiology studies have explored whether brain regions implicated in habitual and goal-directed action contain different types of information. Most of those studies have reported similar information encoding in both DMS and DLS (Kim et al., 2009, 2013; Kimchi et al., 2009; Stalnaker et al., 2010), although one study did find a change in the level of activity in DMS and DLS as animals transitioned from goal-directed to habitual control (Gremel and Costa, 2013). Another approach has been to correlate computational learning models to fMRI or neurophysiology data (O'Doherty et al., 2003; Samejima et al., 2005). Some studies found different types of reinforcement learning (model free vs model based) correlate with activity in DLS compared with DMS, vmPFC, and dorsolateral prefrontal cortex (dlPFC; Hampton et al., 2006; Gläscher et al., 2010; Wunderlich et al., 2012; Lee et al., 2014) while others report more mixed representations (Daw et al., 2011; Simon and Daw, 2011). However, while such analyses reveal the computations that might be operating in a given area, they do not illuminate the type of information encoded in those regions upon which a particular computational process may act.

Received Nov. 12, 2014; revised Jan. 12, 2015; accepted Jan. 16, 2015.

Author contributions: D.M., M.L., and J.P.O. designed research; M.L. and O.Z. performed research; D.M., M.L., and O.Z. analyzed data; D.M., M.L., and J.P.O. wrote the paper.

This work was funded by National Institutes of Health (NIH) Grant DA033077-01 (supported by OppNet, NIH's Basic Behavioral and Social Science Opportunity Network) to J.P.O.

The authors declare no competing financial interests.

*D.M. and M.L. contributed equally to this work.

Correspondence should be addressed to Daniel McNamee, Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA 91125. E-mail: dmcnamee@caltech.edu.

DOI:10.1523/JNEUROSCI.4677-14.2015

Copyright © 2015 the authors 0270-6474/15/353764-08\$15.00/0

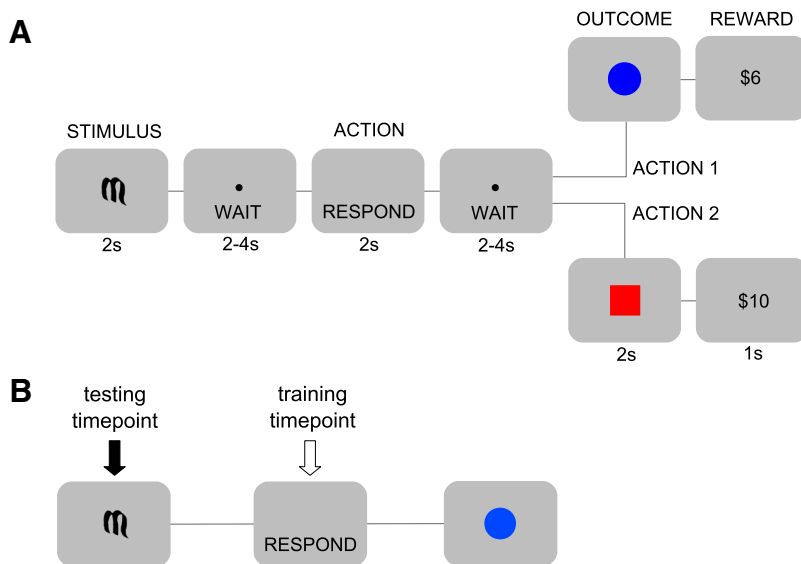


Figure 1. Task design and time-span decoding. **A**, Subjects performed a binary decision task. One of two possible initial stimuli (STIMULUS) was presented, which determined the subsequent deterministic action-outcome contingencies between two possible actions and two possible outcome states. Outcome states were denoted by either a blue circle or a red square and were followed, after a short delay, by one of two distributions of monetary rewards (large or small). Crucially, each possible combination of stimulus, action, and outcome was permuted across conditions, thereby ensuring that representations of these different decision variables were not conflated (see Material and Methods). **B**, These analyses aimed to assess the classification of events that were being associatively retrieved by the initial stimulus. The classifier was trained on the relevant events (e.g., actions) at the time of their execution, but tested on those same events at the time of the initial stimulus (see Materials and Methods for more details).

In the present study, human participants underwent fMRI while performing a binary decision task, in which we carefully manipulated response and outcome identity across experimental sessions. Using a multivariate pattern analysis, we tested for the presence of response and outcome information at the time of stimulus and action performance in different brain regions. We hypothesized that brain regions implicated in habitual control, such as the DLS, would encode response, but not outcome, information indicating a role for this region in supporting stimulus–response associations, while other brain regions such as the vmPFC, dlPFC, and anterior DMS would contain representations of both responses and outcomes, indicative of a role for those regions in goal-directed learning and control.

Materials and Methods

Participants

Nineteen healthy, right-handed volunteers participated in this study (11 male; mean age 22.9; SD 4.1 years). The volunteers were pre-assessed to exclude those with a history of neurological or psychiatric illness. All participants gave informed consent and the study was approved by the Institutional Review Board of the California Institute of Technology.

Task

Participants performed a simple binary decision task (Fig. 1A). At the start of each trial an initial stimulus environment was indicated by one of two Sanskrit characters. The participant then performed one of two possible actions and subsequently entered an outcome state in which they received an associated reward. There were two distinct initial stimuli, two possible outcome states (represented by a blue circle and a red square, respectively), and two reward distributions: high (equal probability of \$8, \$10, and \$12) and low (equal probability of \$2, \$4, and \$6). Participants interacted with the environment using two qualitatively different actions: a double button press and a trackball roll. These actions were performed on the same device using the right hand. Action outcomes were anticorrelated and deterministic such that, within a session, a particular action always led to a given outcome, and that outcome always

yielded a particular reward distribution. The initial stimulus determined the subsequent action-outcome contingencies and thus indicated which of the two actions was highly rewarded on a given trial.

Critically, the relationships between initial stimuli, actions, outcomes, and reward distributions were permuted across four conditions, ensuring that representations of one decision variable (e.g., actions) could not be confounded with those of another (e.g., initial stimuli; ϕ correlation coefficient $< |0.005|$ for all pairs of variables). Thus, if in one condition, the button press produced the blue circle followed by a high reward, in another condition the tracker ball produced the blue circle followed by a high reward, and so on. Since, across the four conditions, each decision variable was paired equally often with each level of every other decision variable, and since subsequent classification was performed across conditions, these permutations ensure that each decision variable is dissociated from all others. A description of the full set of permutations is provided in Table 1. The order of the four conditions was counterbalanced across subjects.

At the beginning of the experiment, before any scanning, participants were given detailed instructions about the task procedure and trial structure. Specifically, they were told that each trial would begin with the presentation of an abstract symbol, which would be followed by a brief “wait” pause, in turn followed by a prompt to respond, using either of the two actions, another wait pause, a picture of a colored shape and, finally, the reward delivered on that trial. The wait pauses, shown in Figure 1, were included to ensure the separation of intertrial events. Participants were further instructed that the relationships between initial stimuli, actions, outcomes, and rewards would change throughout the task, and that they would be notified whenever those changes occurred. Each scanning condition consisted of three consecutive sessions separated by 2 min breaks during which the scanner was turned off (to reduce inter-run variability in image intensity), with 16 instrumental choice trials per session (eight for each initial stimulus, randomly interleaved). At the beginning of each session, participants were instructed either that the relationships between initial stimuli, actions, outcomes, and rewards would be the same in the upcoming session as in the previous (within condition) or that these relationships had changed and would have to be relearned (between conditions). Finally, participants were told to use their knowledge of instrumental contingencies to maximize their monetary gain. Due to the simple nature of the task, participants rapidly achieved a high level of performance in each condition, selecting the action that yielded the high reward distribution on the vast majority of trials (see Results).

fMRI data acquisition

Functional imaging was performed with a 3 T Siemens Trio scanner. Forty-five contiguous interleaved transversal slices of echo-planar T2*-weighted images were acquired in each volume, with a slice thickness of 3 mm and no gap (repetition time, 2650 ms; echo time, 30 ms; flip angle, 90°; field of view, 192 mm²; matrix, 64 × 64). Slice orientation was tilted 30° from a line connecting the anterior and posterior commissure. This slice tilt alleviates the signal drop in the orbitofrontal cortex (OFC; Deichmann et al., 2003). We discarded the first three images before data processing and statistical analysis to compensate for the T1 saturation effects. A whole-brain, high-resolution T1-weighted structural scan (voxel size: 1 × 1 × 1 mm³) was also acquired for each subject.

Data preprocessing and filtering

Slice timing correction, motion correction, and spatial normalization were applied to the data. Before multivoxel sample extraction, low-

Table 1. Illustration of the experimental conditions describing the permutation across sessions

Condition	Stimulus	Action	Outcome	Reward
1	S1	A1	Red	High
	S1	A2	Blue	Low
	S2	A1	Red	Low
	S2	A2	Blue	High
2	S1	A1	Blue	Low
	S1	A2	Red	High
	S2	A1	Blue	High
	S2	A2	Red	Low
3	S1	A1	Blue	High
	S1	A2	Red	Low
	S2	A1	Blue	Low
	S2	A2	Red	High
4	S1	A1	Red	Low
	S1	A2	Blue	High
	S2	A1	Red	High
	S2	A2	Blue	Low

S1, S2, stimuli; A1, A2, button press and trackerball actions. Note that the order of presentation of the experimental conditions was also permuted across participants.

frequency components ($<1/120$ Hz), serial autocorrelations, and head motion were subtracted from the data. To correct for session-related mean and scaling effects, we applied second-order detrending and z scoring on a per voxel per session basis (Pereira et al., 2009). Preprocessing and filtering were performed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>), except detrending and z scoring for which the PyMVPA package was used (Hanke et al., 2009).

General linear model

Eight regressors of interest were included in the general linear model (GLM). Each regressor corresponded to the identity of a particular decision variable (i.e., one regressor for each initial stimulus, each action, each outcome, and each reward distribution). In addition, parametric modulators reflecting the actual reward delivered on a given trial were added to the reward-distribution regressors. Time series of head motion estimated during realignment were included as covariates of no interest.

Classification algorithm

We used a Gaussian Naive Bayes (GNB) classification algorithm (Mitchell, 1997) with an assumption of zero covariance across voxels. To perform binary classification our algorithm first estimates mean activity vectors and covariance matrices from training data for the Gaussian distributions $p(x|A)$ and $p(x|B)$. Then, the algorithm assigns a test sample x_{test} to the condition with the maximum posterior probability at x_{test} based on the estimated distributions: if $p(x_{\text{test}}|A) > p(x_{\text{test}}|B)$ the algorithm infers that x_{test} was sampled under condition A. Generalization accuracy is estimated using cross-validation. This involves training and testing on mutually exclusive subsets of samples and repeating with a different partitioning on each “fold.” Cross-validation was done on a leave-one-session-out basis. On every fold, the classifier was trained on three sessions and tested on the remaining session, thereby avoiding session-related dependencies between training and testing samples (Mitchell, 1997; Kriegeskorte et al., 2009; Pereira et al., 2009). Accuracy scores were averaged to give the generalization accuracy. All preprocessing and filtering was performed on a per session basis. Importantly, the average Spearman correlation between combinations of decision variables across subjects was 2×10^{-5} , 1×10^{-3} , and 2×10^{-4} , respectively, indicating that the classifier could not erroneously decode one decision variable based on correlated representations of another.

Multivariate pattern analysis

A searchlight procedure (Kriegeskorte et al., 2006; Kahnt et al., 2010) provided a spatially unbiased estimator of distributed activity across the brain. This involved the performance of GNB classification on the fMRI data in spheres of voxels of radius 3 throughout the brain. We extracted a sample of fMRI data corresponding to the initial stimulus, action, and outcome time points in each trial (with a shift of 5 s to account for

hemodynamic delay) by averaging the two volumes closest in time (one before and one after) to the relevant time point (Clithero et al., 2011; McNamee et al., 2013). Each fMRI data sample had two task-related characteristics: time point and identity. Our hypotheses required us to decode based on a variety of interactions between these two characteristics, which we detail below. In particular, we performed the following analyses to determine whether neural representations of decision variables are present at time points in a trial other than the moment of perception or action. For all analyses below, cross-fold validation was used, in which training was done on the data from 11 sessions, and testing was done on the data from the 12th session. This was then repeated 12 times, using a different test session on each occasion.

Time-span decoding

In “time-span” decoding analyses, we trained the classifier at one time point in the trial to discriminate activity patterns elicited at another time point in the trial (Fig. 1B).

Action at stimulus time. To detect regions involved in encoding action representations at the time of initial stimulus presentation, we trained our classifier to discriminate between different action representations (double button press vs trackerball roll) at the time of action selection. We then tested the classifier at the time of initial stimulus presentation to assess whether activity in a given brain area during the time of initial stimulus presentation reflected the action that would subsequently be selected on that trial. A successful classification in a given brain region would indicate that information about the to-be-performed action is represented during the initial decision period, suggesting the presence of stimulus–response associations in that region.

Outcome at stimulus time. To detect regions involved in encoding outcome representations at the time of initial stimulus presentation in the trial, we trained our classifier to discriminate different target outcomes, e.g., blue circle versus red square as they were presented at the time of outcome delivery. We then tested the classifier at the time of initial stimulus presentation (i.e., at the onset of the trial) to assess whether activity in a given brain area during the time of initial stimulus presentation reflected the outcome that would ultimately be delivered on that trial (contingent on the subsequent action). A successful classification in a given brain area would indicate that information about the goal of an action is represented in that area during the initial decision period.

Decoding of integrated representations

We were also interested in testing for “integrated representations,” in which distinct combinations of stimuli and actions (e.g., S1-A1 and S2-A2) might be encoded as unique configurations (S1-A1 vs S2-A2), as opposed to encoded as elemental action representations (A1 vs A2). The key distinction is that an integrated representation of an S1-A1 combination would successfully decode only on trials in which A1 is selected in the presence of S1 but not otherwise; in contrast, a unitary representation of action A1 would successfully decode on any trial in which A1 was selected, regardless of whether S1 or S2 was present. To detect such integrated representations we performed the following steps.

Stage 1: Establishing potential regions of interest (ROIs). First we trained the classifier to decode S1-A1 versus S2-A2 configurations at the time of initial stimulus presentation and tested for those representations at the same time point. A significant signal in this analysis is indicative of the encoding of unitary stimulus representations, unitary action representations, or integrated stimulus-action representations.

Stage 2: Ruling out unitary stimulus representation. Second, we used the classifier weights trained up in stage 1 to also test for discrimination between S1-A2 versus S2-A1. If the classifier performs significantly above chance, this would indicate that unitary stimulus information is being decoded (since the only consistent labels between the training and testing data are S1 and S2).

Stage 3: Ruling out unitary action representation. In our third analysis, we again used the classifier weights from stage 1, and tested if the classifier could decode S2-A1 versus S1-A2. Similar logic implies that significant decoding in this analysis is consistent with unitary action representations.

It is also possible that both integrated and unitary representations are present in a region simultaneously. Significant classification in stage 1

Table 2. ROIs

Region	Coordinates	Source
dIPFC	(48, 9, 36)	(Gläscher et al., 2010)
vmPFC	(−3, 41, −11)	(McNamee et al., 2013)
cOFC	(21, 38, −11)	(McNamee et al., 2013)
Caudate (anterior)	(6, 10, 20)	(Tanaka et al., 2008)
Putamen/GP (posterior)	(−33, −24, 0)	(Tricomi et al., 2009)
Hippocampus	(18, −6, −20) (−34, −14, −18)	(Simon and Daw, 2011)
PPI analysis only		
Thalamus	WFU PickAtlas mask	(Maldjian et al., 2003)
Motor cortex	WFU PickAtlas mask	(Maldjian et al., 2003)

Correction was performed within a 10 mm radius sphere surrounding the corresponding coordinates.

but not in stages 2 and 3 is indicative of integrated stimulus-action representations only. Thus, to attribute decoding signals specifically to integrated representations, we consider the conjunction between two statistical maps obtained from per voxel paired *t* tests between (1) accuracy scores in stage 1 versus stage 2 and (2) accuracy scores in stage 1 versus stage 3. The only explanation for a signal that survives this stringent criterion is that it is generated by an integrated stimulus-action representation, since the first paired *t* test rules out stimulus-only decoding and the second rules out action-only decoding.

Significance testing

For the searchlight analyses, the percentage of correctly identified samples, averaged across folds in the cross-validation, was used as the classification score in each searchlight and this score was assigned to the voxel at the center of the searchlight sphere. This defined a classification accuracy map for each subject, which was then smoothed with an 8 mm FWHM kernel. A second-level analysis was implemented by performing voxelwise *t* tests comparing the distribution of accuracies across participants against 50%, which is the expected performance of an algorithm randomly labeling samples. Since multivariate classification is susceptible to optimistic classification biases, we performed permutation tests to validate our decoding procedure (McNamee et al., 2013).

All results were significant at familywise error rate (FWE)-adjusted $p < 0.05$ corrected for multiple comparisons by controlling the FWE with a 10 voxel extent threshold. We had strong prior hypotheses regarding action and outcome representations in posterolateral and anteromedial striatum, and in ventromedial and dorsolateral prefrontal cortex. Thus, in these areas, corrections were performed within small volumes defined a priori based on relevant functional imaging studies (see Table 2). Small volume corrections are denoted throughout by SVFWE and whole-brain corrections by FWE. For display purposes, we present overlays thresholded at $p < 0.005$ uncorrected.

Psychophysiological interactions. Blood-oxygen-level dependent (BOLD) time courses were extracted from ROIs using SPM's Volume of Interest functionality correcting for an *F*-contrast composed of all effects of interest in the GLM. ROIs were defined as the set of voxels within a 6 mm radius of seed coordinates, which were independently defined based on related functional imaging studies (see Table 2). A GLM was then constructed with three regressors in the following order: the BOLD time course from the seed region (the physiological term), an indicator regressor encoding the initial stimulus onset on each trial (the psychological term), and the corresponding interaction regressor. Once the GLMs were estimated for all participants, a second-level contrast (i.e., across participants) was specified for the interaction regressor. The resulting statistical map details the degree of coupling, modulated by the psychological regressor, between the seed region and voxels throughout the brain. It does this by measuring how much BOLD activity in the target location is accounted for by the interaction term in the GLM.

Results

Behavioral performance

Due to the relatively simple nature of the task, and the training conducted before each experimental condition, participants were expected to perform close to optimally (defined as choosing ac-

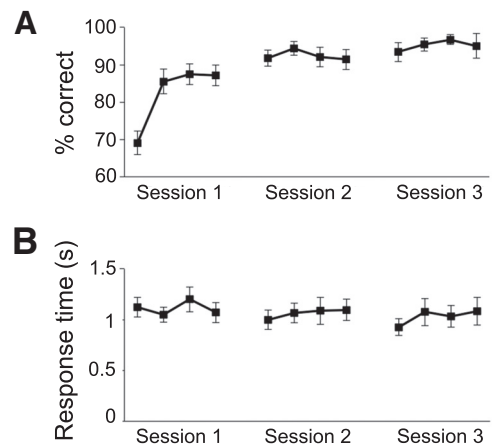


Figure 2. Behavior. **A**, Choice accuracy. **B**, Response times across training (Session 1) and experimental sessions. Each time bin represents four trials, and each data point is averaged across conditions and participants.

tions associated with the high reward distribution in each condition). To assess the acquisition of instrumental contingencies, we divided the three sessions making up each condition into 12 bins of four trials and computed the percentage of “correct” (i.e., optimal) responses in each bin, averaging across conditions. As can be seen in Figure 2A, participants quickly reached asymptotic performance; on average achieving and maintaining $>90\%$ correct responding by the beginning of the second session. When all trials were considered, the mean percentage of optimal action selection was 90%. Performance ranged from 85.4 to 97.4%, except in one participant who was an outlier in terms of having a performance level of 60.1%. The individual with the outlier performance level was nonetheless included in the fMRI analysis, as a sufficient number of trials were still available for the classification performance. A condition(4)-by-bin(12) repeated-measures ANOVA performed on the percentage correct scores yielded a significant effect of training ($p < 0.001$), but no effect of condition ($p = 0.99$) and no condition-by-bin interaction ($p = 0.16$). Pairwise comparisons of bin scores, averaged across conditions, revealed significant differences between the first and all subsequent bins (all $ps < 0.01$), as well as between the second bin and bins 6 through 12 (all $ps < 0.05$). Furthermore, the third and fourth bins both differed significantly from bins 6 and 9 through 12 (all $ps < 0.05$) and significant differences were also found between bin 7 and 11, as well as between bin 8 and 10 and 11 (all $ps < 0.05$). In contrast, response times (shown in Fig. 2B) were relatively constant across bins, with an analogous ANOVA yielding no main effects or interactions (smallest $p = 0.48$).

Neuroimaging results

Goal-directed associative encoding

We first tested for areas surviving individual tests for outcome information or action information during the initial stimulus period, and then finally report a conjunction across those tests, which is the key criterion for a region involved in goal-directed associative encoding (Balleine and Ostlund, 2007).

Outcome at stimulus time

We first tested for brain regions involved in encoding outcome identity at the time of the initial stimulus, as such representations would be indicative of regions having access to the goal or outcome at the time of decision making. We used a time-span analysis to train the classifier on outcome representations at the time

of outcome delivery, and tested at the time of stimulus presentation. Prospective representations of the predicted outcome state at the time of stimulus were identified in right dlPFC ($p < 0.05$ SVFWE, $t_{(18)} = 4.55$, $x = 60$, $y = 20$, $z = 34$) and in central OFC ($p < 0.05$ SVFWE, $t_{(18)} = 3.11$, $x = 18$, $y = 32$, $z = -20$).

Outcome at action time

We also tested for regions encoding outcome information at the time of action performance. For this we used a time-span analysis to train the classifier on outcome representations at the time of outcome delivery, and then tested at the time of action execution. We found significant signals in dlPFC ($p < 0.05$ SVFWE, $t_{(18)} = 5.15$, $x = 51$, $y = 17$, $z = 37$), vmPFC ($p < 0.05$ SVFWE, $t_{(18)} = 6.02$, $x = 0$, $y = 53$, $z = -20$), central OFC ($p < 0.05$ SVFWE, $t_{(18)} = 5.15$, $x = 30$, $y = 38$, $z = -11$), and caudate ($p < 0.05$ SVFWE, $t_{(18)} = 4.02$, $x = 9$, $y = 20$, $z = 16$).

Action at stimulus time

We also expected regions involved in goal-directed control to encode action information at the time of initial stimulus presentation. Out of the regions identified above as containing outcome information at the time of either initial stimulus presentation or action execution, two regions also contained action information at the initial stimulus time: the dlPFC ($p < 0.05$ SVFWE, $x = 57$, $y = 8$, $z = 34$, $t_{(18)} = 3.85$) and vmPFC ($p < 0.05$ SVFWE, $x = 0$, $y = 53$, $z = -20$, $t_{(18)} = 6.02$).

Regions containing both action and outcome information at stimulus time

To formally identify voxels containing both outcome and action information at the time of the initial stimulus, we performed a conjunction analysis on the results of the outcome at stimulus time and action at stimulus time statistical maps. This analysis yielded significant effects only in right dlPFC (conjunction: $p < 0.05$ SVFWE, $x = 60$, $y = 17$, $z = 34$, $t_{(18)} = 3.47$; Fig. 3A). Although we did not specify the dorsomedial prefrontal cortex (dmPFC) as an a priori region of interest, activity was also found in this region at an uncorrected threshold. Given that this region was identified as being involved in model-based RL algorithms in a previous study (Lee et al., 2014), we performed a *post hoc* small volume correction using dmPFC coordinates identified in that study ($x = 12$, $y = 32$, $z = 37$), which revealed a significant cluster ($p < 0.05$ SVFWE, $x = 21$, $y = 35$, $z = 40$, $t_{(18)} = 3.36$). As this was a *post hoc* inference, we refrain from discussing it further, but report it for completeness.

Regions containing both action at stimulus time and outcome at action execution

We also performed a conjunction analysis to pinpoint regions in which action information is available at the initial stimulus time, while outcome information is represented during action execution. This contrast revealed significant effects in the vmPFC ($p < 0.05$ SVFWE, $x = 3$, $y = 53$, $z = -20$, $t_{(18)} = 5.54$), as well as the left ($p < 0.05$ SVFWE, $x = -42$, $y = 26$, $z = 49$, $t_{(18)} = 4.78$) and

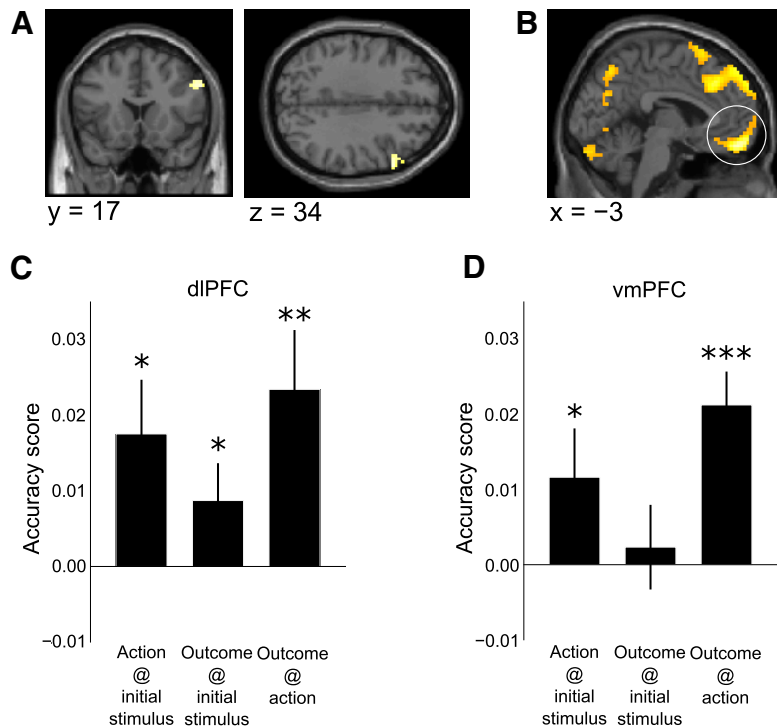


Figure 3. Goal-directed representations. **A**, Right dlPFC encoded both action and outcome representations at the time of the initial stimulus presentation (conjunction analysis, $x = 60$, $y = 17$, $z = 34$, $t_{(18)} = 3.47$). **B**, vmPFC encoded action at the time of initial stimulus presentation and outcome information at the time of action performance (conjunction, $x = 3$, $y = 53$, $z = -20$, $t_{(18)} = 5.54$). **C**, Bar plot depicts accuracy score distributions in an independently defined dlPFC ROI. This score is the decoding accuracy minus 0.5, which is the expected accuracy of a random algorithm, $*p < 0.05$, $**p < 0.005$. **D**, Bar plot depicts accuracy score distributions for vmPFC, $***p < 0.0005$.

right ($p < 0.05$ SVFWE, $x = 51$, $y = 29$, $z = 43$, $t_{(18)} = 3.97$) dlPFC (Fig. 3B).

Habitual encoding of stimulus–response associations

To identify brain regions that could potentially be involved in habitual action selection, we tested for areas that encoded action information at the initial stimulus time but that were not encoding outcome information at either the stimulus time or during action performance. Of the areas identified in the analysis testing for significant decoding of actions at the time of stimulus, two regions in particular were identified as containing action representations that did not also contain outcome representations: the posterior lateral putamen ($p < 0.05$ SVFWE, $x = -27$, $y = -22$, $z = 7$, $t_{(18)} = 3.24$; peak within same cluster $x = -21$, $y = -19$, $z = 7$, $t_{(18)} = 4.79$), and the supplementary motor cortex ($p < 0.05$ FWE, $x = 15$, $y = 32$, $z = 61$, $t_{(18)} = 7.95$). In an independent follow-up analysis using anatomically defined regions of interest centered on the posterior putamen and supplementary motor cortex, we tested whether these regions contained on average significantly better predictions of actions compared with outcomes at the time of stimulus. In a paired t test we found that action representations were significantly more strongly represented than outcome representations in both these regions (Fig. 4; putamen: $p = 0.001$, $t_{(18)} = 3.9$; supplementary model cortex: $p = 0.005$, $t_{(18)} = 2.86$). In addition to these paired t tests, we performed one-sample t tests against a random-chance accuracy score, which indicated that only action information, but not outcome information, was present in the putamen and supplementary model cortex at the time of the initial stimulus presentation.

If putamen is driving motor activity during the performance

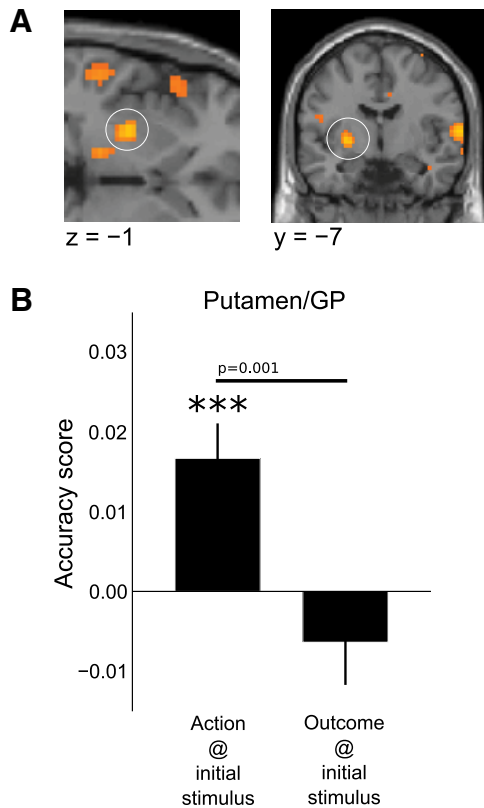


Figure 4. Habitual representations. **A**, A region of DLS (posterior putamen), extending into the globus pallidus (GP) was found to encode information about the action to be performed at the time of initial stimulus presentation ($p < 0.05$ SVFWE, $x = -27, y = -22, z = 7, t_{(18)} = 3.24$), but critically, no significant information about outcome. **B**, The distribution of accuracy scores for actions and outcomes at the time of initial stimulus in an independently defined putamen/GP region of interest, $***p < 0.0005$.

of habitual actions, one would expect this area to be functionally connected to thalamus, and the thalamus in turn to motor cortex in the contralateral (left) hemisphere, as dictated by the anatomy of corticostriatal loops. We tested for psychophysiological interactions (PPI) between an indicator variable for the onset of the initial stimulus and neural activity seeded at putamen and thalamus. The putamen-based PPI resulted in a significant correlation with activity in the left thalamus ($p < 0.05$ SVFWE, $x = -12, y = -19, z = -2, t_{(18)} = 4.07$) and the thalamus PPI correlated significantly with activity in left premotor cortex ($p < 0.05$ SVFWE, $x = -39, y = -7, z = 46, t_{(18)} = 7.39$). A weaker effect was also found in the ipsilateral (right) premotor cortex ($p < 0.05$ SVFWE, $x = 42, y = -10, z = 58, t_{(18)} = 5.45$). All seed coordinates used in the PPI were defined independently of results of the other analyses in this study (see Table 2).

Integrated stimulus-action representations

We also tested for integrated stimulus-action representations—encoding specific stimulus-action pairs as unique configurations (e.g., S1A1) rather than encoding actions independently of their antecedent stimulus—at the time of the initial stimulus (see Materials and Methods). Integrated stimulus-action representations were identified (Fig. 5) in the anterior dorsomedial striatum (caudate nucleus; $p < 0.05$ SVFWE, $x = 15, y = 11, z = 22, t_{(18)} = 4.16$) and hippocampus ($p < 0.05$ SVFWE, $x = 24, y = -1, z = -20, t_{(18)} = 4.35$).

Integrated action-outcome representation

We also tested for evidence of integrated action-outcome representations at the time of stimulus, using a similar approach. No significant decoding of integrated action-outcome representations was found.

Ruling out response time confounds

There was a significant difference in response times for the two actions (two-sided paired t test, $t_{(18)} = 3.415, p = 0.003$); in contrast, no difference was found in response times as function of the identity of the initial stimulus ($t_{(18)} = 0.561, p = 0.582$) or outcome state (two-sided paired t test, $t_{(18)} = 0.577, p = 0.571$). To ensure that response times were not confounding our results, we ran additional analyses assessing action-dependent decoding. Specifically, in these analyses, we included individual trial reaction times as a covariate of no-interest in the fMRI design matrix and re-ran all of the classification analyses involving actions as described above. We filtered out any variance in voxel activity explained by trial-by-trial response times at the initial stimulus and action execution time points. This was accomplished by estimating a GLM, which included trial-by-trial reaction times as parametric modulators, time locked to the initial stimulus and action execution trial events. Following GLM estimation, β -values for these parametric modulators were multiplied by the corresponding regressors and linearly subtracted from the data. All of our results remained significant after inclusion of the reaction time covariate, indicating that our classifier is not relying on differences in reaction times to decode action information.

Discussion

Contemporary associative theory distinguishes between habitual S-R associations and a combination of S-O, O-R, and R-O associations thought to mediate goal-directed performance (Balleine and Ostlund, 2007). In this study, we used multivariate pattern analysis to assess whether dissociable regions of the human brain encode these distinct associative structures. Unlike previous work in humans, contrasting qualitatively different experimental conditions designed to encourage different action-selection strategies, or comparing largely parameter-driven value signals generated by RL algorithms, our approach sought to identify a neural implementation of the associative content of goal-directed versus habitual behavioral control. We found evidence for stimulus-elicited response representations but no outcome representations, indicative of habits, in the DLS (posterior putamen). Conversely, in the vmPFC, dlPFC, and anterior caudate nucleus, both response and outcome representations were present, indicative of goal-directed decision making.

Our finding of stimulus-elicited response, but not outcome, representations in the DLS suggests that this area is especially involved in encoding S-R associations. While previous studies have found evidence that activity in this area increases over time as habits come to control behavior (Tricomi et al., 2009), and that activity in this region correlates with model-free value signals (Wunderlich et al., 2012; Lee et al., 2014), the present study illuminates the associations encoded in the region. A previous report found that the degree of structural connectivity between the posterior putamen and the premotor cortex predicts susceptibility to habit-like “slips-of-action” (de Wit et al., 2012). Our connectivity analysis suggests a potential mechanism by which stimulus—response-related activity in the putamen—is ultimately transferred to the motor cortex via the thalamus, to implement habitual motor control.

Whereas habits depend on a reflexive retrieval of a previously reinforced response, goal-directed behavior involves selecting,

evaluating, and initiating an action based on the probability and utility of its consequences. The “associative cybernetics theory” (Balleine and Ostlund, 2007; de Wit and Dickinson, 2009) postulates that the retrieval of potential outcomes, of the actions that produce them, and of the values of those actions, is mediated respectively by S-O, O-R, and R-O associations. Critically, to allow for sensitivity to sensory-specific outcome devaluation and contingency degradation, defining features of goal-directed performance, the associations relating the probabilities and utilities of potential outcomes to the stimuli and actions that produce them must be flexible and current, suggesting a dynamic binding of features.

One area well suited for the dynamic binding of stimuli, actions, and outcomes is the dlPFC, given prior evidence for a role of this structure in working-memory and goal-directed behavior more generally (Goldman-Rakic, 1996; Miller and Cohen, 2001). We found that activity in this region reflected representations of both action and outcome identities at the time of initial stimulus presentation, indicative of a key role for this region in encoding the information necessary to guide goal-directed actions at the time of decision making. Specifically, the finding that dlPFC activity reflects information about action and outcome identities, necessary for computing goal-directed action values, is consistent with a contribution of this area to encoding the model component of a model-based RL algorithm. Previous findings reported state-prediction errors in this region that could underpin learning of the underlying associations needed to form such a model (Gläscher et al., 2010). The present findings suggest that dlPFC is involved in learning or updating such a model as well as encoding (or at least retrieving) the model itself.

The contribution of dlPFC to the encoding of associative information necessary for computing goal-directed actions at the time of initial stimulus presentation can be contrasted with our findings in the vmPFC. Whereas vmPFC did encode information about the action at the time of initial stimulus presentation, information about the outcome identity was not present until later in the trial, during action execution. However, in the central orbitofrontal cortex (cOFC), an area adjacent to and highly connected to the vmPFC (Carmichael and Price, 1996), we did find that outcome identity information was represented at the time of initial stimulus presentation. One possibility, therefore, is that the cOFC encodes the identity of a goal at the time of initial decision making and that this outcome-identity representation is then used to retrieve outcome value signals in the vmPFC. Consistent with this interpretation, a previous study by our group reported activity in cOFC extending to vmPFC correlating with the categorical identity of the goal at the time of decision making (McNamee et al., 2013). In that previous study, information about the value of the goal was most prominently represented in vmPFC. An important feature of our experiment is that we have controlled for value (i.e., kept value constant throughout, with high and low value outcomes assigned equally often to every possible combination of stimuli and actions), to ensure that outcome identity information is not confounded with the outcome value. Thus, we cannot test in the present design when value information about outcomes emerges in vmPFC. However, previous studies have reported such information to be present in both the

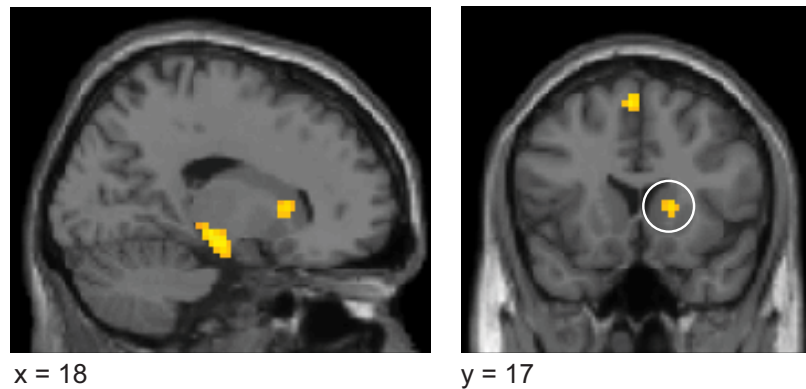


Figure 5. Integrated stimulus-action representations were localized in DMS (anterior caudate nucleus; $p < 0.05$ SVFWE, $x = 15, y = 11, z = 22, t_{(18)} = 4.16$) and hippocampus ($p < 0.05$ SVFWE, $x = 24, y = -1, z = -20, t_{(18)} = 4.35$).

vmPFC and in the dlPFC at the time of decision making (Plassmann et al., 2007; Chib et al., 2009).

Our findings provide new insight into the differential functions of DLS versus DMS. While the posterior DLS (posterior putamen) was found to encode representations of responses elicited by discriminative stimuli, a different type of stimulus–response coding was present in the anterior dorsomedial striatum (anterior caudate), as well as in the hippocampus, such that the encoding of response associations was integrated with stimulus identity: a unique distributed representation was present in these areas for each stimulus–response pair. In contrast, in the DLS, each response was coded independently of the stimulus that elicited it. The binding of stimulus–response associations into a single representation found in the DMS and hippocampus could underpin a form of abstraction of stimulus–response codes, which could potentially be part of a mechanism for chunking stimulus–response configurations. The hippocampus in particular has been proposed to support the unitization, or chunking, of stimulus elements into unique configurations (Sutherland and Rudy, 1989). Our results suggest that the DMS may also play an important role in this process. Our finding of a difference in the type of response coding present in the DMS versus DLS is important, given that some previous neurophysiology studies have not found clear differences in information encoding between these regions (Kim et al., 2009; Kimchi et al., 2009; Stalnaker et al., 2010).

Our results also differentiate between the DLS and DMS in that only the DMS was found to encode outcome identity. An important feature of our experimental design is that differences found in outcome identity representations between these regions could not be accounted for by potential differences in the value of the outcomes. While we did have actions leading to high versus low rewards in our experiment, we trained the classifier to distinguish between different outcome states leading to the same high-valued reward. This is necessary because, differences in outcome value, i.e., between high- and low-valued goal states, could drive differences in outcome-related neural activity in a brain region even if that area is not explicitly representing outcome identity: indeed, even a pure S-R learning system would discriminate high- and low-valued states as the high-valued state would be associated with stronger S-R associations through trial-by-trial reinforcement.

Naturally, the absence of significant decoding from the BOLD signal in a given brain area does not imply the absence of that information at the level of single neurons. fMRI and single-unit data may capture different aspects of neural activity in any event, with the BOLD signal suggested to be correlated more closely

with input into a region and intrinsic processing therein as opposed to output (Logothetis et al., 2001). Nevertheless, it is striking that our current findings about information content do accord very well with previous evidence about the differential role of dorsolateral striatum in habitual control, and a corresponding role for dorsomedial striatum and prefrontal cortex regions in goal-directed actions (Hikosaka et al., 1989; Yin et al., 2004, 2005; Valentin et al., 2007; Gläscher et al., 2010).

To conclude, our present results suggest that different brain areas are involved in encoding different kinds of information about responses and outcomes, consistent with a differential role for these regions in goal-directed and habitual learning and control. Whereas cortical areas including the dorsolateral prefrontal cortex and the ventromedial prefrontal cortex alongside the anterior dorsomedial striatum contained associative information about the identities of both responses and outcomes, necessary for goal-directed control, the dorsolateral striatum contained only information about stimuli and responses, which would be sufficient for habitual but not goal-directed control.

References

- Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407–419. [CrossRef Medline](#)
- Balleine BW, Ostlund SB (2007) Still at the choice-point. *Ann N Y Acad Sci* 1104:147–171. [CrossRef Medline](#)
- Balleine BW, Daw ND, O'Doherty JP (2008) Multiple forms of value learning and the function of dopamine. In: *Neuroeconomics: decision making and the brain* (Glimcher PW, Camerer C, Fehr E, Poldrack RA, eds), pp 367–385. New York: Elsevier.
- Carmichael ST, Price JL (1996) Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *J Comp Neurol* 371:179–207. [CrossRef Medline](#)
- Chib VS, Rangel A, Shimojo S, O'Doherty JP (2009) Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J Neurosci* 29:12315–12320. [CrossRef Medline](#)
- Clithero JA, Smith DV, Carter RM, Huettel SA (2011) Within- and cross-participant classifiers reveal different neural coding of information. *Neuroimage* 56:699–708. [CrossRef Medline](#)
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204–1215. [CrossRef Medline](#)
- Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* 19:430–441. [CrossRef Medline](#)
- de Wit S, Dickinson A (2009) Associative theories of goal-directed behaviour: a case for animal–human translational models. *Psychol Res* 73:463–476. [CrossRef Medline](#)
- de Wit S, Watson P, Harsay HA, Cohen MX, van de Vijver I, Ridderinkhof KR (2012) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J Neurosci* 32:12066–12075. [CrossRef Medline](#)
- Dickinson A (1985) Actions and habits: the development of behavioural autonomy. *Philos Trans R Soc Lond B, Biol Sci* 308:67–78. [CrossRef](#)
- Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585–595. [CrossRef Medline](#)
- Goldman-Rakic PS (1996) Regional and cellular fractionation of working memory. *Proc Natl Acad Sci U S A* 93:13473–13480. [CrossRef Medline](#)
- Gremel CM, Costa RM (2013) Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat Commun* 4:2264. [CrossRef Medline](#)
- Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci* 26:8360–8367. [CrossRef Medline](#)
- Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S (2009) PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7:37–53. [CrossRef Medline](#)
- Hikosaka O, Sakamoto M, Usui S (1989) Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J Neurophysiol* 61:814–832. [Medline](#)
- Kahnt T, Heinzle J, Park SQ, Haynes JD (2010) The neural code of reward anticipation in human orbitofrontal cortex. *Proc Natl Acad Sci U S A* 107:6010–6015. [CrossRef Medline](#)
- Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex* 13:400–408. [CrossRef Medline](#)
- Kim H, Sul JH, Huh N, Lee D, Jung MW (2009) Role of striatum in updating values of chosen actions. *J Neurosci* 29:14701–14712. [CrossRef Medline](#)
- Kim H, Lee D, Jung MW (2013) Signals for previous goal choice persist in the dorsomedial, but not dorsolateral striatum of rats. *J Neurosci* 33:52–63. [CrossRef Medline](#)
- Kimchi EY, Torregrossa MM, Taylor JR, Laubach M (2009) Neuronal correlates of instrumental learning in the dorsal striatum. *J Neurophysiol* 102:475–489. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540. [CrossRef Medline](#)
- Lee SW, Shimojo S, O'Doherty JP (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81:687–699. [CrossRef Medline](#)
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157. [CrossRef Medline](#)
- Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19:1233–1239. [CrossRef Medline](#)
- McNamee D, Rangel A, O'Doherty JP (2013) Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat Neurosci* 16:479–485. [CrossRef Medline](#)
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202. [CrossRef Medline](#)
- Mitchell TM (1997) *Machine learning*. New York, USA: McGraw-Hill.
- O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. *Neuron* 38:329–337. [CrossRef Medline](#)
- Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45:S199–209. [CrossRef Medline](#)
- Plassmann H, O'Doherty J, Rangel A (2007) Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J Neurosci* 27:9984–9988. [CrossRef Medline](#)
- Rudy JW, Sutherland RJ (1989) The hippocampal formation is necessary for rats to learn and remember configurational discriminations. *Behav Brain Res* 34:97–109. [CrossRef Medline](#)
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340. [CrossRef Medline](#)
- Simon DA, Daw ND (2011) Neural correlates of forward planning in a spatial decision task in humans. *J Neurosci* 31:5526–5539. [CrossRef Medline](#)
- Stalnaker TA, Calhoun GG, Ogawa M, Roesch MR, Schoenbaum G (2010) Neural correlates of stimulus–response and response–outcome associations in dorsolateral versus dorsomedial striatum. *Front Integr Neurosci* 4:12. [CrossRef Medline](#)
- Tanaka SC, Balleine BW, O'Doherty JP (2008) Calculating consequences: brain systems that encode the causal effects of actions. *J Neurosci* 28:6750–6755. [CrossRef Medline](#)
- Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. *Eur J Neurosci* 29:2225–2232. [CrossRef Medline](#)
- Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci* 27:4019–4026. [CrossRef Medline](#)
- Wunderlich K, Dayan P, Dolan RJ (2012) Mapping value based planning and extensively trained choice in the human brain. *Nat Neurosci* 15:786–791. [CrossRef Medline](#)
- Yin HH, Knowlton BJ, Balleine BW (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci* 19:181–189. [CrossRef Medline](#)
- Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* 22:513–523. [CrossRef Medline](#)