

Dynamics of Frontal, Striatal, and Hippocampal Systems during Rule Learning

Carol A. Seger and Corinna M. Cincotta

Department of Psychology and Program in Molecular, Cellular, and Integrative Neuroscience, Colorado State University, Fort Collins, CO 80523, USA

We examined interactions between frontal, striatal, and hippocampal systems during a rule-learning task. Nineteen healthy young adults solved multiple rule-learning problems requiring hypothesis testing while functional magnetic resonance images were obtained. Activity in the head of the caudate peaked early after the beginning of each problem and then dropped rapidly. In contrast, activity in prefrontal cortex areas reached peak values later. These results are in accordance with theories suggesting that the striatum identifies the behavioral context necessary for the frontal lobe to select an appropriate strategy. Striatal and hippocampal systems showed antagonistic patterns of activity: Activation in the anterior hippocampus decreased, whereas caudate activity increased. Good learners showed higher activity in the body and tail of the caudate than poor learners, whereas learning success correlated negatively with activity in the hippocampus. Activation in the head of the caudate correlated negatively with hippocampal activation, indicating a potential mechanism for hippocampal activity reduction.

Keywords: basal ganglia, categorization, concept learning, medial temporal lobe

Introduction

The striatum has been shown to play an important role in human learning in conjunction with frontal and medial temporal lobe systems. The striatum, and in particular the caudate nucleus, is required for learning associations or rules that link stimuli with responses or categories; this learning has been termed procedural or “habit” learning (Mishkin and others 1984) in contrast with the “memory” processes subserved by the medial temporal lobe. Rodents (Packard and Knowlton 2002), monkeys (Fernandez-Ruiz and others 2001), and humans (Channon and others 1993; Knowlton, Mangels, and Squire 1996; Knowlton, Squire, and others 1996; Filoteo, Maddox, Simmons, and others 2001; Maddox and Filoteo 2001; Keri and others 2001, 2002; Witt and others 2002; Myers and others 2003; Sage and others 2003; Shohamy and others 2004; Filoteo, Maddox, Salmon, and Song 2005) with striatal damage are impaired at learning rules or associations linking stimuli with categories or responses. Furthermore, functional imaging studies have found striatal activity during verbalizable rule (Rao and others 1997; Monchi and others 2001; Filoteo, Maddox, Simmons, and others 2005) and implicit association learning (Poldrack and others 1999, 2001; Toni and others 2001, 2002; Seger and Cincotta 2002, 2005).

Prefrontal cortex is commonly active along with the striatum during rule-learning tasks (Rao and others 1997; Monchi and others 2001; Toni and others 2001, 2002; Filoteo, Maddox, Simmons, and others 2005). Neuropsychological studies show that frontal lobe damage leads to impairment in tasks involving

learning explicit rules for categorizing stimuli (Milner 1963; Cicerone and others 1983; Stuss and others 2000; Reverberi and others 2005). However, impairment is not always found following frontal lobe damage on implicit association-learning tasks such as the probabilistic classification task (Knowlton, Mangels, and Squire 1996).

The prefrontal cortex and the striatum are often simultaneously active during learning. However, it is unclear whether or not the 2 brain regions interact during learning and, if so, whether they play qualitatively different roles. Anatomically, the striatum interacts with cortex, particularly prefrontal cortex, in corticostriatal “loops” (Alexander and others 1986; Lawrence and others 1998). One theory is that prefrontal cortex and striatum work together as “teacher” and “student.” Prefrontal cortex is responsible for initial rule or association induction. The striatum then mediates rule application during extended task performance (Packard and Knowlton 2002), possibly by recoding the information into a “chunked” action plan (Graybiel 1998). This theory predicts that frontal activity should be high early in learning and decrease as rules are learned, whereas striatal activity should increase with learning in step with behavioral indexes of learning.

An alternative prediction comes from computational models of frontostriatal systems. Houk and Wise (1995) postulated that the convergence of projections from widespread cortical areas onto single striatal spiny neurons puts the striatum in a position to identify the behavioral context the organism is in by integrating information across many features of the environment (both internal and external). They proposed that the striatum transmits this context information to cortex to aid the cortex in selecting the appropriate strategy or behavior for the circumstances. Similarly, Frank and colleagues (Frank and others 2001; Frank 2004) in their model of frontostriatal interaction in working memory proposed that the striatum selectively gates the updating of representations in the frontal cortex. Both theories predict that striatal cells will fire transiently when a change in context or need to update working memory is detected; frontal areas will become active in response and maintain activity across the time the context remains relevant. A recent study in monkeys supports this model of frontostriatal interaction. Pasupathy and Miller (2005) taught monkeys associations between saccades and visual stimuli. They found early and transient changes in striatal neuron activity when the associations were changed, in combination with a slower trend in frontal neuron activity that followed closely the time course of learning.

Striatal and medial temporal lobe systems are thought to underlie qualitatively different forms of learning and memory: the striatum mediates learning consistent patterns across

stimuli, whereas the medial temporal lobe enables the representation of individual episodes. Consistent with these opposing functions, research across species finds an antagonistic relationship between striatal and medial temporal lobe systems (Poldrack and Packard 2003). This antagonism includes reciprocal activation and suppression of each system, as well as differences in recruitment of each system across tasks with differing demands for generalization or individual item memory. For example, Poldrack and others (2001) found greater striatal activity during a probabilistic classification task, in which subjects learned to classify stimuli into categories, but greater hippocampal activity during a paired associates task, in which subjects memorized relationships between stimuli and category names. Poldrack and Rodriguez (2004) argue based on connectivity analyses that the interaction between striatum and medial temporal lobe is mediated by prefrontal cortex.

There is also evidence that striatal subregions participate in dissociable functions during learning tasks. Although individual single striatal spiny cells receive synaptic input from an extended portion of cortex (Parent and Hazrati 1995), different regions of the striatum have predominant interconnections with particular cortical regions. The cortex and striatum together participate in processing loops in which information travels from cortex to striatum to GPI/SNr to thalamus and back to cortex (Middleton and Strick 2000). The dorsolateral prefrontal cortex is most strongly interconnected with the head of the caudate (Alexander and others 1986; Lawrence and others 1998), whereas extrastriate and inferotemporal visual cortex project to the ventral body and tail of the caudate nucleus (Saint-Cyr and others 1990; Webster and others 1993; Yeterian and Pandya 1995) and receive projections back from these areas through the SNr and thalamus (Middleton and Strick 2000). Thus, frontal executive processes present in learning tasks (setting context, receiving and processing feedback, shifting between hypotheses) should recruit the head of the caudate, whereas visual processing (detecting visual similarity, potentiating acquired visual stimulus–response associations) should recruit the body and tail of the caudate (Ashby and others 1998). Seger and Cincotta (2005) found evidence for this dissociation in a classification-learning task: activity in the head of the caudate was sensitive to feedback processing, but not successful classification, whereas activity in the body and tail of the caudate showed the opposite pattern. Specifically, activity in the body and tail of the caudate increased across classification training and correlated with classification accuracy across subjects.

The goal of this study was to compare frontal, striatal, and hippocampal recruitment across time within a single task. The task was a rule-learning task that was extensively developed in the 1960s and 1970s by Levine and colleagues (Levine 1975). Neuropsychological studies have shown that people with frontal lobe damage (Cicerone and others 1983) and those with Parkinson's disease (Channon and others 1993) are impaired in performance on this task. Previous functional imaging research found both frontal and striatal activity during performance of this task in comparison with a low-level baseline (Rao and others 1997).

Subjects solved multiple rule-learning problems in a single scanning session, which allowed us to examine the time course of learning across problems. The early trials of a typical rule-learning problem are shown in Figure 1. Subjects viewed 2 stimuli on each trial, indicated which one they believed followed the rule, and then received feedback. The stimuli

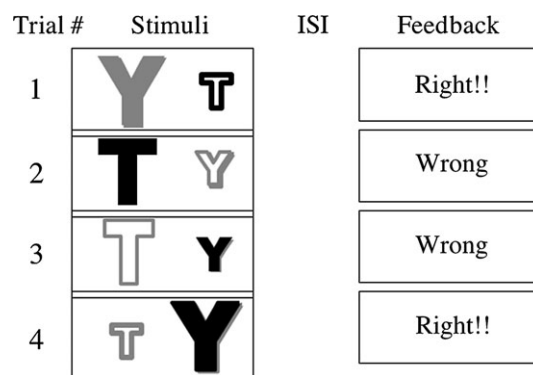


Figure 1. Diagrammatic representation of problem structure. On each trial, subjects viewed 2 stimuli and pressed the key corresponding to the location of the one (left or right) that followed the rule. After making a response, the subject received either positive (Right!!) or negative (Wrong) feedback.

differed on 4 features: letter identity, color, size, or font; features were repaired on each trial. Subjects were told that the correct rule would be a simple unidimensional rule, for example, “choose the blue stimulus” or “choose the larger stimulus.” Using a task in which each problem can be solved relatively quickly allowed us to examine learning across multiple problems and gave us more power in identifying patterns of activity characteristic of early and late stages of problem solving.

We compared 2 phases of learning, rule learning and rule application: rule learning encompasses the initial trials of each rule-learning problem, during which the subject is still unsure of the correct rule, whereas rule application (which starts after the last error trial) reflects the period in which subjects have determined the rule and are simply applying it to each new stimulus pair. Behavioral research indicates that during rule learning, subjects engage in a process of hypothesis testing, in which they propose a hypothesis, make a selection on the basis of the hypothesis, examine the feedback following the response, and if feedback is negative, shift to a new hypothesis (Levine 1975). We identified regions of interest (ROIs) in frontal, striatal, and medial temporal areas that were differentially active during rule learning and rule application and examined the time course of their activity. We also performed event-related analyses to examine differences in neural recruitment associated with feedback processing.

Materials and Methods

Subjects

A total of 19 people participated in the study, 12 in version 1 and 7 in version 2. Subjects were Stanford community members, 8 males and 11 females, with an average age of 24.0 years (range: 19–33). Subjects met criteria for magnetic resonance imaging (MRI) scanning (no metallic implants, no claustrophobia, head size compatible with the custom head coil) and were neurologically healthy (no known neurological or psychiatric injury or disease, not taking any psychoactive medication or drugs). One subject's behavioral data had several errors introduced due to noise in the response box system and had to be excluded from the individual differences analyses.

Materials

Rule Task

The rule task was based on a task developed by Levine (1975). In the rule task, subjects were presented with pairs of letters, one on the left side of the screen and one on the right side, and had to choose one of

them as following the rule by pressing the response button corresponding to the side of the screen that the correct letter appeared on. The letters differed on 4 dimensions: letter identity, letter color, letter size, and letter font. Letters and colors were selected from all 26 letters of the alphabet and a set of 8 different easily named colors (red, orange, yellow, blue, green, turquoise, purple, and brown); for each rule-learning task, 2 of the letters and 2 of the colors were selected randomly. For letter size, there were only 2 options (small, 48 points vs. large, 72 points), and there were 2 fonts used (outline and solid). The letters were presented in a sans serif block font (Helvetica). Each stimulus feature appeared in 1 of the 2 letters on each trial, but the pairings of features were randomized across trials, as shown in Figure 1.

Baseline Tasks

During the baseline task, subjects viewed pairs of letters as in the rule-learning task. In version 1, the baseline task was the Choose task. Subjects were told a particular feature (e.g., red, outline, or large) at the beginning of the block of baseline trials and had to press the response button corresponding to the item that contained that feature. In version 2, a lower level baseline task was used, called the Observe task. Subjects merely observed the pairs of letters and made no responses. The Baseline task blocks served only to separate Rule task blocks and were not included in the analyses presented here.

Procedure

In version 1, each subject completed 2 scans, each consisting of 10 alternating Rule and Baseline (Choose task) blocks. In version 2, subjects completed a single scan of 14 alternating Rule and Baseline (Observe task) blocks. At the beginning of each block, a phrase was presented that indicated the task to perform on the following blocks: For Rule task, the phrase was "Determine Rule." This phrase was presented for the same amount of time as a trial. The instruction phrase was then followed by trials of stimuli. For the Rule task, there were 12 trials per block in version 1 and 10 trials in version 2. Each trial lasted for 3000 ms in version 1 and 3250 ms in version 2; the letters appeared for 2000 ms, during which the subjects made their responses. After the letters were cleared, the screen was blank for 250 ms and then feedback (either "Right!!" or "Wrong") appeared for 500 ms. For the Baseline-V1 task, the instruction phrase was "Choose _____," where _____ was filled by a single feature from 1 of the 4 manipulated dimensions (e.g., Choose Red, Choose Small, Choose Outline, Choose B). For the Baseline-V2 task, the instruction phrase was "Observe letters," and on subsequent trials, subjects viewed the stimulus displays without making responses. For the Choose task, there were 6 trials per block, and for the Observe task, there were 5 trials per block. Each scan lasted for 10 min in version 1; the single scan in version 2 lasted for 13 min and 53 s.

Imaging was performed with a custom-built whole head coil in a 3.0-T MRI Signa LX Horizon Echospeed (General Electric Medical Systems, Milwaukee, WI, USA). Head movement was minimized for subjects using a "bite-bar" formed with the subject's dental impression. In addition to the functional scans, 3 anatomical scans were performed: a coronal T_1 -weighted localizer scan, a three-dimensional high-resolution T_1 -weighted spoiled gradient echo scan with 124 contiguous 1.5-mm slices (minimum full echo time [TE], 30 degree flip angle, 24 cm field of view), and T_1 -weighted spin-echo anatomical images in 28 contiguous 4-mm axial slices (30-ms TE, 1500-ms repetition time [TR]). Functional scanning was performed using a T_2^* -sensitive gradient echo spiral pulse sequence (Preston and others 2004) (30-ms TE, 76 degree flip angle, 20-cm field of view, 64×64 acquisition matrix). The TR was 2000 ms in version 1 and 1500 ms in version 2.

Stimuli were presented using a magnet-compatible projector (Resonance Technology, Inc., Van Nuys, California) that back-projects visual images onto a screen mounted above the subject's head. The rule-learning task was programmed using PsychoScope software for version 1 and E-Prime software (Psychology Software Tools, Pittsburgh, Pennsylvania) for version 2. Responses were obtained using a magnet-compatible response system.

Image Processing

Image preprocessing was performed using BrainVoyager 2000 4.9, and statistical analyses were performed using BrainVoyager QX 1.0.9 (Brain

Innovation, Maastricht, The Netherlands). The functional data were first subjected to preprocessing, consisting of three-dimensional motion correction, slice scan time correction, and temporal data smoothing with a high-pass filter of 3 cycles in the time course and linear trend removal. Each subject's high-resolution anatomical image was normalized to the Talarach and Tournoux (1988) brain template. The normalization process in BrainVoyager consists of 2 steps, an initial rigid body translation into the anterior commissure-posterior commissure plane followed by an elastic deformation into the standard space performed on 12 individual subvolumes. The resulting set of transformations was applied to the subject's functional image volumes to form volume time course representations to be used in subsequent statistical analyses. Finally, the volume time course representations were spatially smoothed with a Gaussian kernel, full width at half maximum of 6.0 mm.

BrainVoyager QX was used to analyze contrasts between conditions. A model of the hemodynamic response for each condition was formed by convolving a prototypical hemodynamic response function with the time course of the condition. Conditions were compared using the general linear model with separate subject predictors and subjects treated as random effects. The false discovery rate method (Genovese and others 2002) of correcting for multiple comparisons was used.

Differences in brain activity between conditions were examined within frontal, striatal, and hippocampal ROIs. ROIs were defined based on functionally activated clusters using the tools in BrainVoyager; the ROI is identified via a region-growing process, which starts with a voxel indicated by mouse click and which spreads to suprathreshold adjacent voxels, stopping at the boundaries of the functional cluster. Percent signal change within each ROI was calculated for each subject across rule-learning problems using the event-related averaging functions and the ROI analysis function. Percent signal change was measured from the onset of the first trial of a rule-learning problem across a course of 35 s, approximately the entire length of the rule-learning problem.

Results

Behavioral Results

The primary measure of learning was the trial of last error: the number of the last trial within each rule-learning problem on which an error was committed. The mean trial of last error across rule-learning problems and across participants was 4.8 (standard deviation [SD] = 0.7); for version 1, it was 4.7, and for version 2, it was 5.0. The range across participants extended from 3.3 to 6.0. Subjects became more efficient at finding the rule across problems: average trial of last error for first half (mean [M] = 5.5, SD = 1.0) was significantly greater than for the second half (M = 3.9, SD = 0.76), $t(18) = 6.4$, $P < 0.0001$.

Reaction times (RT) differed across the time course of learning: mean RT was 940 ms (SD = 189) during rule learning (trials preceding and including the trial of last error) and 685 ms (SD = 84) during rule application (trials following the trial of last error), $t(18) = 6.1$, $P < 0.0001$. Reaction time for Choose task trials for version 1 subjects was 629 (SD = 65); version 2 subjects made no responses during Observe trials. There were very few missed trials in which a response was not made within the time allotted: across subjects, the average percentage of trials on which responses were omitted was 0.23%, and no individual subject had more than 2 missed trials total. For purposes of the analyses, missed trials were considered to be incorrect responses.

Imaging Results

Rule Learning versus Rule Application

In order to examine the brain areas associated with the rule-learning and rule-application phases, we modeled the

hemodynamic response function for separate epochs defined individually for each subject and each problem. Rule learning was defined as extending from the onset of the first trial of a rule-learning problem through the trial of last error on that problem. Rule application extended from the trial following the trial of last error through the last trial of the rule-learning problem.

In comparison with rule application, rule learning activated a wide network of frontal areas. These areas of activation were used to define ROIs as described in Materials and Methods. As described in Table 1 and depicted in Figure 2, ROIs for the frontal lobes were in the bilateral inferior frontal gyri, bilateral middle frontal gyri, bilateral medial frontal areas, and left anterior frontal cortex. Due to the blocked nature of this comparison, this analysis is sensitive to all neural activity differences between the rule-learning and rule-application tasks and may include areas of activation corresponding to processes such as working memory that may not be specific to rule learning.

Four areas of the striatum were more active during rule learning than rule application. These 4 ROIs, which are depicted in Figure 2, encompass the right and left heads of the caudate and the right and left body and tail of the caudate. Other brain regions significantly more active for rule learning than rule application are listed in Table 2.

Rule application activated a smaller set of regions in comparison with rule learning. Most notably, rule application led to greater activity in the right anterior hippocampus, which was identified as an ROI (see Table 1 and Fig. 3).

Changes in Activity Across the Time Course of Rule Learning: Frontal and Striatal Patterns

We compared the dynamics of the hemodynamic response across the time course of problem solving in the 6 frontal ROIs, the 4 striatal ROIs, and the right anterior hippocampal ROI. As shown in Figure 2, activity increased gradually across the course of problem solving in all 6 ROIs within the frontal lobe, with peak activity being reached between 12 and 20 s after the onset of the problem. Activity in the middle frontal gyri and left anterior frontal ROIs peaked at a later time point and reached

higher percent signal change than in the inferior frontal gyri and medial frontal cortex.

Within the striatum activity in the head of the caudate followed a different pattern than in the body and tail. Activity in the right and left heads of the caudate increased quickly to a peak between 7 and 11 s after the beginning of the rule-learning task and then declined rapidly from those levels. This pattern is consistent with the observed response being primarily driven by neural changes occurring at the very beginning (first 1 or 2 trials) of the rule-learning problems; previous research has shown that responses to short lasting bursts of neural activity typically show maximal hemodynamic response at a latency of approximately 6–10 s (Boynton and others 1996). In contrast, activity in the body and tail ROIs rose quickly but remained at consistently elevated levels across the course of the rule-learning problem.

To directly compare the time course of activity in the frontal and striatal regions, we defined 2 critical time periods, one “early” and one “late.” The early time period was defined as the average across 7–11 s, which encompassed the average peak of the hemodynamic response in the striatum. The late time period was defined as 15–19 s and encompassed the average peak within the frontal ROIs. We chose the head of the caudate and the middle frontal gyrus for further analyses both because these areas showed the most pronounced peaks and also because they are the closest to the human analogs of the areas examined in monkeys by Pasupathy and Miller (2005). A $2 \times 2 \times 2$ within-subjects analysis of variance (ANOVA) with factors of brain area (head of caudate vs. middle prefrontal cortex), laterality (right vs. left hemisphere), and time (early, 7–11 s, and late, 15–19 s) revealed a significant interaction between brain area and time, $F_{1,18} = 17.9$, $P < 0.05$. As shown in Figure 4, activity in the caudate decrease from early to late time periods, whereas activity in the frontal ROIs increased. The ANOVA also revealed a significant main effect of brain area, $F_{1,18} = 39.2$, $P < 0.0005$, with higher activity in the frontal lobes than caudate.

Finally, we compared peak latency across subjects for the head of the caudate and middle frontal gyri ROIs. For each subject in each ROI, peak latency was defined as the highest percent signal change value between 4 and 25 s after the beginning of the rule-learning problem. Three subjects did not have the percent signal change exceed 0.1% in 2 or more of the ROIs and were excluded from the analysis. Average peak latency is shown in Figure 5. A 2×2 ANOVA with factors of brain area (head of caudate vs. middle prefrontal cortex) and laterality (right vs. left hemisphere) found a main effect of brain area, $F_{1,15} = 13.2$, $P < 0.005$, such that the peak occurred earlier in the head of the caudate ($M = 10.6$ s) than in the middle frontal areas ($M = 14.6$ s). There was no effect of laterality, $F < 1.0$, and no significant interaction, $P > 0.05$.

Changes in Activity Across the Time Course of Rule Learning in the Hippocampus

Within the right anterior hippocampus (Fig. 3), activity decreased from the beginning of the rule-learning problem and reached its lowest value 15 s later.

Correlations between Frontal, Striatal, and Hippocampal Regions

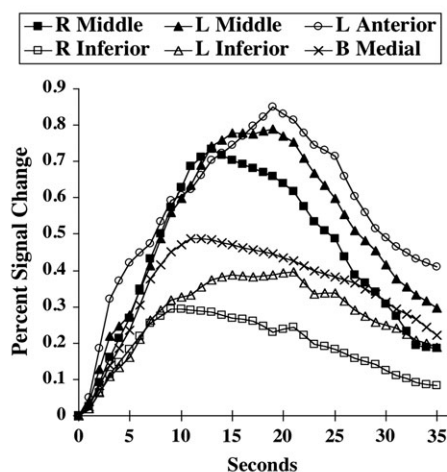
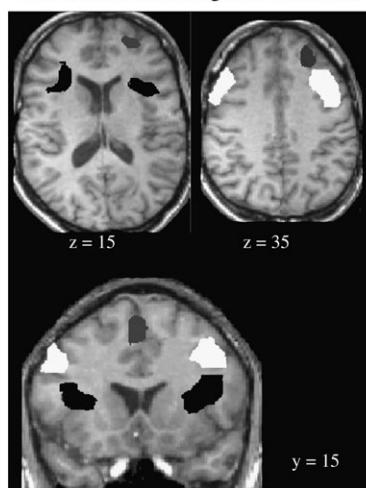
We examined the degree to which peak activity in the ROIs correlated across subjects. For the frontal ROIs we chose the

Table 1
ROIs in prefrontal cortex, striatum, and medial temporal lobe

	BA	x	y	z	Voxels
Rule learning > rule application					
Prefrontal cortex					
L inferior frontal	44, 45	−34	17	19	5236
R inferior frontal	44, 45	33	22	13	3680
L middle frontal	9	−35	19	40	6972
R middle frontal	9	46	21	35	5214
L anterior frontal	9, 10	−22	50	26	2890
B medial frontal	6, 8	1	17	52	3880
Striatum					
R head caudate, putamen		20	−4	18	2568
L head caudate		−6	5	12	309
R body and tail caudate		20	−11	27	3001
L body and tail caudate		−17	−28	30	881
Rule application > rule learning					
Medial temporal lobe					
R anterior hippocampus		28	−9	−14	658

Note: The false discovery rate threshold was set to $q < 0.01$ for regions in the frontal lobe and to $q < 0.05$ for regions in the striatum and medial temporal lobe. L, left; R, right; B, bilateral. BA: Brodmann's areas; x, y, z: coordinates of the centroid of the region in Talairach coordinates; Voxels: volume of the region in mm^3 .

Prefrontal Cortex Regions of Interest



Striatal Regions of Interest

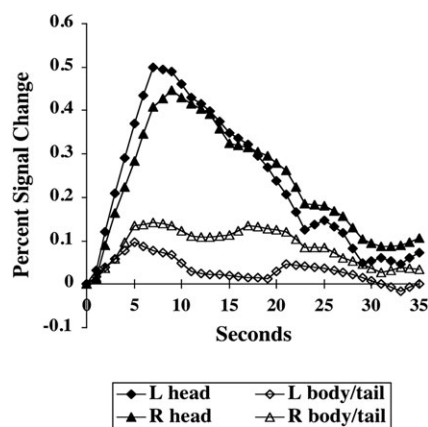
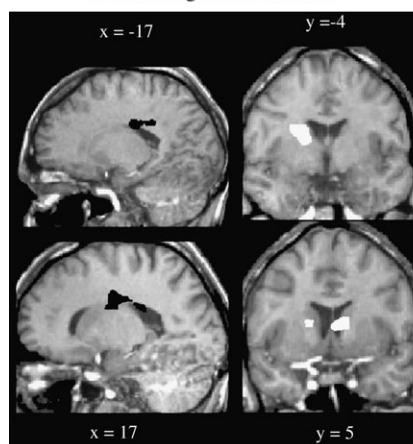


Figure 2. Left top: frontal lobe ROIs defined in Table 1. White: left and right middle frontal. Black: left and right inferior frontal. Dark gray: bilateral medial frontal (bottom coronal slice) and left anterior frontal (top axial slices). Left bottom: striatal ROIs defined in Table 1. Black: left and right body/tail of the caudate. White: left and right head of the caudate. Right: percent signal change across the entire learning problem, including both rule-learning and rule-application phases for frontal ROIs (top) and striatal ROIs (bottom). On average, the transition from rule learning to rule application occurred at approximately 15 s after the start of the problem and would thus be reflected in the BOLD signal at approximately 21–24 s.

Table 2

Additional areas of activation during rule learning and rule application

	BA	x	y	z	Voxels
Rule learning > rule application					
R parietal/lateral occipital	19, 39, 7	35	-61	38	16 388
L parietal/lateral occipital	19, 39, 7	-31	-64	34	25 224
R lingual gyrus	19	29	-58	-3	840
R middle temporal/insula	21	40	-31	-5	1277
R middle temporal	19	34	-77	20	1788
L fusiform gyrus	37	-24	-42	-16	216
L middle occipital	18	-33	-77	12	1075
L superior temporal	39	-38	-57	21	1993
B cerebellum		0	-38	-18	2163
Rule application > rule learning					
L insula		-44	-7	11	158
R insula		35	-3	-3	319

Note: The false discovery rate threshold was set to $q < 0.01$. L, left; R, right; B, bilateral. BA: Brodmann's areas; x, y, z: coordinates of the centroid of the region in Talairach coordinates; Voxels: volume of the region in mm³.

right and left middle frontal and right anterior frontal ROIs for analysis because they showed the most pronounced peaks. For frontal and hippocampal ROIs, which showed relatively late activation peaks, we took as our measure the average percent signal change from 15–19 s across all the voxels in the ROI. For the striatal ROIs, which showed early peaks, we took the average across 7–11 s.

As shown in Table 3, there were significant positive correlations between frontal areas and activity in the left and right head of the caudate. This pattern is consistent with the interconnectivity between the head of the caudate and the frontal cortex in the “cognitive” corticostriatal loop. In contrast, there was no trend toward a correlation, either positive or negative, between the body and tail of the caudate and frontal areas.

Significant negative correlations were found between activity in the hippocampus and activity in the right head of the caudate and left head of the caudate, with a trend toward a negative correlation with right middle frontal gyrus activity.

Right anterior hippocampus

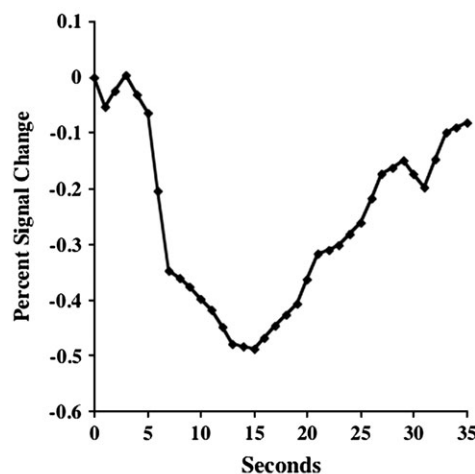
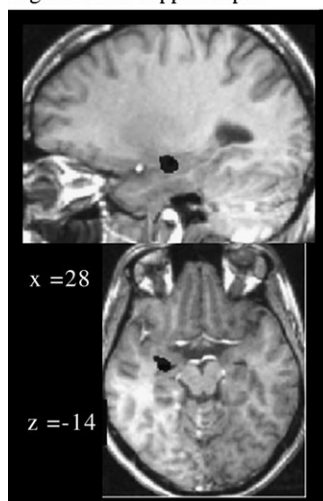


Figure 3. Left: hippocampal ROI defined in Table 1. Right: percent signal change across rule learning for the hippocampal ROI.

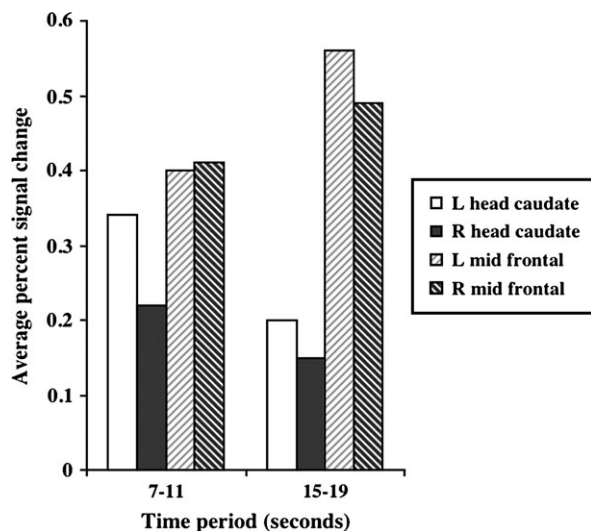


Figure 4. Average percent signal change in left and right head of the caudate and left and right middle frontal gyrus early (7–11 s) and later (15–19 s) in the time course of the rule-learning problems.

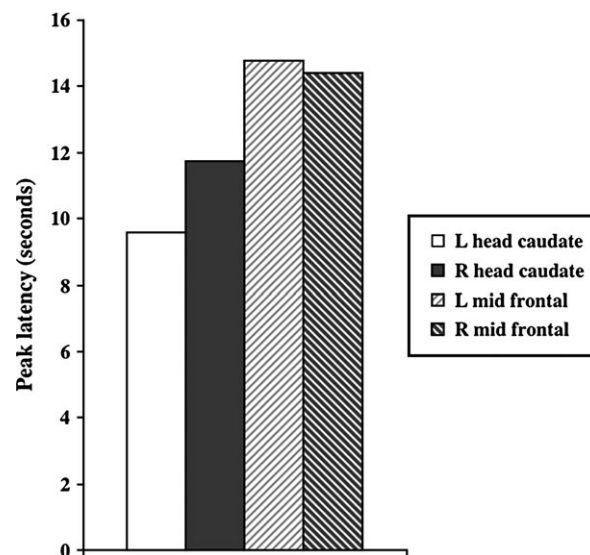


Figure 5. Average BOLD signal peak latency in left and right head of the caudate and left and right middle frontal gyrus across subjects.

Individual Differences in Activity Related to Problem-Solving Ability

Subjects differed in their problem-solving efficiency; as noted in Behavioral Results, the average number of trials required to find the correct rule ranged across subjects from 3.3 to 6.3. Correlation analyses were performed between blood oxygen level dependent activation in the ROIs (as described in the previous section) and mean trial of last error across subjects. Poorer performance (reflected in greater average trial of last error) was positively correlated with activity in the right anterior hippocampus, $r = 0.47$, $P < 0.05$. This pattern of results indicates that subjects who learned slowly were more likely to have higher hippocampal activity. No other striatal or frontal ROIs correlated with accuracy.

Subjects were divided via a median split into 2 groups; the 9 faster learners had a mean trial of last error of 4.3 (range: 3.3–4.9), and the 9 slower learners had a mean trial of last error of

5.4 (range: 5.0–6.3), $t(16) = 5.0$, $P < 0.001$. Faster learners had significantly greater activation in the left tail of the caudate than slower learners, $t(16) = 2.57$, $P < 0.05$, and showed a trend toward greater activation in the right tail of the caudate, $t(16) = 1.91$, $P < 0.1$. No other frontal, striatal, or hippocampal ROI showed a difference in activation between faster and slower learners.

Effects of Positive and Negative Feedback

Trials during the rule-learning phase were divided into those receiving positive feedback (45%) and those receiving negative feedback (55%). A random effects analysis with the false discovery rate set to $q < 0.05$ identified no significant areas of activation. Because of the potential theoretical interest of any differences between positive and negative feedback, we performed an exploratory random effects analysis with a threshold of $P < 0.005$, uncorrected for multiple comparisons. As shown in

Table 3
Correlations of frontal, striatal, and hippocampal activations

	RMF	LMF	Hippocampus
L head caudate	0.53*	0.56*	−0.52*
R head caudate	0.51*	0.46 [^]	−0.60**
L body and tail caudate	−0.05	−0.10	−0.08
R body and tail caudate	0.20	0.20	−0.17
Hippocampus	−0.44 [^]	−0.14	

Note: RMF, right middle frontal; LMF, left middle frontal. [^] $P < 0.1$, * $P < 0.05$, ** $P < 0.01$.

Table 4
Event-related analyses of feedback valence during rule learning

	BA	x	y	z	Voxels
Negative > positive					
B medial frontal	32	−8	35	29	6389
R precuneus	7	11	−46	50	1328
B medial parietal	23, 31	1	−57	19	1328
R inferior parietal lobe	40	56	−37	34	631
R cerebellum		10	−42	−14	709
Positive > negative					
Left putamen		−15	5	−1	416
Right head caudate		15	12	2	237
Right putamen		20	−1	0	145

Note: These comparisons were limited to trials within the rule-learning phase of the Rule task. Statistical threshold was set to $t = 3.2$, $P < 0.005$ uncorrected for multiple comparisons. L, left; R, right; B, bilateral. BA: Brodmann's areas; x, y, z: coordinates of the centroid of the region in Talairach coordinates; Voxels: volume of the region in mm³.

Table 4, this analysis revealed several areas in the striatum that were more active upon receiving positive feedback than negative feedback. These included both the right and left putamen and the head of the caudate. The head of the caudate focus of activity was inferior to that found in the overall analysis that compared rule learning with rule application and bordered on the ventral striatum. Several cortical areas were more active for negative feedback than positive feedback, including bilateral medial frontal areas, several medial and lateral parietal areas, and the cerebellum. It should be noted that in this experiment feedback and correctness are confounded, so these differences may be due to the subjects' producing a correct or incorrect response, rather than just whether they received positive or negative feedback.

We performed an ROI analysis with anatomically defined ROIs of the head and body and tail of caudate nucleus (Seger and Cincotta 2005). These ROIs were defined using a single subjects' high-resolution normalized anatomical image and were confirmed to encompass the structure in all subjects. The ROIs were drawn with generous margins in order to ensure coverage of each structure across subjects but did not extend into the adjacent gray matter areas such as the thalamus, insula, or putamen. The border between the head and the body/tail ROIs was along an oblique plane angled at 45 degrees from horizontal running between the lines defined by $y = 0$, $z = 14$ and $y = 10$, $z = 24$. The head of the caudate ROIs extended inferiorly to $z = -1$ and laterally from the ventricles to $x = \pm 13$. The body/tail of the caudate ROIs extended superiorly to $z = 28$, inferiorly in the body portion to $z = 16$ and in the tail portion to $z = -3$, laterally from approximately $x = 9$ to $x = 22$ in the body region, and posteriorly to $y = -38$ at the tip of the tail. These analyses revealed significantly higher activation in the right head of the caudate for correct feedback trials than incorrect feedback trials, $t(18) = 2.12$, $P < 0.05$, and a trend toward greater activity

in the left head of the caudate, $P = 0.11$. There was no effect of feedback in the body and tail of the caudate ROIs, P values > 0.4 .

Discussion

The results reveal 3 important patterns of frontal, striatal, and hippocampal activity during a rule-learning task. First, striatal and frontal areas were both recruited during rule learning, but they differed in the time course of their recruitment. Striatal activity in the head of the caudate nucleus rose quickly, peaked early, and then rapidly declined, whereas frontal activity reached peak values significantly later, followed by a gradual decline. This pattern is similar to the one found using single-unit recording in monkeys (Pasupathy and Miller 2005). Delgado and others (2005) found greater activity in the caudate when comparing the first 8 trials of classification learning with later trials but it is unknown whether this effect was limited to initial learning trial. The pattern of early striatal with later frontal activity runs counter to theories of frontostriatal interaction that postulate that the striatum is gradually "taught" a procedure by the frontal lobes (Graybiel 1998). One possible alternative is that the head of the caudate may perform an important role early in learning, perhaps in recognizing the context of the situation (Houk and Wise 1995; Frank and others 2001; Frank 2004), which then allows the frontal lobe to choose and implement the proper learning strategy (e.g., hypothesis testing).

The recruitment of frontal areas in this task is consistent with previous studies of the frontal lobes. The middle and inferior frontal gyri are areas that are commonly recruited across tasks with varying demands (Duncan and Owen 2000). Duncan (2001) argues in his adaptive coding model that neurons in these areas dynamically adapt their properties in order to process information relevant to current concerns. In particular, these areas are recruited when working memory demands are high (Curtis and D'Esposito 2003). The medial frontal gyrus/Brodmann's areas 8, part of the supplementary motor area, has been linked in previous studies to hypothesis generation (Elliot and Dolan 1998). Volz and others (2003) argue that this area is active across tasks that involve judgment under uncertainty. Anterior prefrontal cortex activity has been found previously in rule-learning tasks associated with a switch to a new rule (Strange and others 2001). The function of the anterior prefrontal cortex has been variably characterized as processing of branching and subgoals (Koechlin and others 1999; Badre and Wagner 2004), relational integration (Christoff and others 2001), and evaluation of internally generated information (Christoff and others 2003).

Second, we found an antagonistic relationship between frontostriatal and hippocampal activity. Activity in the hippocampus declined as activity in frontostriatal systems increased across the course of rule learning. Across subjects, faster rule learning was associated with greater activity in the body and tail of the caudate and less activity in the hippocampus. The relative reduction of activity in the hippocampus—lower in rule learning and higher during rule application—can be interpreted in 2 ways. First, it may reflect suppression of the hippocampus during rule learning followed by recovery during rule application. Second, it may reflect specific recruitment during rule application, perhaps in encoding and/or retrieving the selected rule or in greater amounts of incidental "resting state" conceptual processing occurring during the relatively simpler rule-application task (Stark and Squire 2001; Law and others 2005).

The sharp decrease in activity at the beginning of rule learning, coupled with the gradual recovery into the rule-application phase, suggests the former interpretation, but it is not possible to distinguish between them with the present data.

Hippocampal-striatal antagonism has previously been found in implicit probabilistic classification tasks (Poldrack and others 1999, 2001). Implicit tasks can be performed without the intention to learn, and the information learned need not be completely accessible to consciousness (Seger 1994). In contrast, rule learning is usually considered an explicit task: it is completely intentional, and all the rules are fully available to consciousness. Although rule learning is explicit, it differs from explicit declarative memory processes that are reliant on the medial temporal lobe in that it involves abstracting commonalities across items, rather than learning individual items. Amnesics with damage to the medial temporal lobe are able to learn rules via hypothesis testing (Janowsky and others 1989) as long as their working memory is intact. The present results provide evidence that striatal systems are important for learning rules and associations regardless of whether the tasks are implicit or explicit. However, it should be noted that rule learning and implicit classification tasks have been dissociated behaviorally (Ashby and Maddox 2005) and are likely to differ in recruitment of other neural systems.

Activity in the bilateral head of the caudate, but not the body and tail, was negatively correlated with hippocampal activity. These correlations indicate that reduction in hippocampal activity during learning may be mediated by the corticostriatal loop connecting the head of the caudate with dorsolateral prefrontal cortex rather than the loop connecting the body and tail of the caudate with the inferior temporal lobe. In contrast, activity in the body and tail of the caudate was related to successful learning. This pattern of results implies that the antagonism between striatum and medial temporal lobe is not due to direct competition between the learning systems but rather is mediated by executive functions subserved by the frontal lobe and head of the caudate and is consistent with results reported by Poldrack and Rodriguez (2004).

Finally, we found evidence that 2 subregions of the striatum, the head and body and tail of the caudate, contribute to rule learning in dissociable ways. Both the head and body and tail regions were more active during rule learning than rule application. However, the body and tail was associated with successful learning: activity in the body and tail remained elevated across rule learning. Furthermore, it was the only brain region that showed higher activity for fast rule learners than slow rule learners. This association of body and tail activity with successful learning is similar to results found in a simple classification task (Seger and Cincotta 2005).

In contrast, activity in the head did not follow the behavioral indexes in learning but instead showed patterns consistent with it being involved in executive functions required for rule learning. The high and transient activity in the head of the caudate at the beginning of each rule-learning task implies that it is involved in executive functions, possibly setting context, important for initiating the rule-learning process. Additionally, the negative correlation between the head of the caudate and medial temporal lobe activity implies that the head of the caudate helps coordinate interactions between the learning systems. Finally, the greater activity in the head of the caudate to positive feedback than to negative feedback indicates that it plays a role in evaluating and responding to feedback.

Several previous studies have found that activation in the head of the caudate is sensitive to the valence of feedback. Typically, activation is greater to positive than negative feedback (Delgado and others 2000; Filoteo, Maddox, Simmons and others 2005; Seger and Cincotta 2005), though one study found greater caudate activation to negative than positive feedback (Monchi and others 2001) and others greater activation to positive than negative feedback. This may reflect differences across tasks in what feedback is expected; many studies have shown the caudate to be sensitive to expectancy (Tremblay and others 1998). In addition to the head of the caudate, several cortical areas were affected by feedback valence but in the opposite direction with greater activity for negative than positive feedback. Of particular note is the medial frontal cortex around the anterior cingulate. This area has often been shown to be sensitive to negative feedback and is the proposed neural generator of the error-related negativity event-related potential component (Ullsperger and von Cramon 2003).

Notes

This work supported by a grant from the National Institute of Mental Health (MH63784) to CAS. We would like to thank John Gabrieli and Richard M. Lucas Center for Magnetic Resonance Spectroscopy and Imaging at Stanford University for assistance and support in performing this study. We would also like to thank Ed Delosh who first introduced us to the task developed by Marvin Levine.

Address correspondence to Carol A. Seger, Department of Psychology, Colorado State University, Fort Collins, CO 80523, USA. Email: seger@lamar.colostate.edu.

References

- Alexander GE, DeLong MR, Strick PL. 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9:357-381.
- Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. 1998. A neuropsychological theory of multiple systems in category learning. *Psychol Rev* 105:442-481.
- Ashby FG, Maddox WT. 2005. Human category learning. *Annu Rev Neurosci* 56:149-178.
- Badre D, Wagner AD. 2004. Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron* 41:473-487.
- Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207-4221.
- Channon S, Jones M-C, Stephenson S. 1993. Cognitive strategies and hypothesis testing during discrimination learning in Parkinson's disease. *Neuropsychologia* 31:75-82.
- Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, Holyoak KJ, Gabrieli JDE. 2001. Rostrocaudal prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14:1136-1149.
- Christoff K, Ream JM, Geddes LPT, Gabrieli JDE. 2003. Evaluating self-generated information: anterior prefrontal contributions to human cognition. *Behav Neurosci* 117:1161-1168.
- Cicerone K, Lazar R, Chapiro W. 1983. Effects of frontal lobe lesions on hypothesis sampling during concept formation. *Neuropsychologia* 21:513-524.
- Curtis CE, D'Esposito M. 2003. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci* 7:415-423.
- Delgado MR, Miller MM, Inati S, Phelps EA. 2005. An fMRI study of reward-related probability learning. *Neuroimage* 24:862-873.
- Delgado MR, Nystrom LC, Fissell C, Noll DC, Fiez JA. 2000. Tracking the hemodynamic responses to reward and punishment. *J Neurophysiol* 84:3072-3077.

- Duncan J. 2001. An adaptive coding model of neural function in prefrontal cortex. *Nat Rev Neurosci* 2:820-829.
- Duncan J, Owen AM. 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23:475-483.
- Elliot R, Dolan RJ. 1998. Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *Neuroimage* 8:17-29.
- Fernandez-Ruiz J, Wang J, Aigner TG, Mishkin M. 2001. Visual habit formation in monkeys with neurotoxic lesions of the ventrocaudal neostriatum. *Proc Natl Acad Sci USA* 98:4196-4201.
- Filoteo JV, Maddox WT, Davis J. 2001. A possible role of the striatum in linear and nonlinear categorization rule learning: evidence from patients with Huntington's disease. *Behav Neurosci* 115:786-798.
- Filoteo JV, Maddox WT, Salmon DP, Song DD. 2005. Information-integration category learning in patients with striatal dysfunction. *Neuropsychology* 19:212-222.
- Filoteo JV, Maddox WT, Simmons AN, Ing AD, Cagigas XE, Matthews S, Paulus MP. 2005. Cortical and subcortical brain regions involved in rule-based category learning. *Neuroreport* 16:111-115.
- Frank MJ. 2004. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J Cogn Neurosci* 17:51-72.
- Frank MJ, Loughry B, O'Reilly RC. 2001. Interactions between frontal cortex and basal ganglia in working memory: a computational account. *Cogn Affect Behav Neurosci* 1:137-160.
- Genovese CR, Lazar NA, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870-878.
- Graybiel AM. 1998. The basal ganglia and chunking of action repertoires. *Neurobiol Learn Mem* 70:119-136.
- Houk JC, Wise SP. 1995. Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. *Cereb Cortex* 5:95-110.
- Janowsky JS, Kritchevsky AP, Squire LR. 1989. Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behav Neurosci* 103:548-560.
- Keri S, Beniczky S, Voros E, Janka Z, Benedek G, Vecsel L. 2002. Dissociation between attentional set shifting and habit learning: a longitudinal case study. *Neurocase* 8:219-225.
- Keri S, Szobodnyik C, Benedek G, Janka Z, Gadoros J. 2001. Probabilistic classification learning in Tourette syndrome. *Neuropsychologia* 135:1-7.
- Knowlton BK, Mangels JA, Squire LR. 1996. A neostriatal habit learning system in humans. *Science* 273:1399-1402.
- Knowlton BK, Squire LR, Paulsen JS, Swerdlow NR, Swenson M, Butters N. 1996. Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology* 10:538-548.
- Koechlin E, Basso G, Pietrini P, Panzer S, Grafman J. 1999. The role of anterior prefrontal cortex in human cognition. *Nature* 399:148-151.
- Law JR, Flanery MA, Wirth H, Yanike M, Smith AC, Frank LM, Suzuki WA, Brown EN, Stark CEL. 2005. Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. *J Neurosci* 25:5720-5729.
- Lawrence AD, Sahakian BJ, Robbins TW. 1998. Cognitive functions and corticostriatal circuits: insights from Huntington's disease. *Trends Cogn Sci* 2:379-388.
- Levine MA. 1975. *Cognitive theory of learning*. Hillsdale, NJ: Lawrence Erlbaum.
- Maddox WT, Filoteo JV. 2001. Striatal contributions to category learning: quantitative modeling of simple linear and complex non-linear rule learning in patients with Parkinson's disease. *J Int Neuropsychol Soc* 7:710-727.
- Middleton FA, Strick PL. 2000. Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res Rev* 31:236-250.
- Milner B. 1963. Effects of brain lesions on card sorting. *Arch Neurol* 9:90-100.
- Mishkin M, Malamut B, Bachevalier J. 1984. Memories and habits: two neural systems. In: Lynch G, McGaugh JL, Weinberger NM, editors. *Neurobiology of learning and memory*. New York: Guilford Press. p 65-77.
- Monchi O, Petrides M, Petre V, Worsley K, Dagher A. 2001. Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J Neurosci* 21:7733-7741.
- Myers CE, Shohamy D, Gluck MA, Grossman S, Kluger A, Ferris S, Golomb J, Schnirman G, Schwartz R. 2003. Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *J Cogn Neurosci* 15:185-193.
- Packard MG, Knowlton BJ. 2002. Learning and memory functions of the basal ganglia. *Annu Rev Neurosci* 25:563-593.
- Parent A, Hazrati L-N. 1995. Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res Rev* 20:91-127.
- Pasupathy A, Miller EK. 2005. Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873-876.
- Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA. 2001. Interactive memory systems in the human brain. *Nature* 414:546-550.
- Poldrack RA, Packard MG. 2003. Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia* 41:245-251.
- Poldrack RA, Prabhakaran V, Seger CA, Gabrieli JDE. 1999. Striatal activation during cognitive skill learning. *Neuropsychology* 13:564-574.
- Poldrack RA, Rodriguez P. 2004. How do memory systems interact? Evidence from human classification learning. *Neurobiol Learn Mem* 82:324-332.
- Preston AR, Thomason ME, Ochsner KN, Cooper JC, Glover GH. 2004. Comparison of spiral-in/out and spiral-out BOLD fMRI at 1.5T and 3T. *Neuroimage* 21:291-301.
- Rao SM, Bobholz JA, Hammeke TA, Rosen AC, Woodley SJ, Cunningham JM, Cox RW, Stein EA, Binder JR. 1997. Functional MRI evidence for subcortical participation in conceptual reasoning skills. *Neuroreport* 8:1987-1993.
- Reverberi C, D'Agostini SD, Skrap M, Shallice T. 2005. Generation and recognition of abstract rules in different frontal lobe subgroups. *Neuropsychologia* 43:1924-1937.
- Sage JR, Anagnostaras SG, Mitchell S, Bronstein JM, De Salles A, Masterman D, Knowlton BJ. 2003. Analysis of probabilistic classification learning in patients with Parkinson's disease before and after pallidotomy surgery. *Learn Mem* 10:226-236.
- Saint-Cyr JA, Ungerleider LG, Desimone R. 1990. Organization of visual cortical inputs to the striatum and subsequent outputs to the pallidum-nigral complex in the monkey. *J Comp Neurol* 298:129-165.
- Seger CA. 1994. Implicit learning. *Psychol Bull* 115:163-196.
- Seger CA, Cincotta CM. 2002. Striatal activation in concept learning. *Cogn Affect Behav Neurosci* 2:149-161.
- Seger CA, Cincotta CM. 2005. The roles of the caudate nucleus in human classification learning. *J Neurosci* 25:2941-2951.
- Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA. 2004. Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain* 127:851-859.
- Stark CEL, Squire LR. 2001. When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc Natl Acad Sci USA* 98:12760-12766.
- Strange BA, Henson RNA, Friston KJ, Dolan RJ. 2001. Anterior prefrontal cortex mediates rule learning in humans. *Cereb Cortex* 11:1040-1046.
- Stuss DT, Levine B, Alexander MP, Hong J, Palumbo C, Hamer L, Murphy KJ, Izukawa D. 2000. Wisconsin card sorting test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia* 38:388-402.
- Tailarach J, Tournoux P. 1988. *Co-planar stereotactic atlas of the human brain*. New York: Thieme.
- Toni I, Ramnani N, Josephs O, Ashburner J, Passingham RE. 2001. Learning arbitrary visuo-motor associations: temporal dynamic of brain activity. *Neuroimage* 14:1048-1057.
- Toni I, Rowe J, Stephan KE, Passingham RE. 2002. Changes of cortico-striatal effective connectivity during visuomotor learning. *Cereb Cortex* 12:1040-1047.

- Tremblay L, Hollerman JR, Schultz W. 1998. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J Neurophysiol* 80:964-977.
- Ullsperger M, von Cramon DY. 2003. Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *J Neurosci* 23:4308-4314.
- Volz KG, Schubotz RI, von Cramon DY. 2003. Predicting events of varying probability: uncertainty investigated. *Neuroimage* 19:271-280.
- Webster MJ, Bachevalier J, Ungerleider LG. 1993. Subcortical connections of inferior temporal areas TE and TEO in macaque monkeys. *J Comp Neurol* 335:73-91.
- Witt K, Nuhman A, Deuschl G. 2002. Dissociation of habit-learning in Parkinson's and cerebellar disease. *J Cogn Neurosci* 14:493-499.
- Yeterian EH, Pandya DN. 1995. Corticostriatal connections of extrastriate visual areas in rhesus monkeys. *J Comp Neurol* 352:436-457.