

Cortical and subcortical brain regions involved in rule-based category learning

J. Vincent Filoteo,^{1,2,CA} W. Todd Maddox,^{3,4} Alan N. Simmons,¹ A. David Ing,³ Xavier E. Cagigas,⁵ Scott Matthews¹ and Martin P. Paulus¹

¹Department of Psychiatry, University of California, San Diego, CA; ²Psychology Service 116B, Veterans Administration San Diego Healthcare System, 3350 La Jolla Village Dr, San Diego, CA; ³Department of Psychology, University of Texas, Austin, TX; ⁴Institute of Neuroscience, University of Texas, Austin, TX; ⁵SDSU/UCSD Joint Doctoral Program in Clinical Psychology, USA

^{CA}Corresponding Author: vfiloteo@ucsd.edu

Received 19 October 2004; accepted 1 November 2004

The brain regions contributing to rule-based category learning were examined using fMRI. Participants categorized single lines that varied in length and orientation into one of two categories. Category membership was based on the length of the line. Results indicated that left frontal and parietal regions were differentially activated in those participants who learned the task as compared to those who did not. Further, the head of the caudate displayed

relative decreases in activation on incorrect trials relative to correct trials. The involvement of this latter structure is likely related to (1) processing an error signal, or (2) volitional switching between potential category rules. Results are consistent with theories suggesting that a frontal-striatal circuit is involved in rule-based category learning. *NeuroReport* 16:111–115 © 2005 Lippincott Williams & Wilkins.

Key words: Category learning; Caudate nucleus; Frontal cortex; Parietal cortex

INTRODUCTION

The ability to categorize is a fundamental process that is important for many aspects of our daily functioning. Moreover, it is critical to acquire and establish new categories throughout our lives. New categories are often acquired via the process of hypothesis-testing, a process in which a participant generates a rule that can be used to establish category membership, applies this rule to stimuli in order to associate them with a category, uses feedback to determine whether the association was as proposed, and subsequently examines whether the hypothesis was correct. If the participant's hypothesis is correct, he/she must maintain that rule, whereas if they are incorrect, they must switch to a new rule. Hypothesis-testing approaches to category learning are often invoked when the category structures to be learned are rule-based [1]. In rule-based tasks, optimal performance is often based on the participant placing a decision criterion on a single stimulus dimension and ignoring the other, irrelevant dimensions. Previous behavioral studies have suggested that working memory and/or selective attention processes are necessary for the learning of rule-based categories [2,3]. As such, one would expect that brain structures that are involved in working memory and/or selective attention processes, such as dorsolateral frontal and parietal regions, would also be involved when participants learn rule-based categories, a prediction that has been supported by prior functional imaging studies [4–7].

In addition to cortical areas, subcortical structures such as the caudate nucleus have also been hypothesized to play a

critical role in category learning [2]. Support for this hypothesis comes from a recent functional imaging study by Monchi and colleagues [5] that found that the caudate is differentially activated on incorrect *vs* correct trials when participants performed the Wisconsin Card Sorting Test (WCST). The WCST is a clinical measure of feedback-based learning because the participant must generate potential hypotheses, test a specific hypothesis, maintain that hypothesis if they receive positive feedback (correct trials), or change to a new hypothesis if they receive negative feedback (error trials). This task is also considered to be a rule-based category learning task in that the correct rule is verbalizable.

In the present study, we investigated the acquisition of rule-based categories using fMRI while individuals performed the perceptual categorization task (PCT) [8], a task that has been used extensively in behavioral research (see [1] for a review). In the PCT, participants are presented with simple perceptual stimuli and asked to categorize each as belonging to one of two categories. The participants are then given feedback immediately following their response. Participants in this study were presented with single lines that varied in length and orientation, and they had to learn that category membership was based on the length of the line (Fig. 1). Based on the behavioral results, we identified a subgroup of individuals who learned the rule (learners) and subgroup who did not (non-learners). By contrasting the pattern of activation of these two subgroups, we identified those brain regions involved in acquiring rule-based categories. Based on previous research [4–7], we anticipated

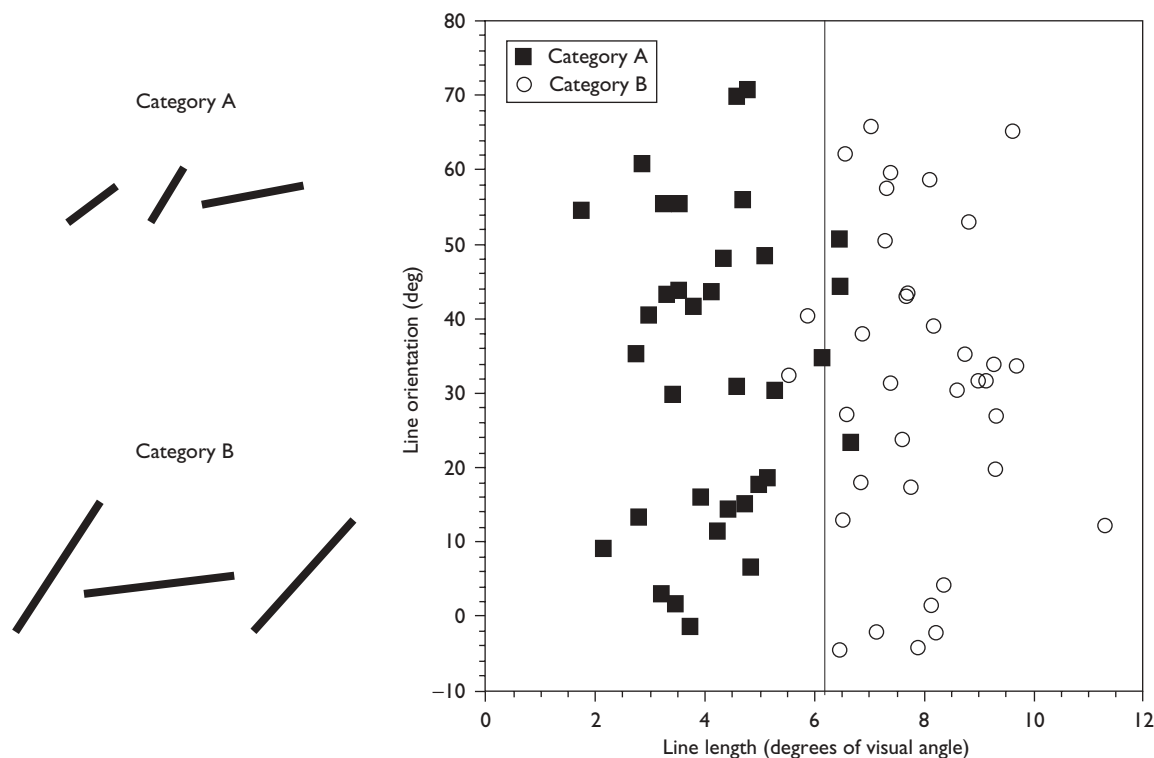


Fig. 1. (a) Sample stimuli used in the perceptual categorization task, and (b) stimulus distributions for the rule-based categorization task. Line lengths are in degrees of visual angle. Line orientation units are in degrees, with zero degrees representing horizontal. The squares denote individual exemplars from Category A, whereas the circles represent individual exemplars from Category B. The solid line denotes the experimenter-defined (optimal) categorization rule that maximized long-run accuracy. The solid line represents the optimal decision bound. The mean values on the two dimensions for the two categories were the following: Category A: length $\mu=4.1$, $\sigma=1.2$, orientation $\mu=30.0$, $\sigma=20.0$; Category B: length $\mu=8.0$, $\sigma=1.2$, orientation $\mu=30.0$, $\sigma=20.0$.

greater activation in the dorsolateral frontal and parietal regions in those individuals who successfully learned the categories. Further, based on the research of Monchi *et al.* [5], we also compared activation patterns between incorrect and correct trials in the learners. This comparison enabled us to determine the brain structures that are differentially involved in the decision-making process(es) during the performance of a rule-based category learning task when a participant responds incorrectly *vs* correctly. Consistent with the findings of Monchi *et al.* we predicted that the caudate would show differential activation during incorrect trials as compared to correct trials.

SUBJECTS AND METHODS

Participants: Fifteen healthy, right-handed individuals signed an informed consent approved by the UCSD Human Research Protection Program and completed this study (10 male and 5 female). Their mean (\pm s.d.) age was 40.73 ± 9.15 years, range 28–56 and mean level of education was 14.87 ± 1.64 years, range 12–18. All participants were screened for a history of psychiatric and neurological disorders.

Perceptual categorization task and procedure: This task is analogous to the behavioral tasks previously described in Ashby and Gott [8]. Briefly, the PCT and a comparator task were presented in a block design along with periods of rest. The stimuli consisted of line segments of varying lengths

and orientations (Fig. 1a). The distribution of the line stimuli for the PCT in terms of length and orientation is displayed in Fig. 1b. Optimal responding required that the participant set a criterion on the length dimension and ignore the orientation dimension. However, participants were not told about this rule and had to learn it based on trial-by-trial feedback provided after each of their responses (see below). Stimuli were white and presented on a black background using a PC. On average, the stimuli subtended a visual angle of 6° .

The comparator task consisted of presenting the same line stimuli to participants, except that the lines were randomly colored blue or yellow, and the participant's task was to identify the color by pressing one button if the line color was blue, or another button if the line color was yellow. Participants were told that when the lines were colored, they simply had to decide whether the line was blue or yellow, but that when the line was white, they had to figure out on their own to which category the line belonged. Thus, in the comparator task, participants were told what constituted a correct response prior to being exposed to the stimuli.

Trial runs for the PCT and comparator task consisted of 6 trials per block. Seventy two trials were given for both the PCT and the comparator task. For each trial a fixation cross was presented for 250 ms followed by a 50 ms blank screen before a sample stimulus was shown. The trial lasted 4000 ms and the sample stimulus was removed from the screen as soon as the individual made a response. Once a

response was made, the word 'correct' or 'wrong' appeared immediately for up to 750 ms or until the trial was over. Prior to entering the scanner, each participant was given pre-training on a parallel version of this task without any information about the rule that was to be learned in the task during MRI scanning.

fMRI: Blood oxygenation level dependent (BOLD) fMRI signal data were collected for each participant using a 1.5 T Siemens (Erlangen, Germany) scanner (T2*-weighted echo planar imaging, TR=2000 ms, TE=40 ms, 64 × 64 matrix, 20 4 mm axial slices, 256 scans) while participants performed the PCT and comparator task. fMRI volume acquisitions were time-locked to the onset of each trial. During the same experimental session, a T1-weighted image (MPRAGE, TR=11.4 ms, TE=4.4 ms, flip angle=10°, FOV=256 × 256, 1 mm³ voxels) was obtained for anatomical reference. For preprocessing, voxel time series were interpolated to correct for non-simultaneous slice acquisition within each volume and corrected for 3D motion.

fMRI analysis pathway: The data were preprocessed and analyzed with the software AFNI [9]. The echo planar images were realigned to the 128th acquired scan and time corrected for slice acquisition order. To exclude the voxels showing an artifact-related signal drop, a combined threshold/cluster-growing algorithm was applied to the average image intensity of all echoplanar images to compute a whole-brain mask. Two primary analyses were conducted. In the first analysis, two orthogonal regressors of interest were contrasted: (1) the PCT trials, and (2) the comparator discrimination task trials. These regressors were convolved with a modified gamma variate function modeling a prototypical hemodynamic response [10] prior to inclusion in the regression model. Three regressors were used to model residual motion (in the roll, pitch, and yaw directions). Regressors for baseline and linear trends were used to estimate slow signal drifts. The AFNI program 3dDeconvolve was used to calculate the estimated voxel-wise response amplitude. A Gaussian filter with FWHM 4 mm was applied to the voxel-wise percent signal change data to account for individual variations in anatomical landmarks. Data of each participant were oriented to Talairach coordinates. The voxel-wise percent signal change data, which was obtained from the estimated beta weight of the regressors divided by the beta weight of the baseline regressor, were entered into a mixed-model ANOVA with task condition (correct, incorrect, and comparator trials) as a fixed factor and participants as a random factor. A threshold adjustment procedure based on Monte-Carlo simulations, which was based on the whole brain mask with a connection radius of 4 mm, a voxelwise probability of $p < 0.05$, and a *posteriori* probability of $p < 0.05$, was used to obtain a minimum volume of 512 μl in order to guard against identifying false positive areas of activation [11]. Participants were then divided based on their behavioral performance on the PCT task into learners *vs* non-learners (see below) and their activation patterns during the PCT trials and the comparator trials were submitted to a mixed-model ANOVA with fixed factors of group (learners *vs* non-learners) by task (PCT *vs* comparator) and participants as a random factor. This analysis enabled us to determine the

brain regions that differentiated those participants who learned the rule *vs* those who did not.

The pathway for the second analysis was the same as that described in the first, except that incorrect and correct trials were divided into two separate regressors. Thus, three orthogonal regressors of interest were contrasted: (1) the correct PCT trials, (2) incorrect PCT trials, and (3) the comparator discrimination task trials. This analysis was conducted in only those individuals who adequately learned the task (see below for details) and allowed us to directly contrast activation patterns associated with correct and incorrect trials.

RESULTS

Behavioral results: A χ^2 test determined that above chance performance was $> 66.7\%$ (> 48 correct responses out of 72 trials) for an individual participant at the $p < 0.05$ level. An initial evaluation of the accuracy data revealed that 8 out of the 15 participants performed above chance throughout the task. These 8 individuals were designated learners and those 7 who performed below chance were designated non-learners. The mean accuracy rates (percent correct) for the learners and non-learners are displayed in Fig. 2 for the 72 trials in 6 trial blocks. A repeated measures ANOVA revealed that learners were more accurate than non-learners ($F=26.99$, $p < 0.001$), that both groups improved across the trial blocks ($F=4.72$, $p < 0.001$), and that learners improved more than non-learners across trial blocks ($F=1.90$, $p < 0.05$). All participants were 100% accurate on the comparator task.

Neuroimaging results: In the first analysis, learners and non-learners were contrasted in order to determine more completely which brain regions were needed to learn the rule-based categories. As shown in Fig. 3, three regions were identified in which there were greater activation differences between the categorization relative to the comparator task for those participants that learned the task as compared to those participants who did not (i.e., in which there was a group × task interaction). These included the left middle frontal gyrus (BA 9; center of mass coordinates: -38, 24, 34, volume 576 μl), the left supramarginal gyrus (BA 40; center of mass coordinates: -41, -46, 36, volume 640 μl), and the right precentral gyrus (BA 6; center of mass coordinates 17, -25, 64, volume 576 μl).

In the second analysis examining incorrect minus correct trials in those participants who performed above chance, the

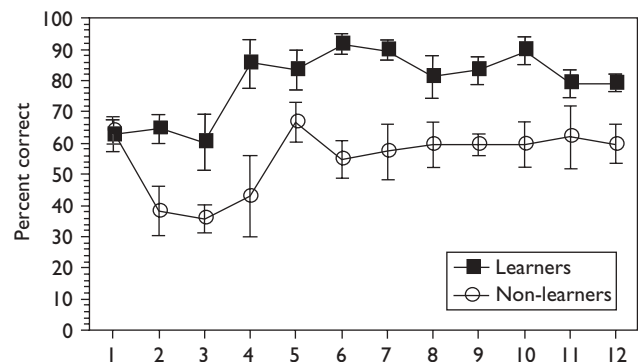


Fig. 2. Mean accuracy rates for the learners and non-learners on the perceptual categorization task. Error bars are s.e.m.

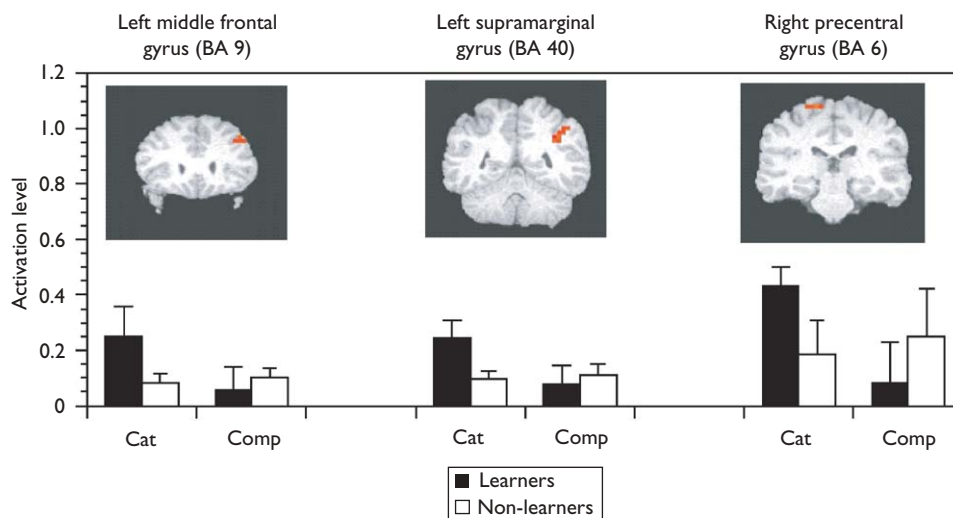


Fig. 3. Activation levels for learners and non-learners associated with the categorization (Cat) and comparator (Comp) tasks, and corresponding brain regions. See text for center of mass coordinates. Error bars are s.e.m.

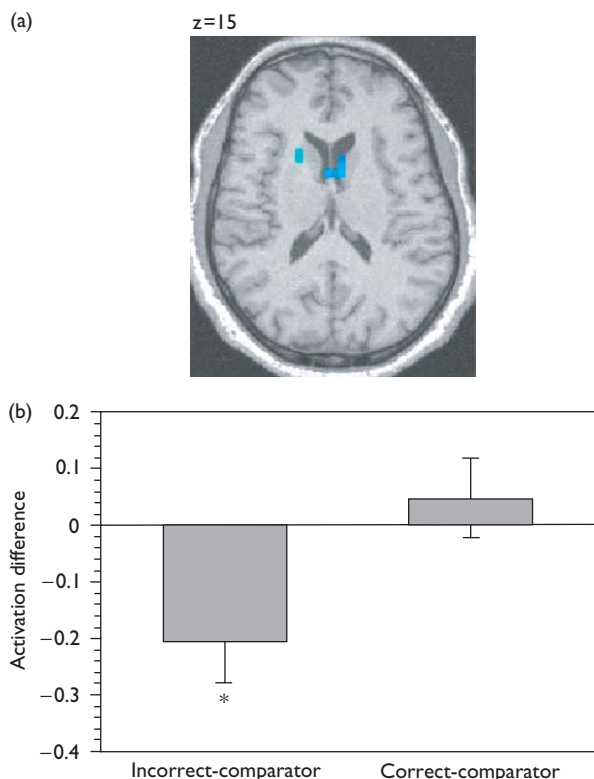


Fig. 4. (a) Activity of significant clusters in the head of the caudate for correct-incorrect trials on the PCT. (b) Activation differences between incorrect trials on the PCT minus the comparator task, and between correct trials on the PCT minus the comparator task. *Significantly different from zero ($p < 0.05$). Error bars are s.e.m.

only brain region to survive the threshold was the head of the caudate bilaterally (center of mass coordinates 1,8,15, volume 1664 μ l; Fig. 4a). An examination of the changes in activation relative to the comparator task indicated that this difference was due to reliable decreases in activity in the head of the caudate on incorrect trials ($t(7)=2.9$, $p < 0.05$)

whereas there was non-significant change in activation in this region during correct trials ($t(7)=0.6$, $p=0.52$; Fig. 4b).

DISCUSSION

This study examined the brain regions involved in learning rule-based categories. Correct responding in this study required that the participant learn to set a criterion on the length dimension of the stimulus and ignore the irrelevant variation on the orientation dimension. Behavioral results indicated that only about half of the participants were able to learn the rule-based task. The primary neuroimaging results indicated that (1) frontal and parietal regions were differentially activated during the performance of the categorization task in those who learned the rule *vs* those who were unable, and (2) in those participants who learned the rule, the head of the caudate displayed a different level of activation when correct trials were subtracted from incorrect trials.

The differential pattern of activation we observed in the learners and non-learners was exactly what would be expected if in fact these brain regions were involved in learning the categories. Specifically, greater activation was observed in learners as compared to non-learners in the dorsolateral prefrontal cortex and posterior parietal cortex during the categorization task relative to the comparator task, whereas there was no difference between the two groups in terms of their activation levels in these regions during the comparator task (Fig. 3). Importantly, these brain regions have been implicated in working memory [12,13], and given that this is an important process involved in rule-based category learning, these findings are consistent with our understanding of dorsolateral and parietal functioning. Furthermore, such findings are also consistent with other studies that have implicated these brain regions in rule-based category learning [4,7,14,15].

An examination of the relative signal change indicated that the caudate experienced decreased activation on incorrect trials relative to the comparator task, whereas there was relatively no change in this brain region on correct trials. Interestingly, previous studies of rule-based category

learning have identified increased activation of the caudate during the performance of rule-based categorization tasks [5–7]. For example, Monchi *et al.* [5] found that the caudate was differentially activated following negative feedback (indicating an incorrect response had been made) when participants performed a rule-based task relative to a control task, whereas this was not the case following positive feedback (i.e. after a correct response). These results are somewhat in contrast to those observed in the present study in that we identified relative decreased activation in the caudate. However, one important difference between the two previous studies and this current study is that participants in the former studies were trained in the categorization task prior to entering the scanner. Although we pre-trained our participants as to the nature of the categorization task, they were not exposed to the actual rule until they were in the scanner. Thus, any differences observed between our results and previous results could be due to the fact that participants in our study were actually required to learn the categorization rule.

Given the finding of differential caudate activation levels during incorrect *vs* correct trials, it appears that this structure could be contributing to one of two different processes: (1) interpreting the negative feedback following an incorrect response, or (2) switching to an alternative rule. In regard to the first possibility, several studies have indicated that this brain region is involved in processing errors when participants perform reward-based learning tasks [16,17]. As such, the caudate has been considered to be part of larger network of subcortical structures involved in processing punishment and reward during learning, so it is very possible that the change in activation we observed in this brain region is secondary to the processing of feedback following an incorrect response.

In regard to the other possibility, that the head of the caudate is involved in rule switching, past research has implicated this brain structure in cognitive switching [18]. Ashby and colleagues [1,2], for example, proposed that the head of the caudate, in conjunction with prefrontal regions, is involved in volitional switching between potential rules during rule-based category learning. In particular, these investigators argue that decreased caudate activation is necessary to disengage from a previously adopted rule (represented by activity between thalamic and frontal structures) so that a new rule can be engaged. Our finding of decreased activation in the head of the caudate (relative to baseline) following incorrect responses provides preliminary evidence in support of this theory, and is consistent with past studies that have demonstrated increased caudate activity under conditions of decreased response switching [18]. However, one important question for future research is whether the decreased activation we observed resulted in the incorrect response, or whether this pattern of activity was in response to the incorrect response.

In conclusion, differential activation was observed between those participants who learned a rule-based category

learning task *vs* those who did not in frontal and parietal regions known to be involved in working memory. Furthermore, the head of the caudate was found to be associated with incorrect responding in participants who successfully learned the rule. The involvement of this structure could be due either to processing negative feedback or in initiating a volitional rule switch.

REFERENCES

- Ashby FG and Maddox WT. Human category learning. *Annu Rev Psychol*, in press.
- Ashby FG, Alfonso-Reese LA, Turken AU and Waldron EM. A neuropsychological theory of multiple systems in category learning. *Psychol Rev* 1998; **105**:442–481.
- Waldron EM and Ashby FG. The effects of concurrent task interference on category learning. *Psychon Bull Rev* 2001; **8**:168–176.
- Grossman M, Smith EE, Koenig P, Glosser G, DeVita C, Moore P *et al.* The neural basis for categorization in semantic memory. *Neuroimage* 2002; **17**:1549–1561.
- Monchi O, Petrides M, Petre V, Worsley K and Dagher A. Wisconsin Card Sorting Test revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J Neurosci* 2001; **21**:7733–7741.
- Rao SM, Bobholz JA, Hammelke TA, Rosen AC, Woodley SJ, Cunningham JM *et al.* Functional MRI evidence for subcortical participation in conceptual reasoning skills. *Neuroreport* 1997; **8**:1987–1993.
- Seger CA and Cincotta CM. Striatal activity in concept learning. *Cogn Affect Behav Neurosci* 2002; **2**:149–161.
- Ashby FG and Gott RE. Decision rules in the perception and categorization of multidimensional stimuli. *J Exp Psychol Learn Mem Cogn* 1988; **14**:33–53.
- Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 1996; **29**: 162–173.
- Boynton GM, Engle SA, Glover GH and Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 1996; **16**:4207–4221.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA and Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 1995; **33**:636–647.
- Awh E, Smith EE and Jonides J. Human rehearsal processes and the frontal lobes: PET evidence. *Ann NY Acad Sci* 1995; **769**:91–117.
- D'Esposito M, Ballard D, Zarahn E and Aguirre GK. The role of prefrontal cortex in sensory memory and motor preparation: an event-related fMRI study. *Neuroimage* 2000; **11**:400–408.
- Fletcher P, Buchell C, Josephs O, Friston K and Dolan R. Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebr Cortex* 1999; **9**:168–178.
- Patalano AL, Smith EE, Jonides J and Koeppel RA. PET evidence for multiple strategies of categorization. *Cogn Affect Behav Neurosci* 2001; **1**:360–370.
- Delgado MR, Lock HM, Stenger VA and Fiez JA. Dorsal striatum responses to reward and punishment: effects of valence and magnitude of manipulations. *Cogn Affect Behav Neurosci* 2003; **3**:27–38.
- Haruno M, Kuroda T, Doya K, Toyama K, Kimura M *et al.* A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. *J Neurosci* 2004; **24**:1660–1665.
- Verny SP, Brown GC, Frank L and Paulus MP. Error-rate-related caudate and parietal cortex activation during decision making. *Neuroreport* 2003; **14**:923–928.

Acknowledgements: This research was supported in part by NINDS Grant (R01-41372) to J.V.F., NIMH Grant (R01-59196) to W.T.M., NIMH (R21-DA13186) and VA Merit Award Grants to M.P.P., and a James McDonnell Foundation Grant.