Behavioral/Systems/Cognitive

Dissociable Prototype Learning Systems: Evidence from Brain Imaging and Behavior

Dagmar Zeithamova,[1,2] **W. Todd Maddox,**[1,2] **and David M. Schnyer**[1,2]
[1]Institute for Neuroscience and [2]Department of Psychology, University of Texas at Austin, Austin, Texas 78712

The neural underpinnings of prototype learning are not well understood. A major source of confusion is that two versions of the prototype learning task have been used interchangeably in the literature; one where participants learn to categorize exemplars derived from two prototypes (A/B task), and one where participants learn to categorize exemplars derived from one prototype and noncategorical exemplars (A/non-A). We report results from an fMRI study of A/B and A/non-A prototype learning that allows for a direct contrast of the two learning methods. Accuracy in the two tasks did not correlate within subject despite equivalent average difficulty. The fMRI results revealed neural activation in a network of regions consistent with episodic memory retrieval for the A/B task while greater activation of a nondeclarative learning network was observed for the A/non-A task. The results demonstrate that learning in these two tasks is mediated by different neural systems and that recruitment of each system is dictated by the context of learning rather than the actual category structure.

Key words: **category learning; declarative memory; functional MRI; medial temporal lobe; perceptual learning; striatum**

Introduction

Category learning is an essential cognitive function. Evidence suggests that different forms of category learning are supported by different memory systems (Poldrack and Foerde, 2008) with each memory system being associated with different neural circuits (Schacter, 1987; Squire, 1992; Poldrack and Packard, 2003). For instance, rule-based learning relies on prefrontal cortex-mediated working memory while information integration relies on striatum-mediated procedural learning (Ashby and O'Brien, 2005; Nomura et al., 2007).

An important form of category learning is prototype learning—prototypes provide the abstract representation for many natural categories (Rosch, 1973; Rosch and Mervis, 1975) and form the basis of much categorization in young children (Strauss, 1979; Ross, 1980). However, the neural underpinnings of prototype learning remain unclear and contradictory findings exist, with an ongoing debate over whether prototype learning relies on declarative or nondeclarative memory systems (cf. Knowlton and Squire, 1993; Palmeri and Flannery, 1999). Ashby and colleagues (Ashby and Maddox, 2005; Ashby and O'Brien, 2005) suggest that the lack of clarity may be due to the use of two different tasks to study prototype learning: an A/B task and an A/non-A task. In the A/B task, participants learn to categorize exemplars derived from two prototypes. In the A/non-A task, participants learn to categorize exemplars derived from one prototype against noncategorical exemplars.

When task type is taken into account, the neural basis of prototype learning may be clearer. A/non-A prototype learning is intact in patients with Parkinson's disease (Reber and Squire, 1999), schizophrenia (Kéri et al., 2001b), and amnesia and Alzheimer's disease (Knowlton and Squire, 1993; Bozoki et al., 2006). Neuroimaging studies with the A/non-A task report learning-related activity reductions in occipital cortex for category A exemplars compared with noncategorical exemplar (Aizenstein et al., 2000; Reber et al., 1998a,b), although the activation pattern also depends on the intentionality of learning (Reber et al., 2003). The results suggest that the perceptual representation memory system (Schacter, 1990) might mediate A/non-A learning.

In contrast, the A/B task is impaired in Alzheimer's disease and amnesia (Zaki et al., 2003). Neuroimaging studies with the A/B task primarily report learning-related changes in prefrontal and parietal cortices (Seger et al., 2000), when comparing task activation to that of fixation baseline. Vogels et al. (2002) used a hybrid A/B/neither task and found prefrontal and parietal activation, but also orbitofrontal and neostriatal activation with no task-related changes in occipital cortex. These findings suggest that explicit reasoning and/or declarative memory processes might mediate A/B prototype learning.

Firm conclusions regarding the neural basis of A/non-A and A/B prototype learning are complicated by the methodological differences between the two tasks and by the different fMRI contrasts typically used. To date, no neuroimaging study has examined A/non-A and A/B prototype learning using the same stimuli, participants, and fMRI contrasts. The overriding goal of this study is to address this significant shortcoming and examine the

Received June 24, 2008; revised Oct. 29, 2008; accepted Oct. 30, 2008.

This work was supported by Army Grant W911NF-07-2-0023, through The Center for Strategic and Innovative Technologies at The University of Texas at Austin. We thank Andrea Bozoki for providing us with the basis for the stimulus sets and Sean Maddox, Borami Lee, and Cristina Benavides for developing the rest of the stimulus sets. We also thank Tori Williams for valuable help with scanning and data analysis.

Correspondence should be addressed to Dagmar Zeithamova, 1 University Station A8000, University of Texas, Austin, TX 78712. E-mail: zeithamova@mail.utexas.edu.

DOI:10.1523/JNEUROSCI.2915-08.2008

 0270-6474/08/2813194-08$15.00/0

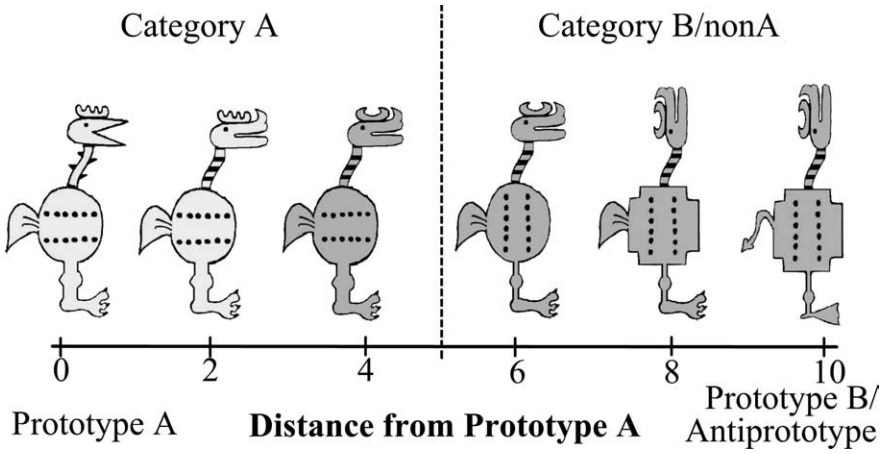

Figure 1. Example stimuli from one stimulus set. The leftmost stimulus represents the prototype of category A, stimuli to the right of the prototype represent examples of stimuli with increasing distances from the A prototype. The rightmost stimulus is the category B prototype. Stimuli with distance 0–4 from prototype A were considered category A members, stimuli with distance 6–10 were considered category B (non-A) members.

neural underpinnings of A/B and A/non-A prototype learning using a well controlled paradigm.

Materials and Methods

Participants. Twenty-seven young adult volunteers (age 18–30; 13 females) participated in the study. Data from 3 participants (1 female) were excluded due to excessive head motion, leaving 24 participants for analysis. Each participant provided signed informed consent to participate in the study and all procedures were approved by the IRB of The University of Texas, Austin. Volunteers received $50 compensation for a 2 h session.

Stimuli. The stimuli were cartoon animals that varied along 10 binary dimensions, such as body shape (round or square), head position (facing forward or upward), tail shape (feathery or pointy), etc. (Fig. 1), adapted from a prototype learning study of Bozoki et al. (2006). For each run, one stimulus served as the category A prototype with all 10 of its feature values being referred to as prototypical features. All other stimuli can be defined relative to the prototype and can differ on 1–10 of the prototypical feature values. The stimulus with all 10 nonprototypical features is the B prototype (in the A/B task) and the anti-prototype (in the A/non-A task). The number of nonprototypical features in each stimulus determines its distance from the prototype (see Fig. 1). Category A stimuli were defined as those with a distance of 0–4 from the A prototype and category B (or non-A) stimuli were defined as those with a distance of 6–10 from the A prototype. Stimuli equidistant from the two prototypes were excluded from the study.

A second set of cartoon animal stimuli with different dimensions were also generated, and each prototype learning task was tested with both sets of stimuli. Note that in this study, unlike in a typical A/non-A experiment, all non-A stimuli were internally consistent and constructed from a fixed prototype. Thus, the only difference between the A/non-A and A/B tasks was in the stimuli presented during training (only A stimuli in the A/non-A task, and A and B stimuli in the A/B task), and the category labels used during the testing phase. Critically, the same stimuli were used in the test phase for both the A/non-A and A/B tasks. Thus, any differences observed in the A/non-A and A/B brain activations cannot be attributed to differences between the structures of non-A versus B category, nor to any stimulus-specific differences.

Experimental design. A within subject design was used. Each run consisted of a training phase and a test phase, with functional MRI scans acquired during the testing phase of each run. Each participant completed two A/B runs (one run with stimulus set 1 and one run with stimulus set 2), a 10 min structural scan, and two A/non-A runs (again one run with each stimulus set). The order of stimulus sets and the order of the tasks were counterbalanced between participants. Importantly, although the training phases differed across tasks, the test phases were identical.

Training design (not scanned). Participants were in the scanner, but no fMRI was recorded during the training phase of each run. During training for the A/B task, participants were asked to categorize 10 A and 10 B items (presented one by one in a random order) with corrective feedback. On each trial, 2 s after stimulus onset, the participant was prompted to give an A or B response. After each response, the participant was informed whether they were correct or wrong. Within each category, 2 training stimuli differed from the category prototype on 1 feature, 3 differed on 2 features, 3 differed on 3 features, and 2 differed on 4 features. Across all 10 stimuli within each category, the category typical features were presented 7 or 8 times and the opposite category typical features were presented 2 or 3 times. Neither prototype was presented. The training stimuli were presented in a random order, different for each of the four runs, but identical across participants.

Before A/non-A training, participants were informed that they will need to learn to discriminate members of category A from nonmembers (non-A). During A/non-A training, participants were shown stimuli from category A only. Twenty training stimuli from category A were passively viewed one by one for a minimum of 2 s, after which a prompt asked a participant to press any button to proceed to a next example of a category member. There were 5 training stimuli that differed from the A prototype on one feature, 5 differed on two features, 5 differed on three features, and 5 differed on four features. Across all 20 stimuli, the prototypical value on each dimension was presented 15 times and the nonprototypical value on each dimension was presented 5 times.

Test design (fMRI recorded). The testing phase was identical for both tasks, with only the label of the second category (B versus non-A) differing between the tasks. Participants were presented with 42 stimuli, one at a time that included both prototypes and five stimuli selected from each distance from the prototype (except distance 5—ambiguous stimuli). None of the stimuli were previously used in the training phase. An event-related design was used to examine the neural activity to specific trials during the testing phase. Four possible orders of A and B stimuli and their onsets including 30% of null time (to interject temporal jitter) were predetermined using the "optseq2" program. Each stimulus onset time and order was used in one experimental run. On each trial, a stimulus was presented for a maximum of 3.5 s, during which time the participant needed to indicate the category membership of the stimulus. No feedback was provided. A fixation cross was presented between each stimulus onset lasting 0.5, 2.5, or 4.5 s.

MRI acquisition, processing, and analysis. Functional and structural images were acquired using a 3T GE Signa MRI scanner with an 8-channel phased array head coil. Functional images were acquired during the testing phase of each run, using a multiecho GRAPPA parallel imaging EPI sequence that reduces typical EPI distortions and susceptibility artifacts. Images were collected using whole-head coverage with slice orientation to reduce artifact (~20° off the AC-PC plane, TR = 2 s, 3 shot, TE = 30 ms, 35 axial slices oriented for best whole-head coverage, acquisition voxel size = 3.125 × 3.125 × 3 mm with a 0.3 mm interslice gap). The first four EPI volumes were discarded to allow scans to reach equilibrium. Stimuli were viewed through a back projection screen and a mirror mounted on the top of the head coil. Responses were collected with an MR compatible button box that was placed under the right hand.

In addition to the EPI images collected during task performance, one or two high-resolution T1 SPGR scans that have been empirically optimized for high contrast between gray matter (GM) and white matter (WM) and between GM and CSF were acquired. These images were acquired in the sagittal plane using a 1.3 mm slice thickness with 1 mm^2 in-plane resolution.

Table 1. Regions commonly activated in both the A/B and A/non-A task

Brain region	Volume	Size	Max *Z*	*x*	*y*	*z*
Whole brain cluster corrected ($p < 0.05$)						
L lateral occipital (BA 19)	37,752	4719	6.93	−36	−86	−4
R lateral occipital (BA 19)	27,096	3387	6.75	42	−66	−12
Calcarine (BA 17)	6456	807	3.59	10	−72	6
L postcentral (BA 3/40)	29,640	3705	5.71	−44	−26	48
R inferior parietal (BA 7/40)	28,104	3513	5.69	36	−54	46
R fusiform/inferior temporal (BA 37)	13,896	1737	6.83	40	−54	−20
L fusiform (BA 37)	7168	896	6.73	−38	−64	−20
L inferior frontal (BA 44/48)	18,664	2333	6.1	−48	6	28
Medial frontal (BA 24/32)	15,416	1927	5.79	−4	8	46
R inferior frontal (BA 44)	13,920	1740	5.61	52	10	24
R middle frontal (BA 6)	3912	489	4.08	30	−4	46
Small volume corrected ($p < 0.05$)						
R hippocampus	984	123	5.15	20	−30	−4
L hippocampus	672	84	4.56	−20	−30	−8
R striatum	1040	130	4.35	16	16	−2
L striatum	488	61	3.23	−20	10	−4

L, Left; R, right; BA, Brodmann area; Max, maximum. Volume is given in mm^3, and size is given in voxels. Coordinates reflect standard MNI space.

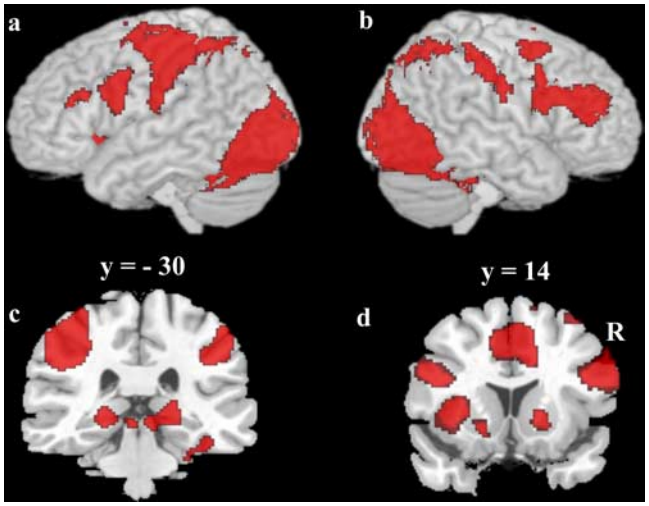

Figure 2. Commonly activated regions in both tasks versus baseline. ***a***, ***b***, Whole-brain 3D rendering with cortical activation overlay. ***a***, Left hemisphere. ***b***, Right hemisphere. ***c***, ***d***, Coronal slices with activations overlays. ***c***, Bilateral hippocampus. ***d***, Bilateral striatum and medial frontal cortex. Activation maps were overlaid upon a canonical brain in standard MNI space using MRIcro software (www.sph.sc.edu/comd/rorden/mricro.html).

Preprocessing and data analysis were conducted using FEAT (FMRI Expert Analysis Tool) version 5.63, part of FSL (www.fmrib.ox.ac.uk/fsl) software. Preprocessing included motion correction using MCFLIRT (Jenkinson et al., 2002), non-brain removal using BET (Smith, 2002), high-pass temporal filtering with a 60 s cutoff, and spatial smoothing with a Gaussian kernel of 5 mm FWHM, and normalization to a 2 mm resolution MNI template brain. Data from each run of each participant were analyzed separately at a first level of analysis. Category A and category B/non-A trials were modeled separately as two predictors. Each category stimulus time onsets were convolved with a canonical hemodynamic response function and together with their temporal derivatives were entered as predictors into a general linear model to estimate β-weights. Data from all four runs from each participant were combined at a second level using a fixed effects analysis. Group level analysis combined data from each participant in a random effects analysis using OLS. For all analyses, individual voxels were considered active when reaching $Z > 2.3$ and survived a whole-brain cluster-size threshold set at $p < 0.05$ (Worsley, 2001). In the FSL implementation of random field theory, the minimal cluster size is determined by both the set cluster size p value and the smoothness of the data estimated directly from the contrast image. The determined minimal cluster size to satisfy cluster size probability threshold of $p < 0.05$ thus varied slightly from contrast to contrast around 240 voxels (1920 mm^3).

Additionally to the whole-brain analysis, we defined two regions of interest (ROI): medial temporal lobe (MTL) and striatum. We were especially interested in area MTL because its involvement in prototype learning has been controversial and in striatum because it has been implicated in other kinds of category learning and is thought to operate complementary to area MTL (Poldrack et al., 2001; Poldrack and Packard, 2003). The MTL ROI was defined by combining the FSL Harvard–Oxford atlas hippocampus and parahippocampal regions for the left and right hemispheres. The striatum ROI consisted of the combined putamen and caudate from the FSL Harvard–Oxford atlas, again for both the left and right hemispheres. Activation in each ROI was assessed using a small volume correction at $p < 0.05$ based on Monte Carlo simulation, accounting for both smoothness of the data and the shape and size of each ROI. During each simulation, uniform random numbers were assigned as activation p values to individual voxels in a mask of the same shape and size as the ROI of interest, representing a possible pattern of "activation" that could be recorded under the null hypothesis of no real activation in the region. The simulated voxel activations were then smoothed with the same kernel as the actual data and the maximal cluster size that occurred under the null hypothesis by chance was recorded for each simulation. Cluster size that occurred with probability <0.05 across 5000 simulations was then considered significant at the cluster-size threshold of $p < 0.05$. The simulations were performed using "AlphaSim" tool in AFNI and determined a minimal required cluster size of 33 voxels (264 $mm^{3)}$ in the normalized space for the MTL ROI and 30 voxels (240 mm^3) for the striatum ROI.

Results

Behavioral performance

For the main behavioral and fMRI analyses, test phase data were pooled across the two runs (the two stimulus sets) of each task. There were no differences between accuracies achieved on the two stimulus sets (A/B: 0.68 vs 0.71, $t_{(23)} = 1.186$, $p = 0.248$; A/non-A: 0.67 vs 0.68, $t_{(23)} = 0.441$, $p = 0.664$) and there was no difference between overall accuracy in the A/B task (mean = 0.694, SE = 0.018) and the A/non-A task (mean = 0.673, SE = 0.020; $t_{(23)} = -0.644$, $p = 0.526$).

Interestingly, A/B and A/non-A accuracy rates were moderately negatively correlated ($r = -0.362$, $p = 0.082$), suggesting that distinct cognitive processes may underlie participants performance in the two tasks. Unlike accuracy, mean reaction times

Table 2. Regions from whole-brain and region-of-interest (small volume corrected) analysis that activated differentially during the A/B task and the A/non-A task

Brain region	Volume	Size	Max *Z*	*x*	*y*	*z*
A/B > A/non-A (whole brain cluster corrected)						
R inferior parietal (BA 40)	5968	746	3.79	58	−38	46
L orbitofrontal (BA 47/11)	3048	381	3.9	−36	54	−16
A/B > A/non-A (small volume corrected)						
L parahippocampus (BA 36)	424	53	3.13	−20	−6	−30
A/non-A > A/B (whole brain cluster corrected)						
L Inf lateral occipital (BA 18)	7912	989	4.88	−20	−94	0
R Inf lateral occipital (BA 19)	6208	776	4.37	38	−82	−2
L Sup parietal (BA 7)	3824	478	4.24	−20	−70	36
R Sup parietal (BA 7)	3608	451	3.74	22	−64	48
A/non-A > A/B (small volume corrected)						
R putamen	328	41	3.47	20	10	−8
R caudate head	264	33	3.24	10	10	−2
L caudate body	256	32	4.27	−10	6	18

L, Left; R, right; Inf, inferior; Sup, superior; BA, Brodmann area; Max, maximum. Volume is given in mm^3, and size is given in voxels. Coordinates reflect standard MNI space.

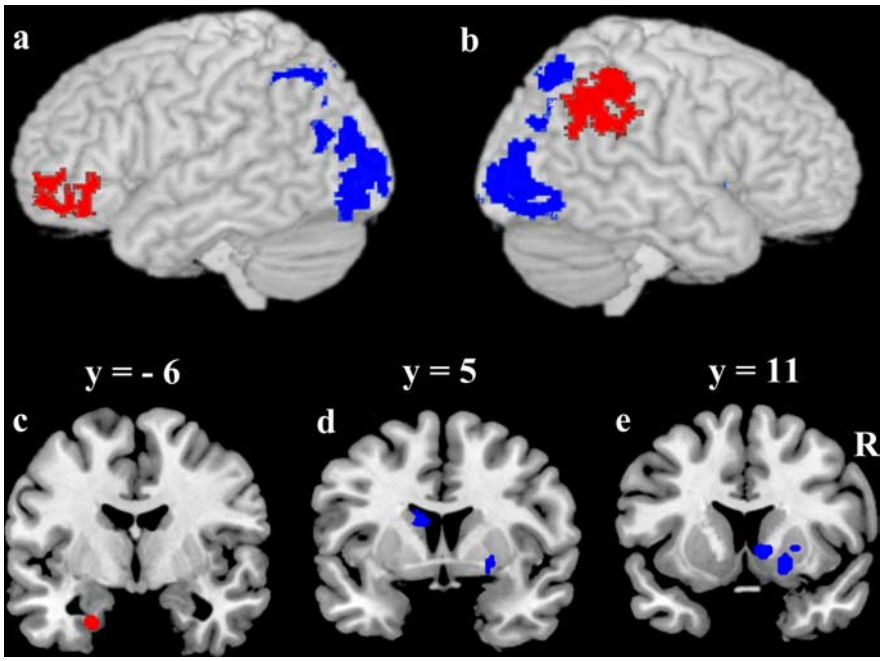

Figure 3. Regions from direct contrast of A/B task versus A/non-A task. In red, A/B > A/non-A; in blue A/non-A > A/B. ***a***, ***b***, Whole-brain cluster corrected contrasts overlaid on a 3D rendering of a canonical brain. ***a***, Left hemisphere. ***b***, Right hemisphere. ***c–e***, Coronal sections illustrating small volume corrected contrast maps in regions of interest. ***c***, Left parahippocampus. ***d***, Left caudate body. ***e***, Right putamen and right caudate head. Statistical maps were overlaid upon a canonical brain in standard MNI space using MRIcro software (www.sph.sc.edu/comd/rorden/mricro.html).

differed between the two tasks by ~0.2 s (A/B: mean = 1.343 s, SE = 0.095; A/non-A: mean = 1.545 s, SE = 0.099; $t_{(23)} = 3.566$, $p = 0.002$) and were positively correlated within subject ($r = 0.831$, $p < 0.001$).

Common neural regions

First, we identified regions that showed common activation in both the A/B task and the A/non-A task compared with the fixation baseline, using overlap masking of the two thresholded *z*-maps. *Z* values for the common activation *z*-map (as reported in Table 1) are the minimum of the two tasks' *z*-maps. A network of regions in which both tasks showed significantly greater activation compared with the fixation baseline (see Fig. 2, Table 1) included occipital and fusiform areas, inferior frontal cortex, and precentral gyrus (Fig. 2*a*,*b*), as well as bilateral posterior hippocampus (Fig. 2*c*) and bilateral striatum (Fig. 2*d*).

Task differences

The primary focus of this research was to directly compare activity during the A/B task and the A/non-A task within subject using the same stimuli. On direct contrast between test trials, a number of regions exhibited greater activity in one task compared with the other. The list of identified regions is provided in Table 2, with corresponding statistical maps provided in Figure 3.

The direct contrast revealed that the A/B task involves to a greater degree frontal and parietal cortices and parahippocampus (Fig. 3, red overlay), areas that have been previously implicated in explicit episodic memory. In contrast, regions that demonstrated greater activity in the A/non-A task than the A/B task included primarily posterior cortices and striatum (Fig. 3, blue overlay), areas previously implicated in nondeclarative perceptual (e.g., Slotnick and Schacter, 2006) and procedural learning (e.g., Packard and Knowlton, 2002; Poldrack and Packard, 2003). Because reaction times were not perfectly equated in the two tasks, it is possible that some of the regions identified in the A/non-A > A/B contrast may reflect longer processing times in the A/non-A task than in the A/B task. However, adding the reaction time differences as a covariate at the group level analysis did not eliminate the activation differences.

Behavioral relevance of differential neural regions

For each of 10 clusters identified in the direct contrast, average time courses for each voxel and for each condition were computed using a selective averaging technique (http://www.poldracklab.org/software). A mean activation (average of 4–8 s after stimulus onset) was computed for each cluster and each participant, during the A/B task and during the A/non-A task, and were correlated with the participant's performance in each task. For each task separately, we excluded participants

Table 3. Regions that exhibited greater activation during correct than incorrect trials

Brain region	Volume	Size	Max *Z*	*x*	*y*	*z*
A/B task (whole brain cluster corrected)						
R middle temporal (BA 21/22)	2984	373	3.64	62	−14	−16
L middle temporal (BA 21/22)	2104	263	3.57	−58	−36	4
Posterior cingulate/ precuneus (BA 23)	12,608	1576	4.58	−6	−52	10
Orbitofrontal (BA 10/11)	11,440	1430	4.1	2	62	−14
A/B task (small volume corrected)						
R medial temporal (BA 20)	1904	238	4.18	30	−22	−16
L medial temporal (BA 20)	1608	201	3.69	−32	−22	−14
A/non-A task (small volume corrected)						
L putamen	1416	177	4.25	−32	−10	−2
R anterior hippocampus	304	38	3.8	28	−10	−22

Regions are identified for each task separately. Volume is given in mm^3, and size is given in voxels. Coordinates reflect standard MNI space.

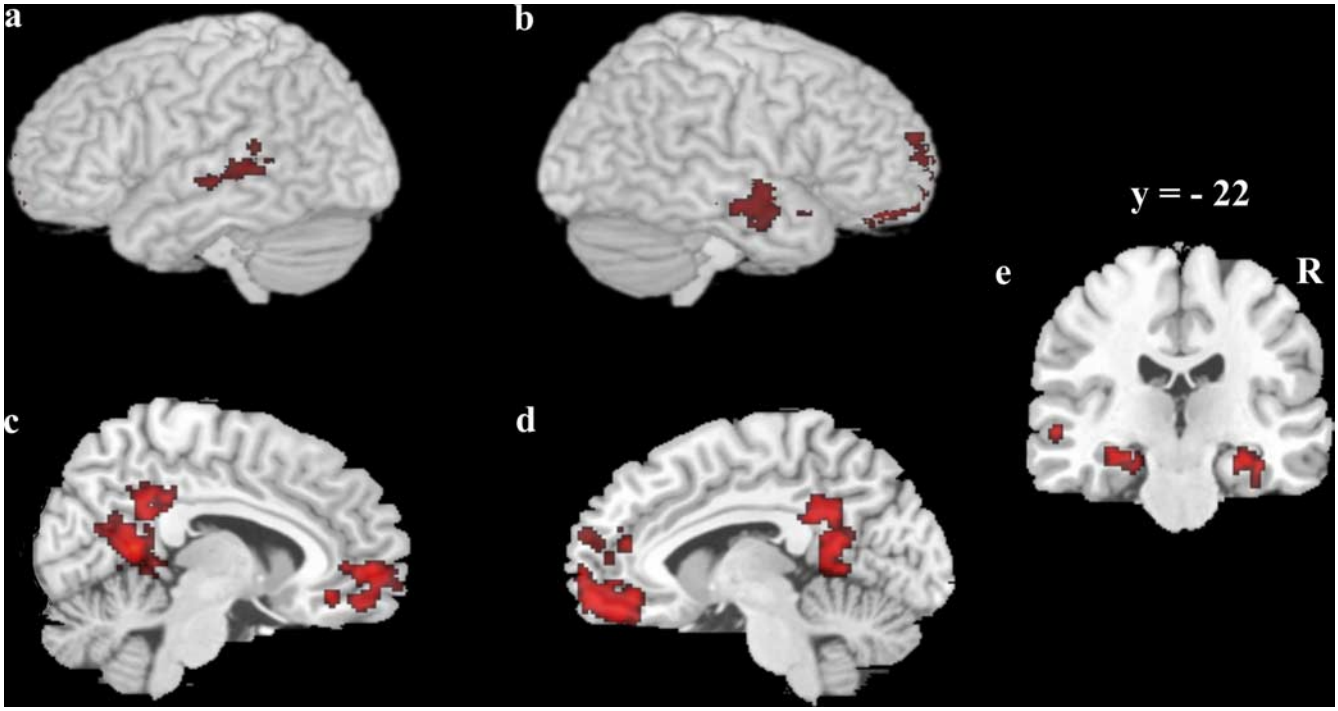

Figure 4. Regions associated with successful categorization during the A/B task. ***a***, ***b***, Lateral view of the left and right hemisphere 3D rendering with activation overlay. ***c***, ***d***, Medial view of the left and right hemisphere. ***e***, Coronal section showing medial temporal lobe activation.

who did perform at chance in both of the two runs of the given task. There were 3 participants excluded from A/B task correlational analysis and four participants excluded from A/non-A correlational analysis. Neural activity in 2 out of 3 clusters identified in the A/B > A/non-A contrast was predictive of behavioral performance in the A/B task—the left inferior orbital frontal cortex ($r = 0.432$, $p = 0.050$) and the left parahippocampus ($r = 0.443$, $p = 0.044$)—indicating that participants who recruited these regions to a larger degree during the A/B task performed better in the A/B task. Neither of the regions predictive of performance in the A/B task was predictive of performance during the A/non-A task. Out of the 7 regions identified in the A/non-A > A/B contrast, none significantly predicted accuracy in the A/non-A task (all $r < 0.30$, $p > 0.19$).

Neural regions predictive of trial accuracy

To identify brain areas predictive of successful categorization on an individual participant's level, we compared activity evoked during correct categorization trials with that evoked during incorrect categorization trials, separately for each task. Identified regions that exhibited greater activation during correct than incorrect trials are listed in Table 3 and are presented in Figures 4 and 5. No region exhibited greater activation for incorrect than correct trials in either task. The activation differences between correct and incorrect trials cannot be accounted for by reaction time differences. For both tasks, there were reaction time differences between correct and incorrect trials, but with longer incorrect than correct reaction times (A/B task: correct mean = 1.39 s, SE = 0.09 s, incorrect mean = 1.51 s, SE = 0.10 s, $t_{(23)} = 3.03$, $p = 0.006$; A/non-A task: correct mean = 1.59 s, SE = 0.09 s, incorrect mean = 1.70 s, SE = 0.10, $t_{(23)} = 3.83$, $p = 0.001$).

Regions that were predictive of correct categorization during the A/B task trials included bilateral middle temporal cortices, posterior cingulate cortex, and orbitofrontal cortex, as well as bilateral medial temporal lobe spanning parts of both parahippocampus and hippocampus. Only two regions were predictive of correct categorization during the A/non-A task,

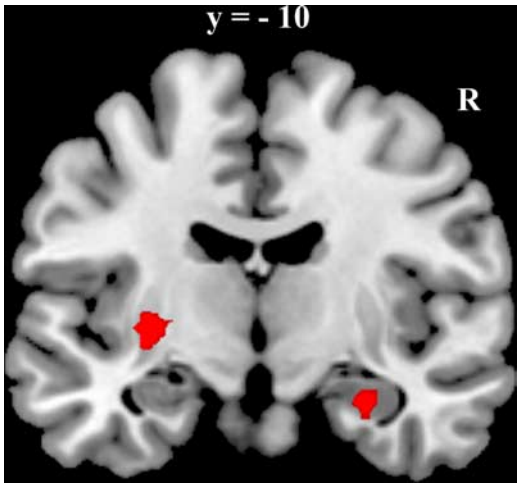

Figure 5. Regions associated with successful categorization during the A/non-A task. Coronal section featuring left putamen and right hippocampal activation.

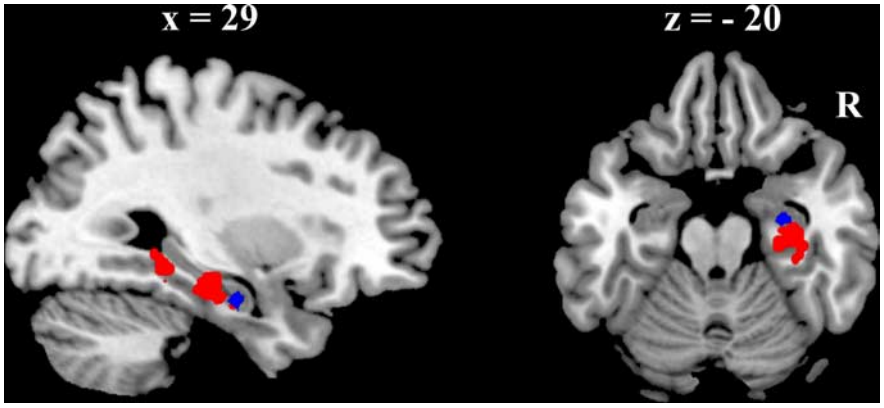

Figure 6. Comparison of MTL regions implicated in the A/B task and the A/non-A task. Sagittal and horizontal section illustrating relative location of the regions of MTL that showed greater activation during correct than incorrect trials during the A/B task (red) and the A/non-A task (blue).

left putamen and right anterior hippocampus. The relative location of the right hippocampal region identified in the A/non-A task and the right MTL region identified in the A/B task is presented in Figure 6. The A/non-A region was located anterior to the A/B region and there was minimal overlap between the regions (3 voxels).

Discussion

We conducted an fMRI prototype learning experiment that used equivalent stimuli, learning mode and a within subject design to examine the neural basis of A/non-A and A/B prototype learning. The results were consistent with the proposition that A/non-A and A/B prototype learning are based on dissociable processes, each with a corresponding neural system. First, there was a negative correlation between A/non-A and A/B task performance even when behavioral data on average showed comparable learning in both tasks. Second, brain regions were identified that were preferentially active during one task versus another. Most notably, the A/B task recruited to a larger degree parahippocampus and inferior parietal and orbitofrontal cortex. Moreover, individual differences in the activation of parahippocampus and orbitofrontal cortex were predictive of participant's accuracy in the A/B task, but not the A/non-A task. Hippocampal and parahippocampal activation was also predictive of correct responses on an individual trial basis. These findings are consistent with the notion that A/B categorization tasks engage mechanisms that are similar to other processes that rely on the MTL, such as declarative or associative memory.

Direct contrast of the tasks also revealed regions that were preferentially recruited during the A/non-A categorization. This included regions of lateral occipital cortex and striatum, with regions of the striatum being predictive of correct responses on individual trials during the A/non-A task. The striatum and posterior cortices have been implicated in multiple studies of non-declarative category learning (Poldrack et al., 2001; Seger and Cincotta, 2002, 2005; Shohamy et al., 2004; Nomura et al., 2007) and these findings are consistent with the idea that the A/non-A task is based to a larger degree on implicit, perceptual, and/or procedural learning.

The preferential involvement of the posterior corticostriatal loops was observed here even though the intentional learning mode and within subject design could bias participants toward applying conscious, explicit strategies, and even though there was no external feedback provided during test or training. Although individual differences in activation of neither striatum nor posterior cortical regions were predictive of participants' accuracy, this could be expected based on previous findings. Occipital and temporal activity in perceptual learning experiments is typically not correlated with behavior (Schacter et al., 2007). In addition, the lack of correlation with striatal activity may be due to the relatively small number of trials used in the current experiment such that it may reflect the early stages of learning in this system. For example, Seger and Cincotta (2005) found the striatum to be significantly involved throughout learning, but becoming predictive of accuracy only later in the learning, after ~300 training trials, while each of our runs involved only 20 training trials. Compared with the striatum, the hippocampal learning system comes online relatively quickly, being dominant early in learning (Poldrack et al., 2001), and providing the basis for the correlation with individual differences in accuracy observed here for the A/B task. The neuropsychological literature also supports the notion that the A/B task depends on MTL and declarative memory while the A/non-A task does not. For example, Knowlton and Squire (1993) found intact A/non-A prototype learning in patients with MTL lesion-based amnesia; Bozoki et al. (2006) and Kéri et al. (2001a) found intact A/non-A learning in patients with Alzheimer's disease. In contrast, Zaki et al. (2003) found impaired learning in amnesic patients in the A/B prototype learning task, but not A/non-A task.

Prototype learning is ubiquitous in everyday cognition. We hypothesized that prototype learning is not mediated by a single neural system, but rather that the system relevant to prototype learning depends critically upon the circumstances of learning—whether the task involves learning to discriminate a single category from other stimuli (A/non-A task), or classification of stimuli into two separate categories (A/B task). In previous studies, the information about which prototype task was used was typically buried deep in the method section and conclusions derived

from one version of the prototype task were readily generalized to the other version. However, there has been no empirical evidence that the two tasks recruit identical cognitive and neural processes and the differential demands of the two tasks suggest that they may not. In the A/non-A task, participants are likely to form a representation of a single prototype and then compare each test item to this single prototype. If the new stimulus is sufficiently similar to the prototype representation, it will be endorsed to the category; otherwise it will be categorized as a nonmember. Novelty or familiarity signals from early processing areas may be used as a basis for successful categorization. In contrast, in the A/B task, participants are likely to form representations of two distinct categories centered on two prototypes. Each new stimulus is then weighed against each prototype and endorsed to the category of the prototype that is closer to the current stimulus. Additionally, participants need to recollect details of learning and extract the appropriate verbal category label. In this case, familiarity or novelty signals are not sufficient for successful performance and additional processes need to be recruited to support learning. Thus, we expected that processes and neural structures supporting prototype learning should depend on the context of learning—whether a category is learned in isolation (as in the A/non-A task), or whether two categories are contrasted with each other (as in the A/B task).

Alternatively, some or all discrepancies in the prototype learning literature could be a by-product of the different learning modes typically adopted in the A/B task and the A/non-A task. The A/B task always involves intentional (conscious) learning where the participants are instructed to learn the characteristics of the categories based on corrective feedback or the provided category label (Little and Thulborn, 2006). The A/non-A task often involves incidental learning where participants passively view category exemplars first while being naive to the purpose of the experiment, with a later "surprise" test on discrimination of categorical from non-categorical exemplars (Reber et al., 1998a). Two fMRI studies compared neural activation in both incidental and intentional version of an A/non-A task (Aizenstein et al., 2000; Reber et al., 2003), showing differential pattern of activation for the two learning modes. However, the current experiment may offer some clarification of this issue since the results presented here demonstrate that even when learning mode is equated, dissociable neural systems emerge that support the A/B task and the A/non-A task. Importantly, only the context of learning (A/non-A vs A/B) differed across the tasks in the current study and thus any differences in the neural signature must be due to context and not methodology, learning mode, or category structure differences.

Both prototype learning systems likely play an important and complementary role in concept acquisition, as everyday prototype learning experience contains elements of both tasks. Each system has its own strengths and limitations. The A/non-A prototype learning system has been shown to work automatically, supporting even incidental learning without supervision (Posner and Keele, 1968). Perceptual coherence of the category exemplars is a major factor in concept learnability; concepts such as carrot or apple can be well learned by the A/non-A system. The A/B prototype learning system depends on supervision (Casale and Ashby, 2008), but allows one to form categories that are less perceptually coherent and make inferences that are not based solely on perceptual similarity. The concept of fruits and vegetables is better suited for the A/B system. While typically operating in parallel, these prototype learning systems are dissociable when demands of the task are tuned to suit one system versus the other (as demonstrated here) or when damage to one system hinders learning of specific tasks versions (as supported by the neuropsychological literature). Importantly, rather than the category structure itself, the framing and context of the task, such as whether a category is learned in isolation or next to another category, play a crucial role in recruiting the complementary learning systems.

References

Aizenstein HJ, MacDonald AW, Stenger VA, Nebes RD, Larson JK, Ursu S, Carter CS (2000) Complementary category learning systems identified using event-related functional MRI. J Cogn Neurosci 12:977–987.

Ashby FG, Maddox WT (2005) Human category learning. Annu Rev Psychol 56:149–178.

Ashby FG, O'Brien JB (2005) Category learning and multiple memory systems. Trends Cogn Sci 9:83–89.

Bozoki A, Grossman M, Smith EE (2006) Can patients with Alzheimer's disease learn a category implicitly? Neuropsychologia 44:816–827.

Casale MB, Ashby FG (2008) A role for the perceptual representation memory system in category learning. Percept Psychophys 70:983–999.

Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.

Kéri S, Kálmán J, Kelemen O, Benedek G, Janka Z (2001a) Are Alzheimer's disease patients able to learn visual prototype? Neuropsychologia 39:1218–1223.

Kéri S, Kelemen O, Benedek G, Janka Z (2001b) Intact prototype learning in schizophrenia. Schizophr Res 52:261–264.

Knowlton BJ, Squire LR (1993) The learning of categories: parallel brain systems for item memory and category knowledge. Science 262:1747–1749.

Little DM, Thulborn KR (2006) Prototype-distortion category learning: a two-phase learning process across a distributed network. Brain Cogn 60:233–243.

Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ (2007) Neural correlates of rule-based and information-integration visual category learning. Cereb Cortex 17: 37–43.

Packard MG, Knowlton BJ (2002) Learning and memory functions of the basal ganglia. Annu Rev Neurosci 25:563–593.

Palmeri TJ, Flannery MA (1999) Learning about categories in the absence of training: profound amnesia and the relationship between perceptual categorization and recognition memory. Psychol Sci 10:526–530.

Poldrack RA, Clark J, Paré-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, Foerde K (2008) Category learning and the memory systems debate. Neurosci Biobehav Rev 32:197–205.

Poldrack RA, Packard MG (2003) Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia 41:245–251.

Posner MI, Keele SW (1968) On the genesis of abstract ideas. J Exp Psychol 77:353–363.

Reber PJ, Squire LR (1999) Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. Behav Neurosci 113:235–242.

Reber PJ, Stark CEL, Squire LR (1998a) Contrasting cortical activity associated with category memory and recognition memory. Learn Mem 5:420–428.

Reber PJ, Stark CEL, Squire LR (1998b) Cortical areas supporting category learning identified using functional MRI. Proc Natl Acad Sci U S A 95:747–750.

Reber PJ, Gitelman DR, Parrish TB, Mesulam MM (2003) Dissociating explicit and implicit category knowledge with fMRI. J Cogn Neurosci 15:574–583.

Rosch E (1973) Natural categories. Cogn Psychol 4:328–350.

Rosch E, Mervis CB (1975) Family resemblances: studies in the internal structure of categories. Cogn Psychol 7:573–605.

Ross GS (1980) Categorization in 1- to 2-yr-olds. Dev Psychol 16:391–396.

Schacter DL (1987) Implicit memory: history and current status. J Exp Psychol Learn Mem Cogn 13:501–518.

Schacter DL (1990) Perceptual representation systems and implicit memory: toward a resolution of the multiple memory systems debate. Ann N Y Acad Sci 608:543–567; discussion 567–571.

Schacter DL, Wig GS, Stevens WD (2007) Reductions in cortical activity during priming. Curr Opin Neurobiol 17:171–176.

Seger CA, Cincotta CM (2002) Striatal activity 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.

Seger CA, Poldrack RA, Prabhakaran V, Zhao M, Glover GH, Gabrieli JDE (2000) Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. Neuropsychologia 38: 1316–1324.

Shohamy D, Myers CE, Onlaor S, Gluck MA (2004) Role of the basal ganglia in category learning: how do patients with Parkinson's disease learn? Behav Neurosci 118:676–686.

Slotnick SD, Schacter DL (2006) The nature of memory related activity in early visual areas. Neuropsychologia 44:2874–2886.

Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17:143–155.

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys and humans. Psychol Rev 99:195–231.

Strauss MS (1979) Abstraction of prototypical information by adults and 10-month-old infants. J Exp Psychol [Hum Learn] 5:618–632.

Vogels R, Sary G, Dupont P, Orban GA (2002) Human brain regions involved in visual categorization. Neuroimage 16:401–414.

Worsley KJ (2001) Statistical analysis of activation images. In: Functional MRI: an introduction to methods (Jezzard P, Matthews PM, Smith S, eds), pp 251–270. Oxford: Oxford UP.

Zaki SR, Nosofsky RM, Jessup NM, Unversagt FW (2003) Categorization and recognition performance of a memory-impaired group: evidence for single-system models. J Int Neuropsychol Soc 9:394–406.