

# Re-evaluating Dissociations between Implicit and Explicit Category Learning: An Event-related fMRI Study

Todd M. Gureckis<sup>1</sup>, Thomas W. James<sup>2</sup>, and Robert M. Nosofsky<sup>2</sup>

## Abstract

■ Recent fMRI studies have found that distinct neural systems may mediate perceptual category learning under implicit and explicit learning conditions. In these previous studies, however, different stimulus-encoding processes may have been associated with implicit versus explicit learning. The present design was aimed at decoupling the influence of these factors on the recruitment of alternate neural systems. Consistent with previous reports, following incidental learning in a dot-pattern classification task, participants showed decreased neural activity in occipital visual cortex (extrastriate region V3, BA 19) in response to novel exemplars of a studied category compared to members of a foil

category, but did not show this decreased neural activity following explicit learning. Crucially, however, our results show that this pattern was primarily modulated by aspects of the stimulus-encoding instructions provided at the time of study. In particular, when participants in an implicit learning condition were encouraged to evaluate the overall shape and configuration of the stimuli during study, we failed to find the pattern of brain activity that has been taken to be a signature of implicit learning, suggesting that activity in this area does not uniquely reflect implicit memory for perceptual categories but instead may reflect aspects of processing or perceptual encoding strategies. ■

## INTRODUCTION

In recent years, a growing number of studies have argued that distinct neural pathways are engaged during category learning depending on the information structure of the category (Smith, Minda, & Washburn, 2004; Ashby, Alfonso-Reese, Turken, & Waldron, 1998), the way in which participants engage with the learning task (Maddox & Ashby, 2004; Reber, Gitelman, Parrish, & Mesulam, 2003; Reber, Stark, & Squire, 1998a, 1998b) and the neuropsychological condition of the learner (Kéri, 2003; Ashby & Ell, 2001; Poldrack et al., 2001; Smith, Patalano, & Jonides, 1998; Knowlton, Mangels, & Squire, 1996; Kolodny, 1994). For example, Squire and Knowlton (1995) and Knowlton and Squire (1993) reported that amnesic patients with selective damage to the medial temporal lobe (MTL) could categorize dot-pattern stimuli at levels comparable to controls but were significantly impaired at recognizing similar patterns in a memory task. The demonstration of preserved categorization performance in the absence of declarative memory has given support to the idea that there are multiple neural pathways for category acquisition and, in particular, that certain forms of learning are implicit (i.e., effectively operate outside conscious awareness; Poldrack & Foerde, 2008; Smith, 2008).

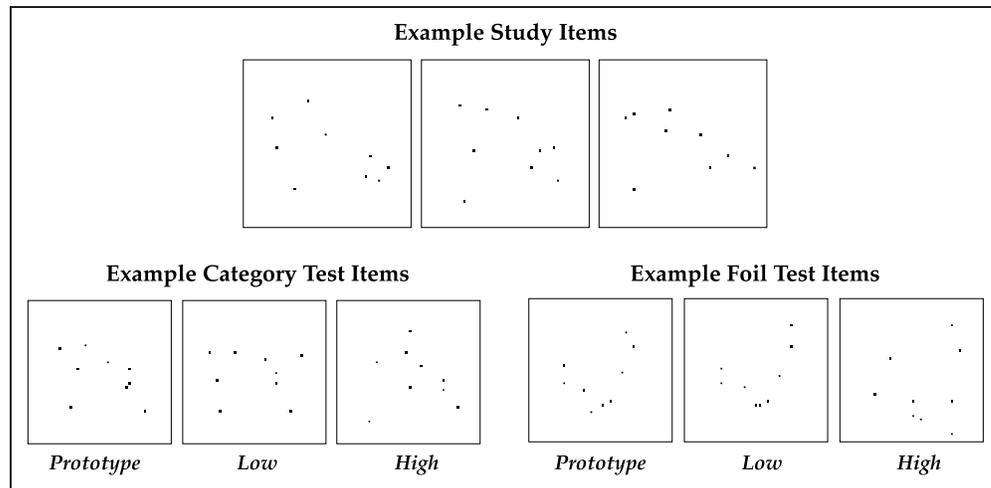
As intriguing as these behavioral dissociations are, other theorists have argued that single-system learning models provide a natural account of the results (e.g., Love &

Gureckis, 2007; Zaki, 2004; Zaki & Nosofsky, 2001; Palmeri & Flanery, 1999; Nosofsky & Zaki, 1998; Shanks & St. John, 1994). For example, Love and Gureckis (2007) and Nosofsky and Zaki (1998) showed that single-system models that made allowances for memory-related parameter differences between normal and amnesic individuals could capture in detailed fashion the classic data from Knowlton and Squire. These alternative accounts have given rise to a debate about whether there are indeed distinct explicit versus implicit category-learning systems. Note that “single-system” models do not argue against the idea that multiple neural systems may be involved in various aspects of cognitive processing (e.g., Love & Gureckis, 2007), but take issue with the more specific claim that completely separate implicit versus explicit memory systems underlie different forms of category learning.

As argued by E.E. Smith (2008), however, the most direct evidence bearing on the idea that distinct neural systems underlie implicit category learning arises from neuroimaging studies (e.g., Koenig et al., 2008; Nomura et al., 2007; Reber et al., 1998a, 1998b, 2003). Highlighted in Smith’s (2008) review was the influential study of Reber et al. (2003), who reported a neural dissociation between implicit (or “incidental”) and explicit (or “intentional”) category learning in healthy individuals using fMRI. In their study, which forms the basis of the present report, participants were assigned to one of two tasks that were identical except for a difference in prestudy phase instructions. In the explicit learning task, participants were told that they would observe a number of dot patterns (see Figure 1 for

<sup>1</sup>New York University, <sup>2</sup>Indiana University

**Figure 1.** Examples of the stimuli used in Reber et al. (2003) and the present study. The top row shows three examples of the study items. The bottom row shows examples of stimuli that were category members (in the sense that they were generated from the same prototypical pattern as the study items), as well as foil patterns (which came from a different prototype). The two classes of test items (category members and foils) both varied in similarity with respect to their underlying prototype (i.e., test patterns were either the actual prototype, a low distortion, or high distortion).



examples), each of which shared a category relationship to one another, and that they should attempt to learn the features that define the category. In the implicit learning task, participants viewed the exact same stimuli. However, in this condition, no mention was made of the categorical nature of the stimuli. Instead, participants were simply told that the study was designed to examine the effects of mental imagery on brain activation, and thus, were asked to mentally imagine pointing to the centrally located dot of each stimulus. After a short study phase, which was identical for both groups, all participants were told that they had just studied a category and, during the test phase, were asked to judge if novel dot-pattern stimuli were category members or nonmembers.

Although both groups achieved similar levels of accuracy in the task, functional imaging of the test phase revealed a dissociation between the neural systems recruited as a function of the two sets of study instructions. In particular, participants who studied under implicit learning conditions showed a significant *decrease* in activity for novel category members relative to novel foils in posterior occipital cortex (extrastriate region V3, BA 19), whereas participants in the explicit study condition showed a pattern of *increased* activity for category members relative to foil patterns in a number of regions associated with declarative memory processing, including right prefrontal cortex, hippocampus, left interior temporal cortex, precuneus, and posterior cingulate. Reber et al. (2003) hypothesized that occipital lobe deactivation for category members following implicit learning reflected a type of repetition priming for perceptually similar items (which they called the categorical fluency effect).

The Reber et al. (2003) study gives powerful support to the idea that separate pathways mediate implicit and explicit category learning. Nevertheless, in our view, questions remain regarding the ultimate basis for the results. In particular, differences between the instructions given to participants in the two experimental conditions were likely to

influence more dimensions of the tasks than simply the implicit–explicit status of the learner. Most critically, in the implicit learning condition, participants were given a specific encoding strategy (i.e., look for the center dot and imagine pointing to it). In contrast, participants in the explicit learning condition were not given a specific encoding strategy. As a result, it is likely that participants in the different conditions adopted different encoding or stimulus processing strategies. For example, participants may have focused attention on a single visual–spatial location in the dot-pointing condition, but spread attention more broadly in the nonspecific condition. The implication is that the observed pattern may have been due to differences in implicit–explicit learning (as suggested by Reber et al.), differences in encoding strategies, or a combination of these factors. Because single-system models of category learning have been strongly challenged on grounds of Reber et al.’s observed implicit–explicit dissociation, it is of crucial theoretical importance to test which of the aforementioned factors are indeed responsible for the effect.

In the present study, we factorially manipulated the nature of the encoding instructions and the implicit–explicit nature of the study task. If category-specific neural activity in occipital cortex is primarily the result of learners not knowing the nature of the category-learning task at the time of study, then our manipulations of encoding instructions should have little effect. In contrast, if manipulations of these study features alter the results in dramatic ways, it suggests that other factors (in addition to or besides implicit learning) contribute to the categorical fluency effect. In addition, our study extends the results from Reber et al. by using an event-related fMRI design. To foreshadow, our results in the conditions that replicate Reber et al. (2003) were qualitatively similar to the original report. However, relative to the implicit–explicit distinction, we find that the encoding strategy suggested to learners prior to study has a stronger effect on the pattern of brain activity observed at test.

## METHODS

### Participants

Data from 44 participants (22 women) were included in the analysis. The Indiana University Institutional Review Board approved the experimental protocol. All participants were paid \$25 for their participation in the study. Data from three participants were excluded due to scanner malfunction, and one participant was excluded for failing to respond on a significant number (20/72) of test trials. Participants ranged in age from 18 to 45<sup>1</sup> years with the majority in their early twenties (mean age = 24.9, median = 24). Participants were randomly assigned to one of four conditions which differed only in the instructions given prior to the study phase (described below): an implicit dot-pointing (ID\*) condition ( $n = 11$ , 6 women), an implicit configural (IC) condition ( $n = 11$ , 6 women), an explicit dot-pointing (ED) condition ( $n = 11$ , 5 women), or an explicit configural (EC\*) condition ( $n = 11$ , 5 women). (Conditions labeled with an asterisk effectively replicate the conditions used by Reber et al., 2003.)

### Stimuli

Stimuli were Posner and Keele (1968)-styled random dot patterns presented on a rear-projection screen visible in the scanner through a mirror positioned above the participant's eyes. Each stimulus consisted of nine dots arranged on the screen (see Figure 1 for examples). The exact coordinate pairs for each study and test pattern were identical to those used in Reber et al. (2003). However, one difference was that the patterns were displayed as black dots on a white display (due to aliasing of white dots on the scanner projection screen). There were two distinct categories of items (A and B), each of which was generated by making random distortions of a prototypical pattern (Posner & Keele, 1968). For some participants, Category A was the studied category and B was the foil, whereas for others the stimulus sets were reversed. Which category of items was studied in the training phase was determined randomly and counterbalanced along with experimental condition.

### Procedure

#### *Study Phase*

In all four conditions, during the initial study phase, participants were positioned in the MRI scanner, the magnet was shimmed, and then participants viewed the instructions and study stimuli presented on the heads-up display. No scanning took place during the study phase. Following Reber et al. (2003), participants viewed five high distortions of an underlying prototype that were presented in two random blocks for 5000 msec each.

Given the fact that our primary experimental manipulation was the study instructions, we have replicated them verbatim in the Appendix. To briefly summarize, partici-

pants in the explicit learning conditions (EC\* and ED) were told that they were about to view a series of dot patterns that shared some relationship to one another such that they all belonged to the same category. These participants were told that they should attempt to learn what made these items a category, and that after the study phase they would be asked to classify new patterns as either belonging to this studied category or not. In contrast, participants in the implicit learning conditions (IC and ID\*) were given no indication of the categorical nature of the study items. Instead, they were told to simply passively evaluate each stimulus.

Critically, the instructions given in each condition also suggested a particular way to evaluate the stimuli during study. Participants in the implicit dot-pointing condition (ID\*) were told that the purpose of the study was to determine the types of brain activity that govern spatial perception and visual/motor imagination. As a result, they were asked to attempt to identify the central dot in each presented stimulus and imagine themselves pointing to that central dot. This ID\* condition effectively replicated the implicit learning condition of Reber et al. (2003). In comparison, participants in the explicit dot-pointing condition (ED) studied under intentional conditions. However, they were told that the category they were attempting to learn was actually defined by the location of the center dot of each stimulus. Thus, to learn the category, the instructions suggested that participants try to identify the central dot in each pattern and imagine pointing to it. Note that the ID\* and ED conditions overlap in terms of the encoding strategy given to participants, but differ as to the degree of awareness participants had concerning the nature of the category-learning task.

In contrast, in the configural encoding conditions (IC and EC\*), participants were encouraged to encode the overall shape and configuration of the stimuli. (Although we cannot know with certainty the type of encoding processes that participants used in Reber et al.'s explicit learning condition, the present configural encoding instructions may provide a reasonable approximation.) In the IC condition, participants were told that the purpose of the experiment was to determine the types of brain activity that govern the perception of shapes and configurations, and thus, they should evaluate the overall shape of each pattern. In the EC\* condition (which effectively replicated Reber et al.'s explicit condition), participants were asked to learn the category defined by the overall shape and configuration of the study phase stimuli. In summary, our design is a  $2 \times 2$  factorial that manipulates learning mode (implicit vs. explicit) and stimulus encoding (dot-pointing vs. configural).

#### *Test Phase (Scanned)*

Immediately after the learning phase, all participants were given a new set of instructions informing them that they had just studied items from a category and that in the following

session they would see new items that either belonged or did not belong to the studied category. Participants then viewed 72 novel dot patterns presented one at a time. One half of the transfer trials were stimuli that belonged to the studied category (in the sense that they were distortions of the same underlying prototype) and half were patterns drawn from the unstudied category. Of the 36 patterns from each category, 4 were presentations of the actual prototype (P), 16 were low distortions (L), and 16 were high distortions (H). The presentation of stimuli followed a rapid event-related design. Stimuli were presented on the screen for a variable amount of time based on random, exponentially distributed jitter (ranging from either 4, 8, 10, or 12 sec). Participants viewed each item and were instructed to respond either yes or no via a response paddle, depending if they believed the item belonged to the studied category. After participants indicated their response, the stimulus disappeared and a blank screen was presented until the next trial began. If a participant did not respond before the end of the trial, the trial was coded as “no response” and the experiment continued normally. The maximum number of trials scored as “no response” was two per participant with the majority ( $n = 28$ ) of participants responding on every trial.

### fMRI Image Acquisition and Analysis

fMRI data were acquired using a Siemens Magnetom TRIO 3-T whole-body MRI located at the Indiana University Bloomington Imaging Research Facility and equipped with an eight-channel phased-array head coil. The field of view was  $22 \times 22 \times 9.9$  cm, with an in-plane matrix of  $64 \times 64$  pixels and 33 axial slices per volume (whole brain), creating a voxel size of  $3.44 \text{ mm} \times 3.44 \text{ mm} \times 3 \text{ mm}$ . Images were collected using a gradient-echo EPI sequence (TE = 30 msec, TR = 2000 msec, flip angle =  $70^\circ$ ) for BOLD imaging. High-resolution T1-weighted anatomical volumes were acquired using Turbo-flash 3-D sequence (TI = 1100 msec, TE = 3.93 msec, TR = 14,375 msec, flip angle =  $12^\circ$ ) with one hundred sixty 1-mm sagittal slices and an in-plane field of view of  $224 \times 256$  (voxel size =  $1 \text{ mm}^3$ ).

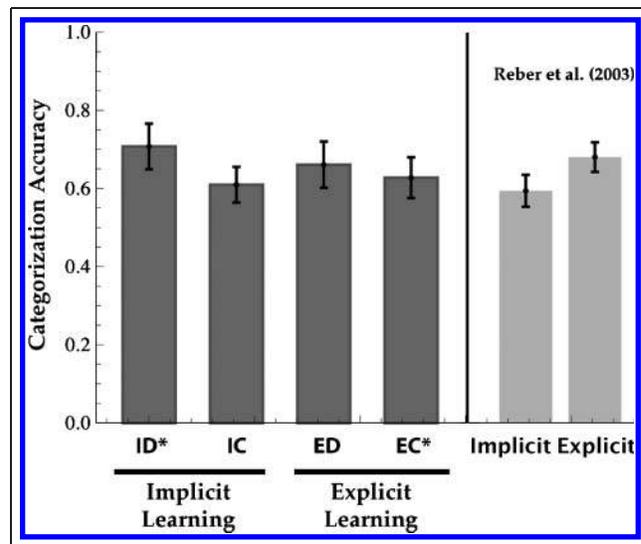
Imaging data were preprocessed using the Brain Voyager 3-D analysis program (Maastricht, The Netherlands). Anatomical volumes were transformed into a common stereotactic space (Talairach & Tournoux, 1988). Functional data were coregistered to the anatomical volumes using an eight-parameter affine transformation, transforming the functional data to a common stereotactic space across participants. Functional data underwent a linear trend removal, 3-D spatial Gaussian filtering (FWHM 7.5 mm), slice scan-time correction, and 3-D motion correction. Transformed functional data were analyzed using Brain Voyager’s multi-subject random-effects general linear model procedure to produce statistical parametric maps. Whole-brain maps were created with a  $p < .001$  voxelwise threshold and a minimum cluster size threshold of 10 voxels as recommended

by Thirion et al. (2007). The cluster threshold correction technique used here is a two-stage analysis that helps to control false positives, but with a relative sparing of statistical power, which was important for studying the small effect sizes seen between our experimental conditions (Forman et al., 1995). Corrections similar to this have been found to successfully manage the multiple testing problem.

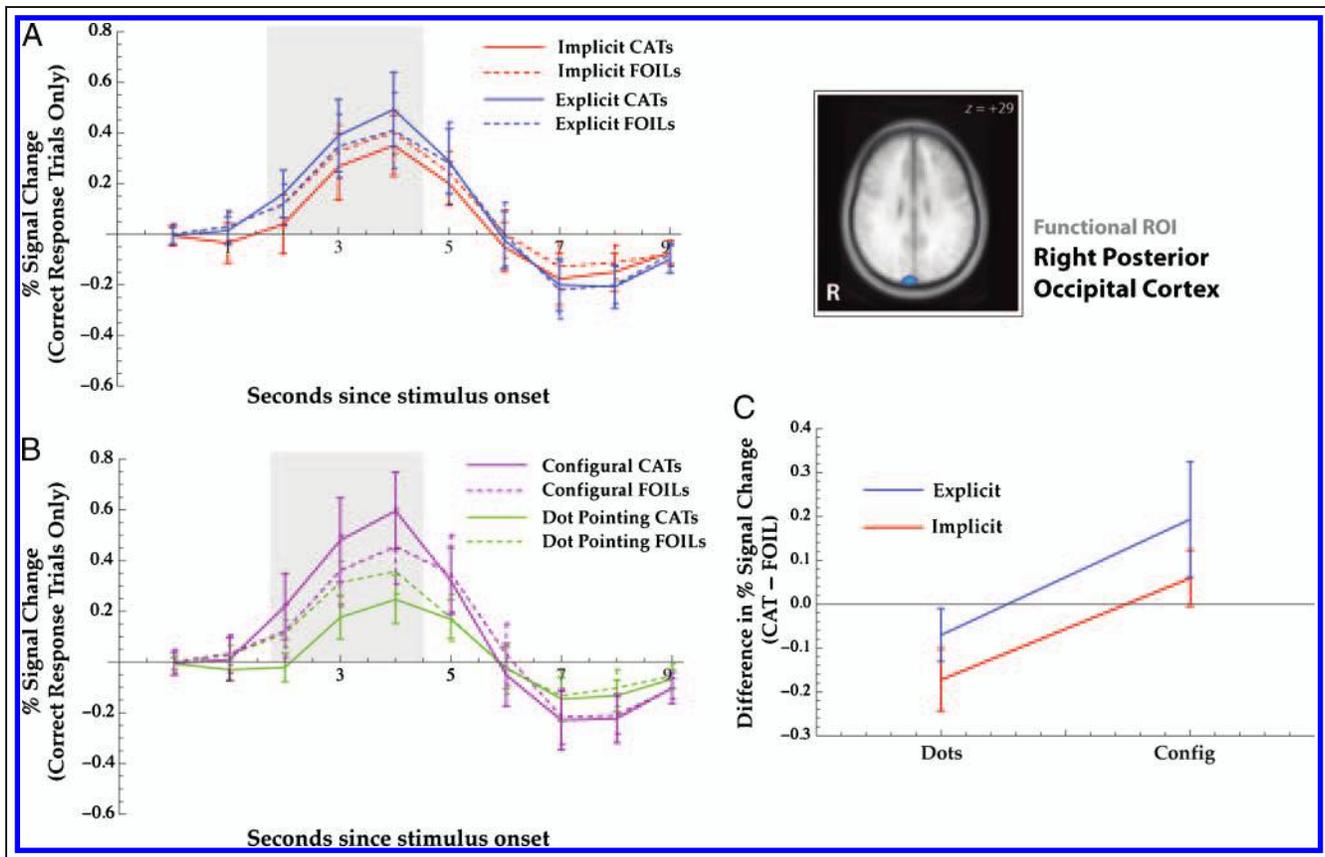
## RESULTS

### Behavioral Results

Figure 2 shows the pattern of behavioral results in the categorization task. A two-way ANOVA on the probability of correct response with both stimulus encoding (dot-pointing, configural) and learning mode (implicit, explicit) as between-group factors failed to detect a significant effect of either stimulus encoding [ $F(1, 40) = 0.069, p > .8$ ], learning mode [ $F(1, 40) = 1.47, p > .23$ ], or an interaction [ $F(1, 40) = .35, p > .5$ ]. All four groups performed significantly better than chance (50%). The ID\* group averaged 70.8% correct [ $SD = 19.4\%$ ;  $t(10) = 3.55, p < .01$ ], the IC group averaged 60.9% [ $SD = 15.1\%$ ;  $t(10) = 2.39, p < .04$ ], the ED group averaged 66.1% [ $SD = 19.5\%$ ;  $t(10) = 2.73, p < .03$ ], and the EC\* group averaged 62.8% [ $SD = 17.3\%$ ;  $t(10) = 2.44, p < .04$ ]. Overall performance in our EC\* condition was virtually identical to what had been observed in that condition in the previous study by Reber et al., whereas performance in our ID\* condition was slightly (but not significantly) better than in Reber et al.’s study.



**Figure 2.** The overall categorization accuracy in each of the four experimental conditions along with the behavioral data from Reber et al. (2003). Conditions are labeled according to the following: ID\* = implicit dot-pointing instructions; IC = implicit configural encoding instructions; ED = explicit dot-pointing instructions; EC\* = explicit configural encoding instructions. The \* indicates conditions that replicate Reber et al. (2003).



**Figure 3.** (A and B) A time-series plot of the deconvolved BOLD signal following stimulus onset for each experimental factor, in the occipital ROI (slice centered at  $z = +28$ ). The solid lines show the BOLD response for presentation of category members that were correctly classified, and the dashed lines are for foil items that were correctly classified. Panel A compares activity for the implicit and explicit learning conditions (IC and ID\*) versus (EC\* and ED). Panel B compares activity for the dot-pointing (ID\* and ED) and configural encoding conditions (EC\* and IC). The shaded regions show the portion of the time series which was analyzed in more detail in Panel C. (C) The difference in percent signal change for category members minus foils in each individual condition. Also shown is a picture of the ROI under consideration. All error bars are standard errors.

As will be seen, however, our brain imaging results will strongly replicate Reber et al.'s findings in the ID\* condition.

We found no significant differences in overall RT between any of the conditions (all  $t < 1$ ). In addition, we did not find any significant effects of RT depending on learning mode or stimulus-encoding instructions (again, all  $t < 1$ ). Finally, in the analyses below, we focus on the subset of trials where participants responded correctly. There was no reliable difference in correct response RT depending on whether the stimulus was a category member or foil for any individual condition (all  $t < 1$ ), or when collapsing across all conditions ( $t < 1$ ). In addition, a  $2 \times 2$  ANOVA on the difference in RT for CAT-FOIL (correct trials only) found no effects of learning mode, stimulus encoding, nor an interaction (all  $F < 1, p > .5$ ).

### Imaging Results: Targeted ROI Analyses

In Reber et al. (2003), the strongest evidence for a dissociation between implicit and explicit learning came from a targeted ROI analysis that compared activity in the func-

tionally defined region of posterior occipital cortex and the anatomically defined region of the left anterior hippocampus. To verify that our results replicated Reber et al., and to directly assess the hypothesis about implicit learning, we conducted a set of analogous ROI analyses. For each condition, and for both category members and foil patterns, time courses measuring the average BOLD response (represented as beta weights) in each ROI over a 10-sec window following stimulus onset were computed via a deconvolution analysis.

#### Posterior Occipital Cortex ROI

Figure 3A and B compare the time course of BOLD activity for category members and foils (correct response trials only) in a posterior occipital ROI selected based on significant voxels in the ID\* condition. This initial region was selected based on significant voxels in the ID\* condition at  $p < .001$ , and was expanded to include all contiguous voxels in the region passing the  $p < .005$  threshold in order to ensure broad coverage. Panel A shows the pattern

of neural response for participants in the explicit learning conditions (EC\* and ED) compared to the implicit learning conditions (IC and ID\*). Panel B shows the same analyses comparing the dot-pointing (ID\* and ED) and configural encoding conditions (EC\* and IC). Finally, Figure 3C shows the percent signal change difference between category members and foils in each condition during a 3-sec window beginning 2 sec after stimulus onset lasting until 4 sec after onset. A  $2 \times 2$  ANOVA on these difference scores revealed no effect of learning mode [ $F(1, 40) = 1.85, p = .18$ ], a reliable effect of stimulus-encoding instructions [ $F(1, 40) = 8.16, p = .007$ ], and no interaction ( $F < 1$ ). Consistent with Reber et al. (2003), we found that differential activity was lower in the ID\* condition relative to the EC\* condition [ $t(20) = 2.44, p = .02$ ]. Also note that although the absolute mean difference in activation was positive for the EC\* condition, this value did not significantly differ from zero [ $t(10) = 1.46, p = .17$ ]. In contrast, differential activity in the ID\* condition was significantly negative [ $t(10) = 2.44, p = .03$ ].

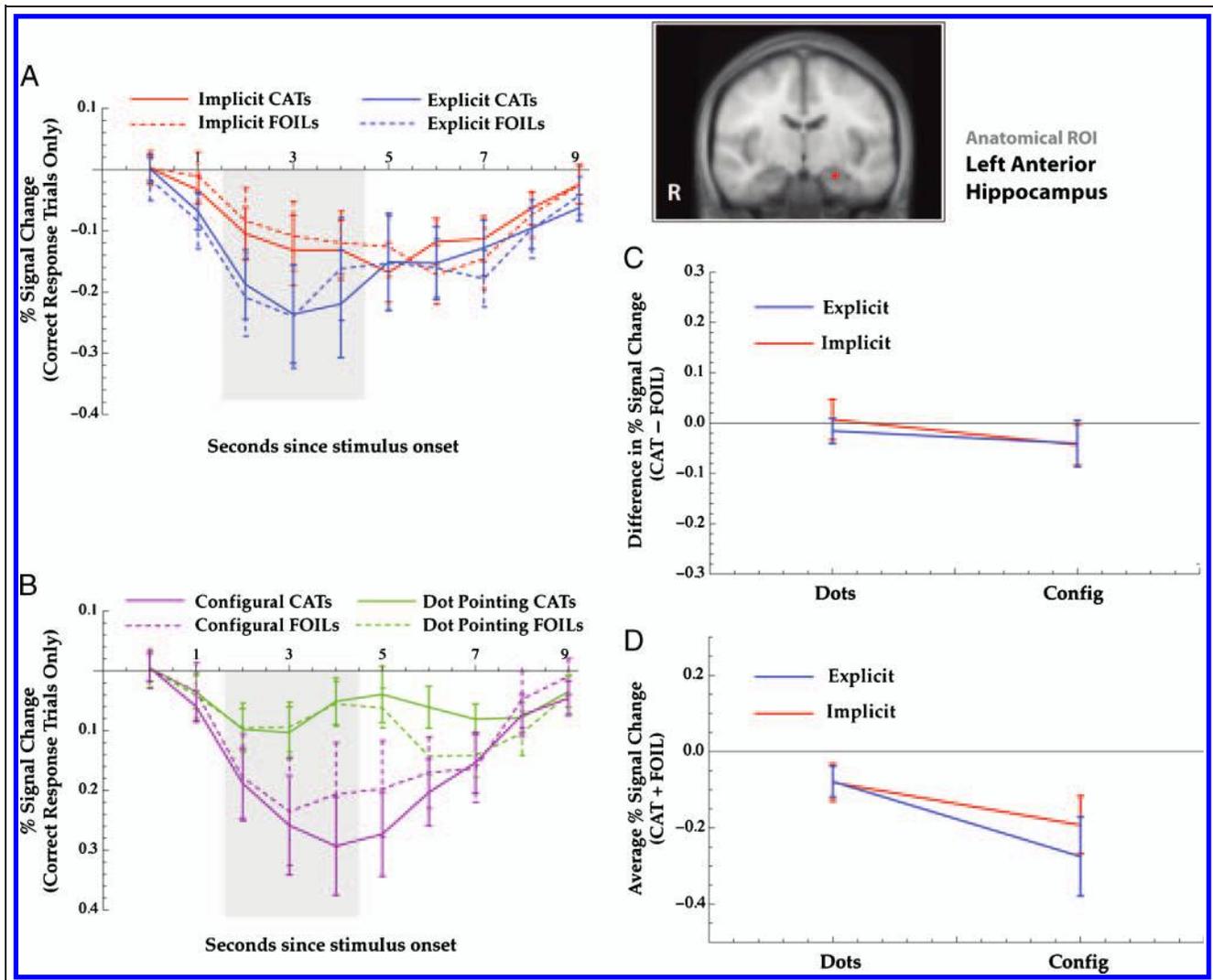
As a further test, within each of the dot-pointing and configural conditions, we considered if pairwise contrasts would reveal an effect of implicit–explicit learning. A contrast between the ID\* and ED conditions was not reliable [ $t(20) = 1.3, p = .2$ ], and neither was the contrast between IC and EC\* ( $t < 1$ ). However, the contrast between the ID\* and IC conditions was significant [ $t(20) = 2.4, p = .02$ ], whereas the contrast between the ED and EC\* conditions was marginal [ $t(20) = 1.81, p = .08$ ]. We also replicated the above analysis considering all trials (rather than only those where participants made a correct response), with similar results.

Note that like Reber et al., the ROI identified in the ID\* condition was selected based on significant voxels observed in a whole-brain contrast. This analysis represents a minor incursion of the nonindependent ROI analysis (Poldrack & Mumford, 2009) because the ID\* member–foil difference is biased to be larger than the differences in the other conditions. In a series of follow-up analyses, we considered independently selected occipital lobe ROIs that establish the robustness of our findings. First, we constructed an ROI in the left hemisphere that paralleled the region of occipital cortex selected based on the significant voxels in the ID\* condition (which was limited to the right hemisphere). An analysis on the differential activity (identical to the above) in this region found no effect of implicit–explicit learning [ $F(1, 40) = 0.039, p = .84$ ], a main effect of encoding instructions [ $F(1, 40) = 7.51, p = .009$ ], and no interaction ( $F < 1$ ). In addition, we replicated our analyses using an ROI based on the occipital lobe coordinates reported by Reber et al. [+11, –91, +18] (a cube volume of  $10 \times 10 \times 10$  voxels was created centered at this location). An analysis here revealed no effect of learning mode [ $F(1, 40) = 2.04, p = .16$ ], a reliable effect of stimulus-encoding instructions [ $F(1, 40) = 6.73, p = .01$ ], and no interaction [ $F(1, 40) = 1.23, p = .27$ ]. As above, in this region, differential activity was more negative in the ID\*

condition relative to the EC\* condition [ $t(20) = 2.46, p = .02$ ]. Also note that although the absolute mean difference in activation was positive for the EC\* condition, this value did not significantly differ from zero [ $t(10) = 1.62, p = .14$ ]. In contrast, differential activity in the ID\* condition was marginally negative [ $t(10) = 1.91, p = .08$ ]. Finally, a pairwise contrast of ID\* and IC was not reliable in this ROI [ $t(20) = 1.2, p = .24$ ], whereas activity in ED condition was significantly lower than in the EC\* condition [ $t(20) = 2.35, p = .028$ ]. In sum, these analyses point strongly to an effect of encoding conditions on patterns of activity in posterior occipital cortex, but to little, if any, effect of learning mode (implicit vs. explicit).

#### *Left Anterior Hippocampus ROI*

Our second ROI analysis considered a region of the left anterior hippocampus, under the hypothesis that this region might show differential involvement in explicit processing of the category members. Given the lack of activation in this region in the whole-brain statistical parametric map, we created an ROI based on established anatomical markers (Talairach & Tournoux, 1988) in the averaged brain which we then verified against each individual subject's anatomical scan for reliability. The actual ROI reported in the present analyses was centered at (–22, –9, –15) and was a cube of  $32 \text{ mm}^3$ . Figure 4A and B compare the time course of BOLD activity for category members and foils (correct response trials only) in this region. In contrast to the occipital area, this region shows a pattern of deactivation following stimulus onset. Figure 4C shows the percent signal change difference between category members and foils in each condition over the same time window considered in the occipital region. A  $2 \times 2$  ANOVA on these difference scores revealed no main effects or interactions (all  $F < 1$ ) and this null pattern was robust to alternative analyses including peak activation or alternative time windows. We also considered a number of alternative hippocampal ROIs (created in a similar way as described above), including bilateral regions, larger and smaller regions, and regions of the posterior hippocampus, all with identical null results.<sup>2</sup> However, as Figure 4 shows, activity in the hippocampus did appear to selectively deactivate following stimulus presentation. In light of this, and the null results involving category-specific activity, we conducted analyses on changes in overall activity (i.e., neural activity in response to the test stimulus independent of category/foil membership). A  $2 \times 2$  ANOVA on the total average percent signal change in response to category or foil trials revealed no effect of learning mode ( $F < 1$ ), a marginal main effect of stimulus-encoding instructions [ $F(1, 40) = 3.28, p < .08$ ], and no interaction ( $F < 1$ ; see Figure 4D). We should note that the marginal effect of encoding reaches significance ( $p < .05$ ) when a slightly larger time window is considered. Thus, although we failed to find a strong effect of category-specific memory in this ROI, overall activity was marginally lower during correct responses in the configural encoding conditions



**Figure 4.** (A and B) A time-series plot of the deconvolved BOLD signal following stimulus onset for each experimental factor in the hippocampal ROI. The details of these panels parallel Figure 3A and B. Again, the shaded regions show the portion of the time series which was analyzed in more detail in Panel C. (C) The difference in percent signal change for category members minus foils in each individual condition. (D) Differences in overall activity independent of the stimulus class (i.e., category members and foil patterns are grouped together). Also shown is an example of the ROI under consideration. All error bars are standard errors.

(IC and EC\*) compared to the dot-pointing conditions (ED and ID\*).

### Whole-brain Analyses

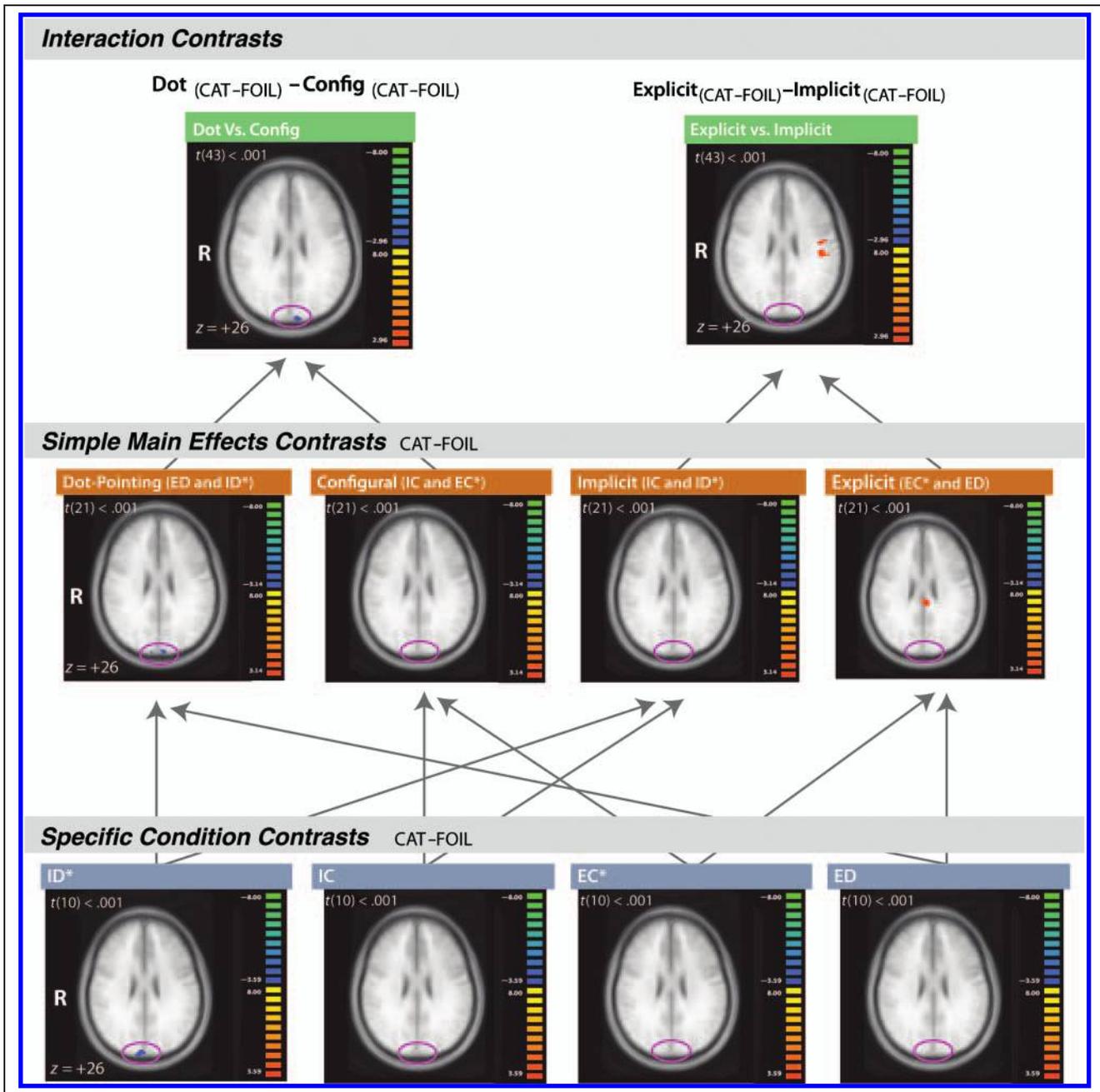
In addition to our targeted ROI analyses, we also conducted a series of supplemental whole-brain analyses, which give a broader sense of the anatomical regions that were differentially engaged during the test phase. Figure 5 shows a single horizontal slice (at  $z = +26$ ) that passes through the occipital lobe ROI identified in the ID\* condition and that was analyzed in the ROI analysis described above. Three distinct classes of statistical contrasts are shown. First, for each individual condition of the experiment (ID\*, IC, EC\*, and ED), we show voxels corresponding to differential BOLD activity for category members compared to foil pat-

terns. These images are shown in the bottom row of Figure 5 (labeled “Specific Condition Contrasts”) and most closely correspond to the analyses reported in Reber et al. (2003). As is visible, we found a strong pattern of deactivation for category members compared to foils in the ID\* condition in right posterior occipital cortex (+6, -89, +29), with no significant activity in this region for the other conditions. Note that the locus of this activity was within millimeters of previous reports of occipital involvement in implicit dot-pattern classification and in the same hemisphere, although slightly dorsal [Reber et al., 2003: (+11, -91, 18); Reber et al., 1998b: (-14, -86, +1), (+1, -92, +7), and (+9, -87, +11); and Reber et al., 1998a: (+1, -88, -4) and (+12, -93, +17); but see Aizenstein et al., 2000: (+9, -71, +41)].

Second, we repeated this analysis on each level of each factor in our  $2 \times 2$  design (e.g., the dot-pointing conditions,

ID\* and ED, and the explicit learning condition, EC\* and ED). These images are shown in the middle row of Figure 5 (labeled “Simple Main Effects Contrasts”). In the dot-pointing conditions (i.e., ID\* and ED), we found a bilateral pattern of deactivation for category members compared to foil patterns in posterior occipital cortex centered at  $(-5, -89, +24)$  and  $(+10, -91, +24)$  (see Table 1). Again, this region is anatomically similar to previous reports of implicit

learning of dot patterns in the ID\* condition. There was no significant activation in this region in the other main effect contrasts. One interpretation of these results is that the dot-pointing effect is driven by the significant voxels in the ID\* condition; however, this account fails to simultaneously explain the presence of the effect in the combined dot-pointing conditions and the absence of the effect in the combined implicit conditions.



**Figure 5.** The regions of significant BOLD activity are shown for a variety of statistical contrasts. The bottom row shows the differential pattern of neural activity for category members and foil patterns (CAT-FOIL) for each individual specific condition. The middle row presents the same contrast for the simple main effects of the  $2 \times 2$  design. Finally, the top row presents a contrasts-of-contrasts (i.e., interaction) analysis that compares activity between the dot-pointing and configural conditions (top left) and the explicit and implicit conditions (top right). All horizontal slices were centered at  $z = +26$ , which passes near the center of the cluster of negative activity in the occipital lobe found in the dot-pointing conditions in various contrasts. The purple ring highlights the pattern of activity in this surrounding region for each contrast. Arrows denote the dependencies between different statistical contrasts (e.g., the dot-pointing condition combines subjects from the ED and ID\* conditions).

**Table 1.** Significant Clusters of Differential Activation for Category Members Relative to Foils for Individual Conditions and for the Analyses That Collapsed across Our Experimental Factors

<i>Anatomic Area</i>	<i>Broadmann's Area</i>	<i>Talairach Coordinates</i>			<i>t</i>	<i>Voxels</i>
		<i>x</i>	<i>y</i>	<i>z</i>		
<i>Pooled Conditions</i>						
Incidental (IC and ID*)						
Lingual gyrus	19	+1	-70	-6	$t(21) = -5.56$	2856
Explicit (EC* and ED)						
Inferior occipital gyrus	18	-25	-87	-4	$t(21) = 5.70$	876
Ant. parahippocampal gyrus	27	+19	-18	-4	$t(21) = 6.46$	1081
Insula	13	-43	-8	-1	$t(21) = 5.69$	629
Medial frontal gyrus	8	-8	-32	+39	$t(21) = 5.72$	2989
Precentral gyrus	4	-30	-15	+49	$t(21) = 5.41$	829
Anterior cingulate cortex	24	0	+3	+45	$t(21) = 4.919$	668
Fusiform gyrus	37	-44	-60	-15	$t(21) = 6.34$	1460
Dot-pointing (ID* and ED)						
R. posterior occipital cortex	19	+10	-91	24	$t(21) = -4.166$	295
L. posterior occipital cortex	19	-5	-89	24	$t(21) = -4.732$	371
Configural (IC* and EC)						
None						
<i>Individual Conditions</i>						
Incidental dots (ID*)						
R. posterior occipital cortex	19	+6	-89	+29	$t(10) = -6.71$	619
Explicit dots (ED)						
Precuneus	7	+6	-54	+54	$t(10) = 8.00$	855
Medial frontal gyrus	6	-17	-8	+54	$t(10) = -6.99$	352
Incidental configural (IC)						
None						
Explicit configural (EC*)						
None						

Although the above analyses are useful for illustrating the whole-brain patterns produced by different combinations of conditions, they do not directly test the most important question, namely, whether there are significant *differences* in activity across conditions. To address this, the statistical maps in the top row of Figure 5 show the results of contrast-of-contrasts tests, which compare the activity between the levels of each factor. Regions of significant activity in these contrasts thus reflect areas that show a category-specific effect for one condition, above and beyond the effect observed in the opposing condition (e.g., dot-pointing vs. configural or explicit vs. implicit). In the statistical contrast comparing category-specific activity for dot-pointing instructions and configural instructions, we

see a pattern of deactivation similar to the one observed in the combined dot-pointing conditions (Figure 5, middle row, far left). In the statistical contrast comparing implicit and explicit learning (Figure 5, top row, right), we see no evidence of differential engagement of the occipital lobe, but a number of regions of significantly positive activity in other brain regions (see also Figure 6 for a complete description of this pattern which is not completely visible on the slice in Figure 5).

The preceding analyses focused on the pattern of category-specific activity in and around the posterior occipital lobe. However, there were a number of other regions of significant category-specific activity identified across the various contrasts that are not visible in Figure 5. First, in the implicit

learning conditions (ID\* and IC), a single cluster of significant activity was detected centered on the lingual gyrus (+1, -70, -6) (see Table 1). Reber et al. (2003) did not report activity in this region, whereas Reber et al. (1998a, 1998b) found a similar region of differential CAT > FOIL activity located at (+1, -88, -4) and (-14, -86, +1), respectively, in an implicit dot-pattern classification task similar to the ID\* condition. Note, however, that in our study no part of this region reached significance in the contrast that directly controlled for activity in this region during explicit learning (i.e.,  $\text{Explicit}_{(\text{CAT-FOIL})} - \text{Implicit}_{(\text{CAT-FOIL})}$ ).

Second, in the explicit learning conditions, we found increased activity in a number of cortical regions including the parahippocampal gyrus, fusiform gyrus, and cingulate (see Table 1 for a full list). Figure 6 shows the activity for the explicit condition (i.e., CAT-FOIL for the combined EC\* and ED conditions, indicated by orange-yellow) and the differences-of-differences contrast that compared implicit and explicit learning (i.e.,  $\text{Explicit}_{(\text{CAT-FOIL})} - \text{Implicit}_{(\text{CAT-FOIL})}$ , indicated in purple). Overall both of these contrasts reveal increased activity across a number of cortical regions, which is consistent with previous studies of explicit learning in dot-pattern classification and memory tasks (Reber et al., 1998a, 1998b, 2003; Aizenstein et al., 2000). However, across all of these studies, there is considerable variability in the regions identified. In addition, few of these regions are uniquely implicated in explicit memory in other memory studies (as opposed to the hippocampal ROI considered above).

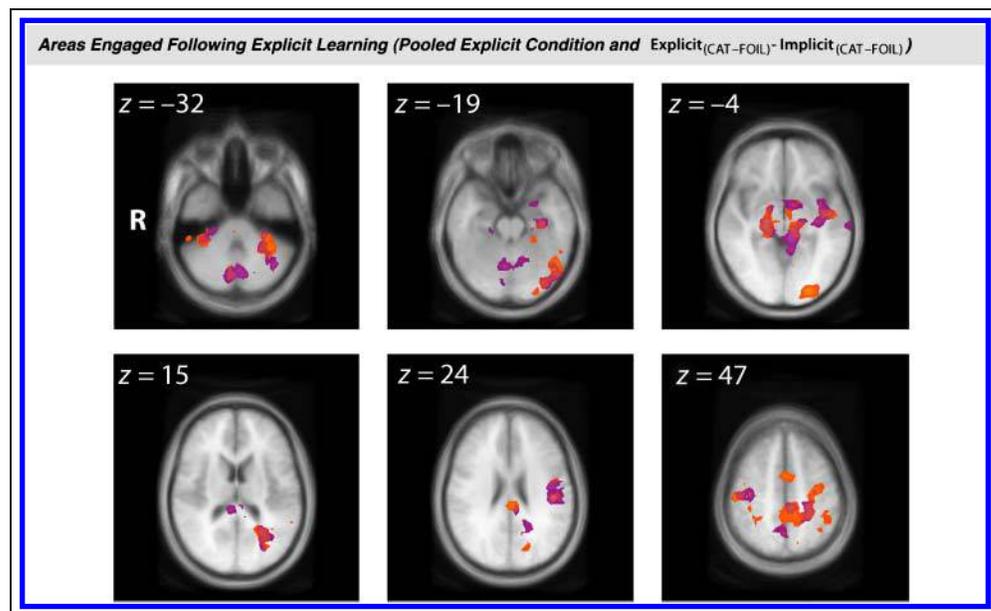
## GENERAL DISCUSSION

In a visual category-learning task, which extends the paradigm developed by Reber et al. (2003), we found that the

encoding instructions given to participants at the time of study strongly influenced the pattern of brain activity observed at test, independent of the implicit-explicit status of the learner. This effect was most pronounced for a region of posterior occipital cortex, which has previously been associated with implicit visual memory for dot-pattern stimuli (Poldrack & Foerde, 2008; Smith, 2008; Reber et al., 1998a, 1998b, 2003). Differential activity in this region for category members compared to foil patterns appeared as deactivations when participants were told to identify the “central dot” of each stimulus presented at study, and was more positively activated when participants were asked to evaluate the “overall shape and configuration” of the dot patterns. This result suggests that category-specific activity in this region may be better explained in terms of how people visually processed or encoded the stimuli than in terms of the explicit-implicit status of the learner.

Overall, this finding has important implications for our understanding of the neural systems underlying visual category learning. A number of recent review articles (Poldrack & Foerde, 2008; Smith, 2008) suggest that the Reber et al. result provides the clearest demonstration to date of a double dissociation between implicit and explicit category learning. Although we cannot rule out the possibility that the implicit-explicit distinction plays a role in the observed pattern of neural activation (because doing so would require arguing from a null result), after controlling for stimulus-encoding effects and the activity during explicit learning, we found no clear evidence that implicit learning differentially modulated activity in these brain regions. Such a result is consonant with the assumptions of single-system models of categorization that do not posit a fundamental distinction between implicit versus explicit modes of learning. At the very least, our results appear to reopen aspects of the debate about the nature of im-

**Figure 6.** Regions of differential activity for category members compared to foils (CAT-FOIL) in the explicit learning conditions (EC\* and ED, indicated in orange) and the differences-of-differences contrast directly comparing explicit and implicit learning ( $\text{Explicit}_{(\text{CAT-FOIL})} - \text{Implicit}_{(\text{CAT-FOIL})}$ , indicated in purple). These areas supplement the analysis in Figure 5 by showing alternative brain slices.



PLICIT learning in certain visual categorization tasks (but see Nomura et al., 2007 for another type of neural dissociation found in procedural learning tasks).

Overall, our findings appear most consistent with a processing account of brain activity whereby the neural systems recruited during encoding participate in subsequent remembrance (Vaidya, Zhao, Desmond, & Gabrieli, 2002; Craik & Lockhart, 1972). Such an account suggests that the signature for subsequent memory processing may be a continuum that is modulated by the nature or depth of encoding processes rather than by strictly dichotomous implicit–explicit learning systems. Note that the claim of “single-system” models of categorization is not that there are no multiple ways of encoding and processing stimuli and representing them in the brain. Instead, the key issue is whether or not distinct implicit versus explicit neural systems underlie different forms of category learning. In our view, a general processing account of brain activity is in keeping with a single-system approach.

Our finding that activity in the occipital lobe can be modulated by the nature of the stimulus-encoding instructions raises questions concerning what aspect of the dot-pointing task is correlated with activation in the posterior occipital lobe. One possibility is that the changes in category-specific activity reflect differences in activity for attending to particular spatial locations (which is, in turn, dictated by the visual properties of the test materials). A spatial processing or working memory account might also explain why category-specific deactivations are reduced in this region when learners adopt a more “configural” encoding strategy, because, in these conditions, participants are less likely to attend to the same spatial location on every test trial. Note that the engagement of posterior occipital cortex (specifically BA 19) has been implicated in spatial working memory processes in numerous studies (Postle & D’Esposito, 1999; Ungerleider, Courtney, & Haxby, 1998; Smith et al., 1995), and is consistent with the view of the dorsal visual pathway representing “where” information. A related account is that participants used different patterns of eye movements to explore the stimuli in the different conditions, which may differentially recruit dorsal occipital and occipito-parietal regions such as V3A and cIPS (Culham, Cavina-Pratesi, & Singhal, 2006; Culham & Kanwisher, 2001).

Finally, note that like Reber et al., we found evidence for increased neural activity following explicit or intentional learning across a variety of brain regions (see Figure 6 and Table 1). Because these positive activations were very widespread and the precise areas of activation appear to vary across studies, they do not seem to be pointing to a specific, dissociable neural learning system. Future research is needed to help interpret the generalized mental processes that appear to be reflected in these widespread and perhaps interacting neural networks.

In sum, previous highly influential findings have supported the hypothesis that implicit and explicit category learning are mediated by separate and specific dissociable brain systems. Our current results suggest instead that

those findings may have reflected alternative processing and stimulus-encoding strategies across conditions. Thus, in our view, our results raise important points of clarification about the degree to which completely dissociable brain systems mediate implicit and explicit category learning of visual patterns.

## APPENDIX: INSTRUCTIONS

### Training Instructions

#### *ID\* Condition (Implicit Dot-pointing)*

In this first phase of the experiment, you will view a series of dot patterns, one at a time. Your job is simply to try to identify the central dot in each pattern and imagine yourself pointing to that central dot. The purpose of the experiment is to determine the types of brain activity that govern spatial perception and visual/motor imagination. Therefore, in this phase, it is important for you to try to identify the location of the central dot in each of the training patterns and to imagine pointing to it on the screen.

#### *ED Condition (Explicit Dot-pointing)*

In this first phase of the experiment, you will view a series of dot patterns, one at a time. Each dot pattern comes from the same category of items. Your job in this first phase of the experiment is to try to learn the category. The category is determined by the location of the central dot in each pattern. To learn the category, simply try to identify the central dot in each pattern and imagine yourself pointing to that central dot. After the first phase is completed, you will be presented with some new dot patterns, some of which belong to the category and some of which do not. Category membership will again be determined by the location of the central dot in each pattern. Therefore, in this phase, it is important for you to try to identify the location of the central dot in each of the training patterns and to imagine pointing to it on the screen.

#### *IC Condition (Implicit Configural)*

In this first phase of the experiment, you will view a series of dot patterns, one at a time. Your job is simply to view and mentally evaluate the entire shape and configuration of each pattern. The purpose of the experiment is to determine the types of brain activity that govern the perception of shapes and configurations. Therefore, in this phase, it is important for you to mentally evaluate the entire shape and configuration of each pattern.

#### *EC\* Condition (Explicit Configural)*

In this first phase of the experiment, you will view a series of dot patterns, one at a time. Each dot pattern comes

from the same category of items. Your job in this first phase of the experiment is to try to learn the category. The category is determined by the overall shape and configuration of dots in the pattern. To learn the category, carefully study each dot pattern and try to evaluate its overall shape and configuration. After the first phase is completed, you will be presented with some new dot patterns, some of which belong to the category and some of which do not. Category membership will again be determined by the overall shape and configuration of the dots in each pattern. Therefore, in this phase, it is important for you to carefully study and evaluate the overall shape and configuration of each pattern that you view.

### Test Instructions (All Conditions)

The dot patterns that you just viewed all came from the same category, in the same sense that if you had viewed a set of pictures of dogs, all would come from the category “DOG.” In this test phase, we will present you with a new set of dot patterns. About half of these new patterns belong to the same category of dot patterns that you just viewed, whereas the other half belong to some different category. For each dot pattern that is presented, please indicate whether you judge it to be a member of the category that you experienced during the first phase. Press the “YES” response button if you judge the pattern to be a member of the category and press the “NO” response button if you judge that the pattern is not a member of the category.

### Acknowledgments

This work was supported by NIH-NIMH training grant TM32 MH019879-12 and startup funds provided by New York University to T. M. Gureckis, a grant to T. W. James from Indiana University’s Faculty Research Support Program administered by the Office of the Vice Provost for Research, and in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc, and MH48494 from NIMH and Grant FA9550-08-1-0486 from the Air Force Office of Scientific Research to R. M. Nosofsky. We thank Shiling Choo and Samantha Brandfon for assistance in subject recruitment, Katherine Duncan for help in constructing our hippocampal ROI, and Paul Reber for providing the stimuli used in his study.

Reprint requests should be sent to Todd M. Gureckis, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, or via e-mail: todd.gureckis@nyu.edu.

### Notes

1. The single 45-year-old subject was in the ED condition. None of our conclusions changes if the single participant is deleted from the analyses.
2. The ROI we settled on was most conservative in that it was the one that appeared to be anatomically consistent across all the individual subjects and aligned most with Reber et al. (2003).

### REFERENCES

- Aizenstein, H., MacDonald, A., Stenger, V., Nebes, R., Larson, J., Ursu, S., et al. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience*, *12*, 977–987.
- Ashby, F., Alfonso-Reese, L., Turken, A., & Waldron, E. (1998). A neuropsychological theory of multiple system in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F., & Ell, S. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*, 204–210.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.
- Culham, J. C., Cavina-Pratesi, C., & Singhal, A. (2006). The role of parietal cortex in visuomotor control: What have we learned from neuroimaging? *Neuropsychologia*, *44*, 2668–2684.
- Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*, *11*, 157–163.
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, *33*, 636–647.
- Kéri, S. (2003). The cognitive neuroscience of category learning. *Brain Research Reviews*, *43*, 85–109.
- Knowlton, B., & Squire, L. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.
- Koenig, P., Smith, E. E., Troiani, V., Antani, S., McCawely, G., Moore, P., et al. (2008). Medial temporal lobe involvement in an implicit memory task: Evidence of collaborating implicit and explicit memory systems for fMRI and Alzheimer’s disease. *Cerebral Cortex*, *18*, 2831–2843.
- Kolodny, J. (1994). Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. *Psychological Science*, *5*, 164–169.
- Love, B., & Gureckis, T. (2007). Models in search of the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 90–108.
- Maddox, W., & Ashby, F. (2004). Dissociating explicit and procedural-learning based system of perceptual category learning. *Behavioural Processes*, *66*, 309–332.
- Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., et al. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, *17*, 37–43.
- Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*, 247–255.
- Palmeri, T. J., & Flanery, M. A. (1999). Learning about categories in the absence of training. *Psychological Science*, *10*, 526–530.
- Poldrack, R., Clark, J., Pare-Blagoev, E., Shohamy, D., Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546–550.
- Poldrack, R., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews*, *32*, 197–205.
- Poldrack, R. A., & Mumford, J. A. (2009). Independence in ROI analysis: Where is the voodoo? *Social Cognitive & Affective Neuroscience*, *4*, 208–213.

- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.
- Postle, B., & D'Esposito, M. (1999). “What”–then–“Where” in visual working memory: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *11*, 585–597.
- Reber, P., Gitelman, D., Parrish, T., & Mesulam, M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience*, *15*, 574–583.
- Reber, P., Stark, C., & Squire, L. (1998a). Contrasting cortical activity associated with category memory and recognition memory. *Learning and Memory*, *5*, 420–428.
- Reber, P., Stark, C., & Squire, L. (1998b). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 747–750.
- Shanks, D. R., & St. John, M. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.
- Smith, E. (2008). The case for implicit category learning. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 3–16.
- Smith, E., Jonides, J., Koeppe, R., Awh, E., Schumacher, E., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience*, *7*, 337–356.
- Smith, E., Patalano, A., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*, 167–196.
- Smith, J., Minda, J., & Washburn, D. (2004). Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General*, *133*, 398–414.
- Squire, L., & Knowlton, B. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *92*, 12470–12474.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., & Poline, J.-B. (2007). Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*, *35*, 105–120.
- Ungerleider, L., Courtney, S., & Haxby, J. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 883–890.
- Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. E. (2002). Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas. *Neuropsychologia*, *40*, 2136–2143.
- Zaki, S. R. (2004). Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bulletin & Review*, *11*, 1048–1054.
- Zaki, S. R., & Nosofsky, R. (2001). A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, *1*, 344–359.