

Dissociating Explicit and Implicit Category Knowledge with fMRI

Paul J. Reber, Darren R. Gitelman, Todd B. Parrish,
and M. Marsel Mesulam

Abstract

■ Neuroimaging of healthy volunteers identified separate neural systems supporting the expression of category knowledge depending on whether the learning mode was intentional or incidental. The same visual category was learned either intentionally or implicitly by two separate groups of participants. During a categorization test, functional magnetic resonance imaging (fMRI) was used to compare brain activity evoked by category members and nonmembers. After implicit learning, when participants had learned the category incidentally,

decreased occipital activity was observed for novel categorical stimuli compared with noncategorical stimuli. In contrast, after intentional learning, novel categorical stimuli evoked increased activity in the hippocampus, right prefrontal cortex, left inferior temporal cortex, precuneus, and posterior cingulate. Even though the categorization test was identical in the two conditions, the differences in brain activity indicate differing representations of category knowledge depending on whether the category had been learned intentionally or implicitly. ■

INTRODUCTION

A fundamental dissociation among memory types is seen in the memory function of amnesic patients who have damage to the medial temporal lobe (MTL). These patients have impaired declarative memory for facts and events (conscious, explicit memory), but exhibit intact nondeclarative (nonconscious) memory (Squire & Knowlton, 2000). Intact nondeclarative memory in amnesic patients implies the existence of memory systems outside the MTL, but the study of amnesic patients does not tell us how these memory systems operate, what brain areas support them, or how they interact with intact declarative memory in healthy people.

When healthy participants are given tasks that depend on nondeclarative memory, they generally appear to be unaware that they have acquired new information or a new skill and participants usually attribute their improved performance to guessing, instinct, or gut feeling. The term “implicit learning” was first used to introduce this idea of learning without awareness (Reber, 1967) and inspired a great deal of research and debate about whether memory could truly occur outside awareness. The character of the debate over awareness was similar in several related research areas. After an initial report in which healthy participants exhibited learning while reporting that they were unaware of what they had learned (e.g., Nissen & Bullemer, 1987; Reber, 1967),

subsequent research suggested other methods of interrogating awareness that seemed to reveal some conscious knowledge (e.g., Perruchet & Amorim, 1992; Dulany, Carlson, & Dewey, 1985). Resolving this issue appeared to depend on proving “null awareness” of acquired knowledge in healthy participants, which is clearly a daunting proposition (Merikle, 1994).

The findings that amnesic patients exhibited normal learning on the tasks for which awareness of learning was being debated (e.g., Knowlton, Ramus, & Squire, 1992; Nissen & Bullemer, 1987) suggested that there is a separate type of memory in the brain that is independent of the MTL and does not automatically afford awareness of what has been learned. A detailed examination of awareness in patients and age-matched controls found that even when some evidence of awareness is seen in healthy participants, amnesic patients exhibit normal learning without awareness (Reber & Squire, 1994).

One criticism of the neuropsychological research is that it has relied largely on single dissociations (with some notable exceptions such as Knowlton, Mangels, & Squire, 1996). A finding of “normal” learning in a patient group is based on a lack of a statistical difference between groups and cannot rule out the possibility of a small impairment that is difficult to detect. Thus, differential test sensitivity could account for some reports in which amnesic patients exhibit unimpaired performance on tasks of nondeclarative memory although they are impaired on declarative memory (e.g., Shanks & St. John, 1994).

A related question is how to connect neuropsychological findings and the studies that demonstrate learning without awareness in healthy participants. Whereas the consistency between the tasks that produce both sets of results suggests a common basis, an alternate hypothesis is that the neurological damage that causes amnesia disrupts a memory system that acts as a more unified whole in healthy participants (Perruchet & Gallego, 1993). This hypothesis suggests that there is normally a single memory system that in healthy participants acquires information with awareness and fragments with neurological damage to produce the phenomenon of apparently isolated nondeclarative memory.

These concerns are addressed here by using fMRI to examine the expression of category knowledge after it has been acquired either in a manner that leads to an implicit, nonconscious representation or in a manner that supports conscious, explicit knowledge. It is expected that implicit learning depends on nondeclarative memory systems whereas explicit knowledge is supported by the MTL memory system. Learning mode was manipulated in a visual categorization task via task instructions. One group of participants learned the category incidentally, using a procedure that leads to nondeclarative memory in amnesic patients. A second group of participants learned the same information intentionally and were expected to develop an explicit, conscious representation of the category.

The visual categorization task is based on learning a category of dot patterns by viewing a set of patterns distributed around a central prototype (Posner & Keele, 1968). Knowlton and Squire (1993) reported that this task is learned at a normal rate by amnesic patients. The task is typically given with incidental learning instructions, and healthy participants often spontaneously report (complain) of a lack of awareness of the category knowledge when given a test of category membership with new stimuli. However, the structure of the category is not necessarily hard to identify if one is attempting to do so. Thus, this task lends itself to simple manipulations that influence the form of learning via specialized instructions during study. One group of participants was told explicitly to try to identify the category when the study items were presented. The other group was shown the patterns without mention of the category and acquired the category incidentally using a standard method from previous work (Reber, Stark, & Squire, 1998a, 1998b; Knowlton & Squire, 1993). After learning, all participants were given an identical categorization test while fMRI data were collected to identify areas exhibiting differential activity for categorical and noncategorical stimuli. If different types of representation are acquired based on the manner of learning, differential activity patterns should be different for categorical and noncategorical stimuli for the intentional and incidental learning groups. If there is one underlying

memory system for category knowledge, the activity pattern should be the same for both groups.

Being able to manipulate the type of learning and compare activity during identical tests addresses the concern about differential test sensitivity for implicit and explicit memory. The same categorization test that assesses nondeclarative memory after implicit learning becomes an assessment of declarative memory when the category has been learned intentionally. Thus, unlike the comparison between recognition and nondeclarative categorization, differences in brain activity (or the performance of amnesic patients) cannot be attributed to differences in the tests. To achieve this situation, it is necessary to test healthy participants who have both declarative and nondeclarative memory systems intact using tasks in which the instructions can be manipulated so that the same information is acquired in either a conscious or nonconscious mode. This also enables a direct comparison of conscious and nonconscious memory in healthy participants. If there are separate and somewhat independent memory systems for the two types of memory, this should be reflected in a qualitatively different pattern of brain activity during categorization after conscious or nonconscious learning.

In two previous reports, contrasts between implicit and explicit learning mode have been examined with fMRI and provided evidence that separate systems are involved in implicit and explicit learning. Aizenstein et al. (2000) examined category learning with a variant of the dot-pattern categorization task in which stimuli were colored so that the color was predicted by whether the stimulus was a category member. Category learning is measured by reduced reaction time to identify the color for categorical stimuli compared with nonmembers. When participants were not informed of the color-category association, categorical patterns were found to evoke less activity than noncategorical patterns in a posterior, superior occipital region. Later, participants performed the task under explicit instructions to identify the color-category association, and the deactivation was not observed. On each trial, participants viewed a dot pattern and received feedback (the color of the pattern), suggesting that the imaging differences could have arisen from the difference in learning mode via either the memory-system-supported learning (nondeclarative for implicit learning, declarative for intentional learning) or the learning mode (specifically the attempt to learn explicitly in that condition). Poldrack et al. (2001) also observed different patterns of brain activity depending on the learning mode used in a probabilistic classification task. When participants learned using a procedure that leads to intact learning in amnesic patients, increased activation was observed in the basal ganglia (consistent with neuropsychological reports), but when participants learned a similarly structured paired-associates task explicitly, increased activity was observed in the MTL.

Both of these reports suggest that implicit or explicit (intentional) learning mode can have a large impact on the brain areas exhibiting activity during a task. A key question for the multiple memory systems model is whether these modes lead to different representations of category knowledge. The representation of knowledge is best assessed during the categorization test, after learning has occurred, when the expression of acquired memory can be observed. One possibility is that although the learning mode changes when participants are told to explicitly deduce the category structure, the representation of the category that occurs as the end product is similar in both cases (as suggested in Nosofsky & Zaki, 1998). In the current study, this possibility is addressed and the effects of learning mode on brain activity are reduced by imaging the categorization test for both groups under identical instructions.

As in previous reports (Reber et al., 1998a, 1998b), participants learned the category during a study phase before scanning by viewing a set of five dot-patterns taken from the category (see Figure 1). One group (IC, $n = 12$) learned the category in the traditional incidental fashion by simply observing the patterns under instructions to identify the central dot in each pattern. For this group, no mention was made of the existence of the underlying category structure. A separate group of participants (EC, $n = 8$) learned the identical category by viewing the same study items after being told that the study items came from the same

category and being instructed to attempt to determine the category.

In the subsequent fMRI phase, participants viewed patterns from the target category and from an unfamiliar category, and were asked to try to indicate whether it came from the same category as the study items for each pattern. The IC group was surprised when told about the category and many participants spontaneously reported (complained) that they were unaware of the category and would therefore have to guess at membership. During the test, the dot patterns were grouped into blocks of mostly categorical or mostly noncategorical patterns (9 patterns per block in a 7:2 ratio; noncategorical patterns belonged to an unfamiliar category). Based on this stimulus grouping, brain activity was compared during categorization judgments for categorical and noncategorical patterns. In previous reports (Reber et al., 1998a, 1998b), this contrast identified reduced activity for the categorical patterns suggesting that category learning produces a change in processing of these stimuli. The reduction in activity for categorical patterns suggests that these patterns are processed more fluently (i.e., requiring less effortful processing) in a manner reminiscent of fluency effects in priming (Schacter & Buckner, 1998). In this case, the fluency effect is generalized to novel members of a learned category and can be termed a “categorical fluency effect” (CFE). The CFE is a learning-based change in stimulus-correlated activity may be the basis of nondeclarative visual category memory. If the CFE is specific for nondeclarative category knowledge, it should be associated with an implicitly learned (IC) category whereas other stimulus-correlated effects should be associated with explicitly learned (EC) category knowledge.

RESULTS

Both groups exhibited knowledge of the category during the categorization test by correctly endorsing novel category members more often than the noncategorical foils. The noncategorical patterns were distributed around an unfamiliar prototype with the same relative structure as the target category, making it impossible to succeed on the test without being influenced by the previously studied patterns. The IC group averaged 59.4% correct ($SE\ 4.1$) and the EC group averaged 68.0% ($SE\ 3.8$) correct, both better than chance (50%): $t(11) = 2.32, p < .05$ and $t(7) = 4.70, p < .01$ for the IC and EC groups, respectively. There was a trend for the EC participants to perform better than the IC participants, $t(18) = 1.44, p < .10$.

For the IC group, reduced activity was observed in the posterior occipital cortex for the categorical patterns compared with the noncategorical patterns (coordinates: $x = +11, y = -91, z = +18$; right middle occipital gyrus, BA 19; see Figure 2). This effect replicates

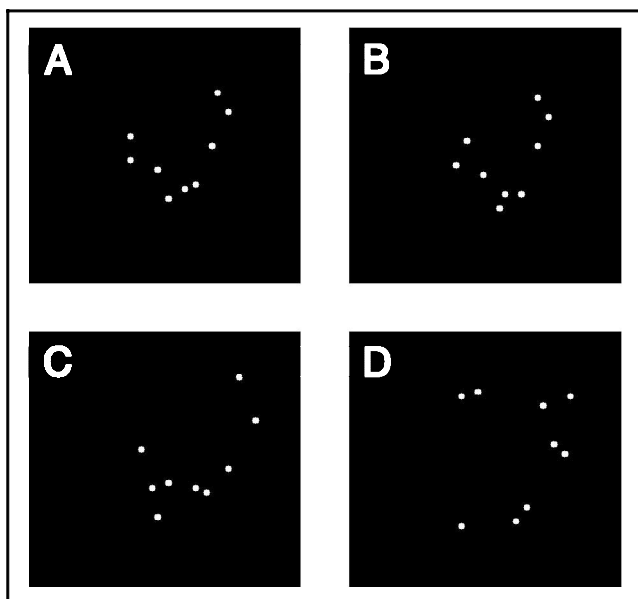
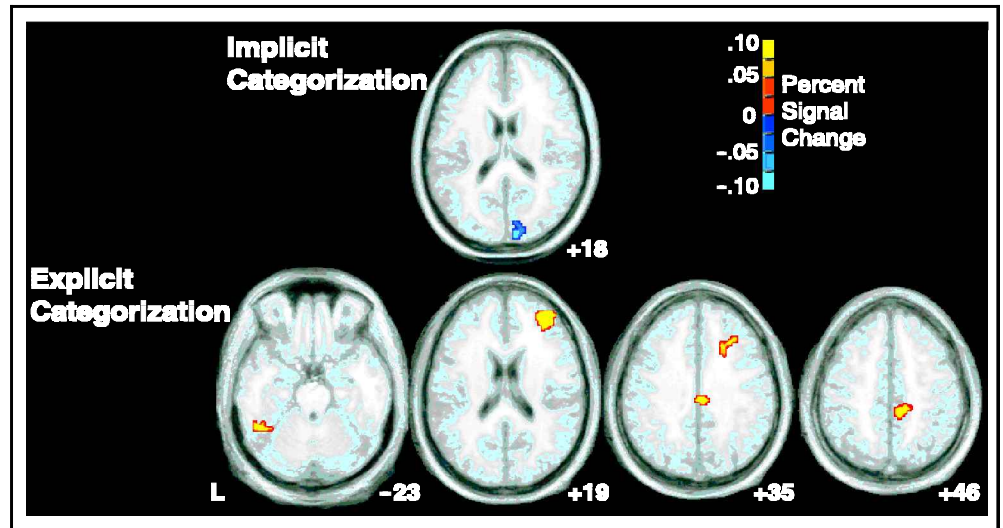


Figure 1. Dot pattern categorization stimuli. (A) Prototype of category to be learned. (B) “Low-distortion” categorical pattern. (C) “High-distortion” categorical pattern. (D) Prototype of unfamiliar category (the basis for the noncategorical patterns used at test). During the study phase, participants saw five high distortion patterns. During the test, participants saw 36 categorical patterns (the prototype four times, 16 low distortions and 16 high distortions) and 36 noncategorical patterns (distributed around the novel prototype, D).

Figure 2. Activity evoked by category members during categorization. Cool colors (blue) indicate areas in which category members evoked less activity than nonmembers. Warm colors (red–orange–yellow) indicate areas in which category members evoked greater activity than nonmembers. (Top) When the category was learned implicitly, decreased occipital activity was observed for categorical patterns—the categorical fluency effect (CFE). (Bottom) After explicit acquisition of the category, increased activity was observed for categorical patterns in the left occipito-temporal cortex, right anterior prefrontal cortex, and medial parietal cortex. All images are oriented according to the radiological convention with the right side of the brain on the left side of the image.



previous reports (Reber et al., 1998a, 1998b) that a CFE occurs after learning the visual category. The CFE reflects a change in visual processing of dot patterns such that the categorical patterns are processed more fluently and thus evoke less activity than noncategorical patterns. Here this finding is extended to the case of a short (5-item) study phase (previous studies used a 40-item study phase).

In the EC group, the same stimulus-based contrast identified five regions in which increased activity was observed for the categorical patterns (Figure 2, Bottom, and Table 1). Although the pattern of increased activity for the categorical patterns is qualitatively similar to the increases in activity in prefrontal and posterior cortical areas observed during recognition of previously presented dot patterns (Reber et al., 1998b), the specific areas of increased activity differ, suggesting that conscious categorization is not driven solely by recognition memory. Activity in two regions, the right anterior prefrontal cortex and the precuneus, has been reported

in conjunction with explicit memory retrieval (Fletcher, Buchel, Josephs, Friston, & Dolan, 1999; Krause et al., 1999; Buckner, Koutstaal, Schacter, Wagner, & Rosen, 1998; Rugg, Fletcher, Frith, Frackowiak, & Dolan, 1996; Rugg et al. 1998). However, activity in the right anterior prefrontal cortex has also been suggested to be sensitive to the testing context (Buckner, Koutstaal, Schacter, Dale, et al., 1998) and may be part of more general conscious memory processes. Conscious categorization also led to increased activity in the left inferior occipito-temporal cortex, which may reflect abstract visual form processing in the left hemisphere (Marsolek, 1995). The anatomical substrate of conscious visual categorization thus seems to rely on a network of areas involved in both explicit memory retrieval and the categorization of visual information.

The two conditions produced qualitatively different patterns of stimulus-correlated activity in these separate analyses. Categorization after intentional learning resulted in a network of increased activity for categorical

Table 1. Increases in Activity for Categorical Patterns After the Category was Learned Consciously

<i>Brain Region</i>	<i>Brodman's Area (BA)</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>Volume (mm³)</i>
Right anterior prefrontal cortex	10	35	45	19	2141
Left inferior occipito-temporal cortex (and cerebellum)	37	-43	-52	-25	1406
Precuneus	7	12	-34	46	906
Posterior cingulate	31	4	-27	34	641
Right dorsolateral prefrontal cortex	9	26	23	34	625

patterns without evidence for decreases. Categorization following incidental learning resulted in decreased activity for categorical patterns in the posterior occipital cortex without evidence for increases. However, the voxel-wise comparison between groups identified no significant clusters of differential activity in the two conditions by traditional statistical thresholds. A targeted direct comparison between the two types of memory was done by examining brain activity in those areas that are most likely to be fundamentally involved in supporting these types of memory. A fundamental characteristic of declarative memory is its dependence on the hippocampus, whereas reduced activity for categorical patterns in the occipital cortex appears to be a reliable marker of implicit memory (the effect observed here is within a few millimeters of the focus of a previously reported category-related deactivation [Reber et al., 1998b]; coordinates $x = +12, y = -93, z = +17$). Thus, the hippocampus and this posterior occipital region were selected for comparison of stimulus-correlated effects for the EC and IC groups.

The differences in activity for categorical and noncategorical patterns for hippocampal and occipital regions of interest (ROIs) are shown in Figure 3. The size of the CFE (reduced activity for categorical patterns) was assessed in each participant as the magnitude of the reduction in activity for categorical patterns that occurred in the posterior occipital ROI that was identified in the IC group analysis. Because this region is so similar in location to the previously reported site of the categorical fluency effect, this ROI should serve as an

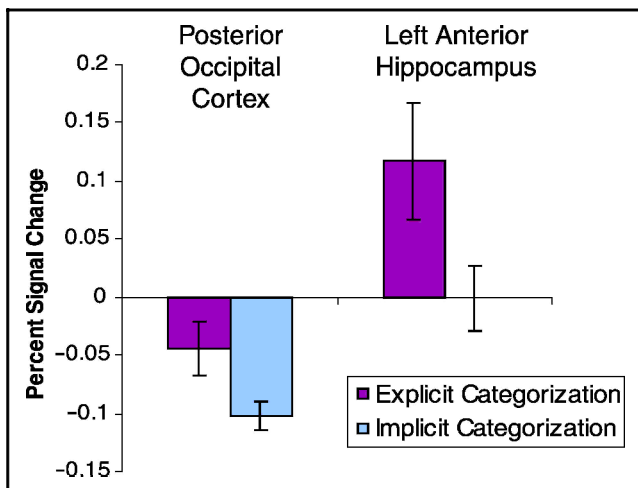


Figure 3. Differences in activity for categorical and noncategorical patterns. Decreased activity was observed in the posterior occipital cortex (left bars) in a functionally defined region of interest (ROI). This CFE was significantly larger when the category was acquired nonconsciously than when the category was acquired consciously, $t(18) = 2.36, p < .05$. Increased activity was observed in the left anterior hippocampus (right bars) in an anatomically defined ROI when the category was acquired consciously. This increase did not occur when the category was acquired nonconsciously. The difference is reliable, $t(18) = 2.22, p < .05$.

effective area for identifying the strength of the fluency effect in each participant for both groups. Activity in this region in the EC group exhibited a trend towards fluency, $t(7) = 1.94, p < .10$, but the fluency effect was significantly smaller than the effect observed in the IC group, $t(18) = 2.16, p < .05$, reflecting a smaller contribution of categorical stimulus fluency to conscious categorization.

To compare activity in the hippocampus, ROIs were defined based on anatomical landmarks for the anterior and posterior hippocampus bilaterally. This type of ROI analysis can detect changes in activity even when standard whole-brain-based analysis did not identify a cluster of increased activity in the MTL (Reber, Buxton, & Wong, 2002) by virtue of the benefit of a specific hypothesis and the increased sensitivity of aggregating data across a larger group of functionally related voxels. By this analysis, greater increased activity was observed in the left anterior hippocampus for the EC group, $t(18) = 2.22, p < .05$, reflecting the role of declarative memory in categorization for this group. No evidence for increased hippocampal activity was observed in the IC group, consistent with previous findings that damage to the MTL region does not impair this type of learning (Knowlton & Squire, 1993). The localization of increased activity for categorical stimuli to the left anterior hippocampus within the MTL suggests that participants who learned the category consciously are engaged in conscious retrieval, possibly verbally mediated, of category knowledge when viewing categorical stimuli during the test. This process appears to be absent (or at least to play a significantly smaller role) when the category is learned outside awareness.

In addition, the regions exhibiting increased activity for categorical patterns in the EC group were examined in the IC group for possible subthreshold activity. None of the five functionally defined regions exhibited differential activity for categorical or noncategorical patterns relative to the other, $ts(11) < -1.84$.

DISCUSSION

Brain activity during categorization indicates that two separate sets of processes occur depending on whether the knowledge of the category was acquired incidentally or intentionally. The CFE is associated with implicit learning of category knowledge, whereas intentionally acquired, explicit category knowledge evoked increased activity in the hippocampus and a network of cortical areas. Of particular importance in this contrast is the fact that the informational content is identical for both groups, that is, both groups have acquired knowledge of the same category based on the identical study items. Previously, categorization has been shown to evoke a different pattern of neural activity than recognition of similar stimuli (Reber et al., 1998b). However, the different task demands of categorization and recognition

allowed for the possibility that other factors than the type of memory influenced the pattern of activity. Here, the only difference between the two groups was the instructions to either consciously extract the category from the study patterns or simply observe the patterns (and the category is extracted incidentally). The different patterns of brain activity that are evoked by categorical and noncategorical stimuli on an identical test indicate that different representations of the category knowledge were acquired in the two conditions.

The CFE results in reduced activity for categorical stimuli after the category has been learned. The posterior occipital region in which this effect is observed is consistent with previous reports (Reber et al., 1998a, 1998b). The magnitude of observed fluency was greater for participants who learned the category implicitly, namely, under incidental conditions (IC group). Some evidence for a weaker fluency effect was observed in the conscious categorization (EC) group, but to a lesser degree. Instead, the participants in the EC group exhibited increased activity in the left anterior hippocampus, suggesting that this region supports explicit retrieval of category knowledge used during the categorization test. A limitation of the comparison of fluency effects is that the region used to investigate fluency was defined by the voxel-based analysis from the IC group, possibly reducing the sensitivity for observing fluency in the EC group. We cannot conclude that the two types of memory are exclusive (as suggested for probabilistic classification by Poldrack et al., 2001). However, the CFE appears to be an effective marker of implicit category knowledge both here and in previous reports (Reber et al., 1998a, 1998b).

Aizenstein et al. (2000) reported a qualitatively similar pattern of decreased cortical activity for categorical patterns during implicit learning and increased cortical activity during explicit learning, although the evoked activity in these regions was not directly compared across conditions. In that report, however, participants were receiving feedback during task performance and difference in learning mode might be part of the activation pattern differences. Feedback can play a significant role in learning mode on categorization tasks (Ashby, Maddox, & Bohil, 2002). The brain area exhibiting fluency in Aizenstein et al. is also several centimeters away from the CFE observed here (and previously). That area is much more superior ($x = +9, y = -71, z = +41$) and appears to be in areas of higher level dorsal visual processing than the area implicated here and previously. The different location of the fluency effect could arise from the differences in the task administration or possibly due to the comparison of a structured category and random patterns in Aizenstein et al. rather than the two similarly structured categories (novel and learned) used here.

In the current experiment, the differential activity evoked for categorical and noncategorical stimuli was

identified in a comparison of blocks of stimuli that alternated between patterns from the learned category and blocks of patterns from a similarly structured novel category. The only difference between the blocks was whether the patterns came from a category that had been seen during the study phase. Learning of the category during the test (as in Palmeri and Flanery, 1999) should lead to an attenuation of the differences between the learned and novel categories as the novel category is learned. The observed stimulus-correlated activity suggests that the novel category was not learned during the test. This might be surprising, given that the category was learned during the study phase after only five exemplars presented twice each, and suggests the possibility that the categorization test somehow interfered with implicit learning of the novel category. If the novel category had been learned implicitly over the 36 exemplars of that category presented during the categorization test, then there might have been little or no fluency differences observed during the test.

The use of a block design was motivated by the fact that the CFE reflects a subtle difference between activity evoked by categorical and noncategorical patterns. It was felt that a block design would provide increased sensitivity for observing this small difference. However, the use of a block design could raise the concern that participants anticipated the fact that the stimuli were being presented in blocks and adjusted their strategy depending on whether they expected a block of categorical or noncategorical stimuli. Although we cannot rule out this possibility entirely, there are three reasons we expect that any expectancy effect would be minimal: (1) The participants were not told that the stimuli were blocked and only completed a single run of four blocks, leaving very little time to learn to anticipate the blocking structure. (2) The blocking ratio only included 77% of the targeted trial type in each block (even balance would be 50%). (3) Participants achieved performance rates only in the 60–70% range, further obscuring their ability to identify the stimulus blocking parameters. We anticipate that future work using an event-related design, instead of stimulus-type blocking, will replicate the CFE and also provide the opportunity to examine the time course of evoked activity for the categorical and noncategorical patterns to better understand the nature of the relative deactivation of the categorical patterns.

The reduction in activity associated with category members after implicit learning of the category is qualitatively similar to reductions in activity seen in studies of repetition priming (Schacter & Buckner, 1998) that have been linked to the phenomenon of repetition suppression (Wiggs & Martin, 1998). However, the CFE is found to occur for novel members of a learned category, suggesting that if repetition suppression is involved, this effect generalizes to novel stimuli that are perceptually

similar to previously studied items. In addition, the CFE is observed in the posterior occipital cortex, rather than the ventral temporal cortical areas (BA 37) in which repetition priming effects are most commonly observed. The differential location may be because the dot pattern stimuli are perceptually simple patterns of white dots on a black background, suggesting that the critical perceptual processing occurs earlier in the visual processing hierarchy. Linking the CFE to repetition suppression implies that repetition suppression effects can generalize across perceptually similar stimuli and may reflect a general mechanism operating throughout the visual stream and not just in the object processing areas of the ventral temporal cortex.

An important aspect of the contrast between the patterns of activity observed for the IC and EC groups is that when the category was learned explicitly (EC group), the preponderance of activity changes reflected increases for the category members. When the categorization decisions were made based on an intentionally acquired representation of the category, evaluation of category members evoked increased activity in prefrontal, ventral temporal and posterior medial parietal cortex. Donaldson, Petersen, and Buckner (2001a) reported a similar contrast between repetition–suppression-like effects for priming and increased activity for “old” stimuli on a recognition test. The posterior medial parietal increases in activity for categorical stimuli bears some resemblance to increases in activity in the precuneus and posterior cingulate have been frequently observed in other studies of old–new recognition using verbal stimuli (e.g., Donaldson et al., 2001a, 2001b; Henson, Rugg, Shallice, Joesechs, & Dolan, 1999; Henson, Rugg, Shallice, & Dolan, 2000). In recognition memory, the area of increased activity in the precuneus has typically been more posterior than observed here ($y = -66$ to -73 compared with $y = -34$) but may reflect some common component processes between old/new recognition and categorization supported by explicit knowledge. It is expected that declarative memory supports both recognition memory and categorization after explicit learning, but it should be noted that the pattern of increased activity observed here during conscious categorization is not identical to that observed during target recognition (Reber et al., 1998b). Areas of common increased activation for recognition and explicit categorization of dot patterns included right anterior and right dorsolateral prefrontal cortex, suggesting that this region may play a general role in conscious memory retrieval. During recognition, however, additional increases were observed in the right ventral occipito-temporal cortex, bilateral lateral temporal cortex, and left anterior prefrontal cortex. During conscious categorization, additional increases were observed in the left ventral temporal cortex and medial parietal areas. While there may be some commonality in mechanisms (right prefrontal cortex, MTL), there are

likely to also be important differences between recognition and categorization that reflect differences in decision criteria and strategies for these two tasks (as suggested in Nosofsky & Zaki, 1998).

The incidental learning procedure of the IC group led to a representation that has been shown to be learned by amnesic patients at a normal rate (Knowlton & Squire, 1993). The incidental learning procedure is thus supported by nondeclarative memory and is theoretically not available to the participants’ conscious awareness. It was not possible with this paradigm to experimentally confirm participants’ awareness or lack of awareness of the category structure as there is no established way of assuring a full verbal report of knowledge of abstract nonverbal stimuli. However, if the instructional manipulation was ineffective, no difference in brain activity during categorization would have been observed. Because the study items and test items are identical, any observed differences are solely attributable to the difference in study instructions given the two groups of participants. Identification of a dissociation in the brain activity between the two groups suggests that they have acquired different types of knowledge and provides evidence that healthy participants have separate brain systems supporting these representations.

The activity observed in the hippocampus during explicit categorization suggests that amnesic patients should exhibit impairments in this type of learning. A recent report of verbal category learning by amnesic patients (Kitchener & Squire, 2000) found that patients were impaired and suggested that this impairment was due to the task affording a declarative, conscious category learning strategy. The consistency of the neuroimaging results with the studies of patients with MTL damage reinforces the connection between awareness and the brain systems supporting declarative and nondeclarative memory.

Dissociations between nondeclarative and declarative memory have been shown numerous times in neuropsychological research (Squire & Knowlton, 2000). Connecting these findings with parallel research examining implicit and explicit learning in healthy participants has been famously difficult (Shanks & St. John, 1994), largely due to the problem of assessing a true lack of awareness of the material to be learned (Merikle, 1994). Here, rather than relying on a verbal report of subjective experience, functional neuroimaging was used to objectively assess differences in brain function depending on whether knowledge was acquired intentionally (explicit) or incidentally (implicitly). When implicit category knowledge was used, evidence was observed for fluent processing of categorical patterns. The expression of explicit category knowledge evoked increased activity in the hippocampus, the brain structure most closely associated with the acquisition and retrieval of conscious knowledge.

METHODS

Subjects

Twenty three participants (12 men, 11 women) were recruited from the Northwestern University community and compensated for their participation. All participants gave informed consent according to procedures approved by the Northwestern University Institutional Review Board. Three participants (of 11) in the EC group were eliminated due to data contamination from artifacts due to motion and/or scanner noise (all three volunteers had been tested during the same scanning day).

Procedure

For the initial learning phase, participants were placed in the MRI scanner and viewed images presented on a rear-projection screen via a mirror mounted above their eyes. For the incidental categorization (IC) group ($n = 12$), participants were shown five dot patterns and asked to identify the center dot in each pattern (no overt response was made). The patterns were shown twice each (10 total). Participants were not told that the patterns were derived from an underlying category defined by a central prototype. The explicit categorization (EC) group ($n = 8$) were shown the same patterns, but were instructed that the patterns all came from the same category and that they should attempt to learn the category while watching the patterns.

After the learning phase, there was a 2-min delay (while localizing scans were performed). All participants were then instructed that the patterns previously seen had been drawn from a category and that they would now be shown new patterns, some of which came from the category and some did not. For each pattern they were instructed to indicate “yes” or “no” (via buttons) whether they thought the pattern came from the category. They were asked to make a response for each pattern even if they had to guess.

The categorization test consisted of 72 novel dot patterns, 36 categorical, and 36 patterns drawn from an unfamiliar category (noncategorical patterns in the sense that they were not drawn from the learned category). Stimuli were grouped into eight blocks of nine patterns each with each block either containing mostly categorical or mostly noncategorical patterns in a 7:2 ratio.

Imaging

Whole-brain images were collected (22 6-mm axial slices) every 3.6 sec (gradient-recalled EPI, TR = 3600 msec, TE = 40 msec, flip angle = 90°, FOV = 240, voxel size = 3.75 mm × 3.75 mm × 6 mm), for 84 repetitions (four initial scans eliminated) over 5 min and 2 sec.

Data Analysis

Using a combination of software within and based on AFNI (Cox, 1996), functional images were coregistered through time (motion correction), spatially smoothed (FWHM = 7.5 mm), and voxels with excessive sudden signal change (>10% in 3.6 sec) were eliminated from analysis. Within each participant, voxels were fit to a model function based on the blocking of stimuli to identify voxels in which activity changed as a result of processing categorical or noncategorical patterns. Hemodynamic delay was fit within a range of 4–8 sec to maximize response on an individual voxel basis. After normalization to the standardized stereotaxic atlas (Collins, Neelin, Peters, & Evans, 1994) (transformed resolution = 2.5 mm³), data were combined across participants with a random effects model to identify voxels in which stimulus-correlated changes in activity were consistent within each group (after removal of linear drift and residual motion correlated activity). Regions of stimulus-correlated change in activity were clusters of at least 500 mm³ in which each voxel was consistently active across participants: $t(11) > 4.1$, $p < .002$ uncorrected for the IC group; $t(7) > 4.5$, $p < .003$ for the EC group; with participants treated as a random effect. The cluster and t statistic thresholds were identified by Monte Carlo simulation as eliminating false positives ($p < .05$ corrected, for a single cluster across the brain at this threshold) in matched noise data. The cluster size of 500 mm³ identifies areas composed of at least 32 voxels (normalized, which would be ~6 in the original EPI if the images maintained size during the normalization transformation).

Signal change within a specific ROI was examined by aggregating the time series data (BOLD signal) from all voxels within the functionally defined (posterior region exhibiting fluency) or anatomically defined (anterior and posterior bilateral hippocampus) areas. Anatomical boundaries for the hippocampal regions that are visible on structural MRI are described elsewhere (Reber et al., 2002; Insausti et al. 1998). The resulting time courses were then examined with correlation analysis to produce an estimate of signal change for each ROI for each participant. The mean signal change within each ROI was compared to chance (no change) within each group and mean signal changes were additional contrasted across groups.

Acknowledgments

The authors thank Ken Paller, Dan Kimberg, and Craig Stark for helpful comments on this report and Sara Polis for assistance with manuscript preparation. This research was supported by a grant from the National Institute of Mental Health, R01 MH58748.

Reprint request should be sent to Paul J. Reber, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60201, USA, or via e-mail: preber@northwestern.edu.

The data reported in this experiment have been deposited in The fMRI Data Center (<http://www.fmridc.org>). The accession number is 2-2002-113D1.

REFERENCES

- Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., & Carter, C. S. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience*, *12*, 977–987.
- Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory and Cognition*, *30*, 666–667.
- Buckner, R. L., Koutstaal, W., Schacter, D. L., Dale, A. M., Rotte, M., & Rosen, B. R. (1998). Functional-anatomic study of episodic retrieval using fMRI: II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. *Neuroimage*, *7*, 163–175.
- Buckner, R. L., Koutstaal, W., Schacter, D. L., Wagner, A. D., & Rosen, B. R. (1998). Functional-anatomic study of episodic retrieval using fMRI: I. Retrieval effort vs. retrieval success. *Neuroimage*, *7*, 151–162.
- Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D inter-subject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, *18*, 192–205.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- Donaldson, D. I., Petersen, S. E., & Buckner, R. L. (2001a). Dissociating memory retrieval processes using fMRI: Evidence that priming does not support recognition memory. *Neuron*, *31*, 1047–1059.
- Donaldson, D. I., Petersen, S. E., Ollinger, J. M., & Buckner, R. L. (2001b). Dissociating state and item components of recognition memory using fMRI. *Neuroimage*, *13*, 129–142.
- Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1985). On consciousness in syntactic learning and judgment: A reply to Reber, Allen, and Regan. *Journal of Experimental Psychology: General*, *114*, 25–32.
- Fletcher, P. C., Buchel, C., Josephs, O., Friston, K., & Dolan, R. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex*, *9*, 168–178.
- Henson, R. N. A., Rugg, M. D., Shallice, T., & Dolan, R. J. (2000). Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. *Journal of Cognitive Neuroscience*, *12*, 913–923.
- Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, *15*, 3962–3972.
- Insausti, R., Juottonen, K., Soininen, H., Insausti, A. M., Partanen, K., Vainio, P., Laakso, M. P., & Pitkanen, A. (1998). MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *AJNR. American Journal of Neuroradiology*, *19*, 659–671.
- Kitchener, E. G., & Squire, L. R. (2000). Impaired verbal category learning in amnesia. *Behavioral Neuroscience*, *114*, 907–911.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 245–254.
- Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*, 172–179.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.
- Krause, B. J., Schmidt, D., Mottaghy, F. M., Taylor, J., Halsband, U., Herzog, H., Tellmann, L., Müller-Gärtner, H. W. (1999). Episodic retrieval activates the precuneus irrespective of the imagery content of work pair associates: A PET study. *Brain*, *122*, 255–263.
- Marsolek, C. J. (1995). Abstract visual-form representations in the left cerebral hemisphere. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 375–386.
- Merikle, P. M. (1994). On the futility of attempting to demonstrate null awareness. *The Behavioral and Brain Sciences*, *17*, 412.
- Nissen, M. J., & Bullemer, P. T. (1987). Attentional requirements for learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.
- 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: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, *10*, 526–530.
- Perruchet, P., & Amorin, M. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 785–800.
- Perruchet, P., & Gallego, J. (1993). Association between conscious knowledge and performance in normal subjects: Reply to Cohen and Curran (1993) and Willingham, Greeley, and Bardone (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1438–1444.
- Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546–550.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863.
- Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning and Memory*, *1*, 217–229.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998a). Cortical areas supporting category learning identified using functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 747–750.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998b). Contrasting cortical activity associated with declarative and nondeclarative memory. *Learning and Memory*, *5*, 420–428.
- Reber, P. J., Wong, E. C., & Buxton, R. B. (2002). Encoding activity in the medial temporal lobe examined with anatomically constrained fMRI analysis. *Hippocampus*, *12*, 363–376.
- Rugg, M. D., Fletcher, P. C., Allan, K., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1998). Neural correlates of memory

- retrieval during recognition memory and cued recall. *Neuroimage*, *8*, 262–273.
- Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, *119*, 2073–2083.
- Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*, 185–195.
- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 376–447.
- Squire, L. R., & Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 765–776). Cambridge: MIT Press.
- Wiggs, C., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *9*, 227–233.