Published in final edited form as:
Neurosci Biobehav Rev. 2008 ; 32(2): 249–264. doi:10.1016/j.neubiorev.2007.07.009.

Multiple Systems of Category Learning

Edward E. Smith[*] and
Columbia University

Murray Grossman
University of Pennsylvania

Abstract

We review neuropsychological and neuroimaging evidence for the existence of three qualitatively different categorization systems. These categorization systems are themselves based on three distinct memory systems: working memory (WM), explicit long-term memory (explicit LTM), and implicit long-term memory (implicit LTM). We first contrast categorization based on WM with that based on explicit LTM, where the former typically involves applying rules to a test item and the latter involves determining the similarity between stored exemplars or prototypes and a test item. Neuroimaging studies show differences between brain activity in normal participants as a function of whether they are instructed to categorize novel test items by rule or by similarity to known category members. Rule instructions typically lead to more activation in frontal or parietal areas, associated with WM and selective attention, whereas similarity instructions may activate parietal areas associated with the integration of perceptual features. Studies with neurological patients in the same paradigms provide converging evidence, e.g., patients with Alzheimer's disease, who have damage in prefrontal regions, are more impaired with rule than similarity instructions. Our second contrast is between categorization based on explicit LTM with that based on implicit LTM. Neuropsychological studies with patients with medial-temporal lobe show that patients are impaired on tasks requiring explicit LTM, but perform relatively normally on an implicit categorization task. Neuroimaging studies provide converging evidence: Whereas explicit categorization is mediated by activation in numerous frontal and parietal areas, implicit categorization is mediated by a deactivation in posterior cortex.

INTRODUCTION

Over a decade ago, researchers began to make a case for the existence of qualitatively different categorization systems; the arguments were primarily confined to behavioral evidence and were insulated from other issues in human learning and memory (e.g., Nosofsky, Palmeri, & McKinley, 1994; Smith, Patalano, & Jonides, 1998). In the past decade, the case for multiple categorization systems has increasingly relied on data from neuropsychology and

[*] Acknowledgment
Preparation of this paper, as well as execution of many of the studies reviewed in it, was supported in part by the U.S. Public Health Service (AG17586, AG15116). The authors are indebted to Mark Gluck and three anonymous reviewers for their comments on an earlier version of the manuscript.

Corresponding author: Edward E. Smith, Columbia University, Department of Psychology, 1190 Amsterdam Ave., MC 5501, New York, NY 10027, Email: eesmith@psych.columbia.edu Phone: (212) 854-1789.

neuroimaging, and has frequently been tied to the broader issue of multiple memory systems. The goal of this paper is to selectively review some of this recent evidence and the newer perspective that motivates it. Although our review emphasizes experiments from our own laboratory, other studies are discussed as well.

Three Kinds of Memory Systems

A useful starting point is the current consensus that there are three distinct learning-and-memory systems: working memory, explicit long-term memory, and implicit long-term memory. *Working memory (WM)* is typically conceptualized as a system for the active maintenance and manipulation of a limited amount of information (e.g., roughly, 4 items) for a limited amount of time (measured in seconds) (see, e.g., Baddeley, 1986; Jonides, 1995). The information in WM is assumed to be represented *explicitly*, as witnessed by the fact that its contents can be described or reported on. The system is further assumed to be mediated neurally, at least in part, by areas in the prefrontal cortex and the posterior parietal cortex (e.g., Smith & Jonides, 1999).

In contrast, *explicit long-term memory* (explicit *LTM*) involves the passive storage of substantial amounts of information on a long-term basis (measured in hours to years). Encoding information into this system is known to be mediated neurally by regions in the medial temporal lobe (e.g., see Squire & Knowlton, 2000 and 2005 for recent reviews), whereas retrieval of information from explicit LTM involves regions in frontal cortex as well (some of which are distinct from those mediating WM; see Wheeler & Buckner, 2003 and Wheeler, Peterson, & Buckner, 2000 for recent reviews). Again the representations are explicit, or reportable.

In sharp contrast to both of these explicit systems is *implicit long-term memory* (implicit *LTM*). Its defining characteristic is that it is used (retrieved) without awareness, and thus the contents of an implicit memory can not be reported. Implicit LTM is essentially defined negatively: it is used with *no* awareness, and is *not* mediated by the prefrontal cortex, the posterior parietal cortex, *nor* the medial temporal lobe (e.g., Schacter, 1987; Schacter & Buckner, 1998). This negative definition allows room for a number of different entities to qualify as implicit memory (i.e., there appear to be more than one kind of implicit memory).

We subscribe to the view that different category-learning and categorization systems are associated with each of these different memory systems (e.g., Ashby & Maddox, 2005). In this article, we will focus on three categorization systems, which can be characterized as follows:

WM, or rule-based categorization system—In principle, WM can support a variety of classification procedures, but it is typically identified with rule-based categorization. When using this procedure, one decides whether a test item belongs to a particular category by determining whether the item fits a rule that defines the category (the rule specifies the necessary and sufficient attribute-values for category membership). Following the analysis offered in Smith et al. (1998), the cognitive operations involved in rule-based categorization presumably include the following:

1. Selectively attending to each critical attribute of the test item (a “critical attribute” is one mentioned in the rule);
2. For each attended-to attribute, determining whether its value matches a value (or feature) specified in the rule;
3. Amalgamating the outcomes of Stage (2) in WM so as to determine the final categorization.

Critically, rule-based categorization requires selective attention and temporary storage (of the rule as well as outcomes of feature tests), and these are standard functions of WM.

Explicit LTM categorization system—Explicit LTM has been associated with two similarity-based categorization processes. In one, a person decides whether a test item belongs to a particular category by determining its similarity to remembered exemplars of the category, whereas in the other process the comparison is to a prototype of the category (Smith & Medin, 1981). It is convenient to treat these two kinds of processes together as recent models of category learning include both exemplars and prototypes as representations (e.g., Love, Medin, & Gureckis, 2004).

Implicit LTM system—As already noted, there may be numerous implicit memories, and hence multiple, implicit categorization systems. Here we focus on one kind of implicit memory, which has been extensively explored, namely the implicit system that underlies perceptual priming and related phenomena (e.g., Roediger & McDermott, 1993; Squire & Knowlton, 2000). In categorization based on this system, one decides whether a test item belongs to a particular category by assessing the ease with which its perceptual features can be processed, with greater "perceptual fluency" indicating a greater likelihood of category membership (e.g., Smith, 2007).

Some comments are in order about what we are leaving out. One proposal about categorization that has some currency in cognitive psychology and cognitive development is called "theory-based" categorization, in which one decides whether a test item belongs to a particular category by determining whether the features of the test item are best explained by the "theory" that underlies the category (see Murphy, 2002 for a review). Detailed, computational accounts of this approach are rare, but the representations involved are generally assumed to be explicit ones. To the best of our knowledge, this approach has rarely been investigated in neuropsychological and neuroimaging studies, which are the focus of this review, and we will have little further to say about this approach.

A potentially more significant omission is that of another implicit system, often referred to as "habit learning" (Knowlton, Mangels, & Squire, 1996; Shohamy, Myers, Grossman, Sage, Gluck, & Poldrack, 2004). To illustrate the system, consider a task often used to study it, the "weather-prediction task". A set of 1 to 3 different visual patterns is presented on each trial, and the participant has to decide in which of two categories, SUN or RAIN, each set belongs. Like many category learning paradigms, feedback is given during learning. But unlike most category learning paradigms, the identical stimulus can be assigned to different categories on different trials; i.e., categorization is probabilistic. Because feedback is given after each trial, one might assume that the learner will test hypotheses about the critical features of the categories, which would suggest that the WM system is recruited; but neuroimaging evidence argues against this assumption (Foerde, Knowlton, & Poldrack, 2006). Presumably, the probabilistic nature of the task defeats any early attempt at rule-based categorization, and an implicit system takes over.

The habit-learning system is important for our understanding of human learning, but it is not clear that what this system learns is best characterized as "categories". Since the empirical analysis of category learning began (Hull, 1920), one criterion for claiming that a category has been learned is that the learner be able to categorize novel items, even those in a different format than the training items. To our knowledge, this criterion has not been demonstrated with the weather-prediction task.

However, this criterion has been satisfied in a rather different paradigm that has been used to study habit learning, namely the "information integration" paradigm developed by Ashby, Maddox and their co-workers (see Ashby & Maddox, 2005, for a recent review). In this paradigm, the stimuli, or the assignments of stimuli to categories, are very difficult to verbalize, and this aspect presumably triggers habit learning. For example, in a number of studies, participants were presented visual sine-waves that varied in frequency (thickness) and orientation (felt); when categorization was based on integrating both attributes in a way that was difficult to verbalize, performance was affected by different factors than when categorization was based on a single, verbalizable attribute (e.g., Ashby & Maddox, 2005). Recent imaging work provides further evidence that the information-integration paradigm recruits a kind of implicit memory that differs from that involved in perceptual priming and perceptual fluency (Nomura et al., 2006).

The relation between the two kinds of implicit categorization is not at all clear. In view of this, and the fact that habit learning is treated in detail in other articles in this volume, we omit this system in the body of our review. At the close, however, we will return to the contrast between our proposal of multiple categorization systems and that of Ashby and Maddox (2005).

In what follows we consider issues about multiple categorization systems. We first deal with the contrast between the WM and explicit LTM systems (both explicit systems), and then focus on the contrast between explicit and implicit LTM. To foreshadow our conclusions, we will argue that the neuropsychological and neuroimaging evidence support the claim that there are three qualitatively distinct systems for categorization. Further, there is some evidence that these systems can operate concurrently.

A CONTRAST OF EXPLICIT SYSTEMS: WM VERSUS EXPLICIT LTM

The studies reviewed in this section are organized on the basis of whether they used artificial (novel) categories or natural ones. The use of novel categories allows the researcher to have control over the contents of the categories, but at the cost of not being able to ensure that structure of the category mirrors that of a natural concept. The use of natural categories has the reverse pattern of costs and benefits. For each kind of category, first we consider evidence from neuroimaging studies, and then results from behavioral studies with neurological patients. While this overview provides a roadmap for the bulk of the section, we begin with a behavioral study of normal participants that provided the inspiration for our initial neuroimaging experiment with novel categories.

Using Novel Categories: A Behavioral Study

Previous reviews of behavioral evidence have supported the hypothesis that rule-based categorization is qualitatively different than similarity-based categorization, as the former is based on WM and the latter on explicit LTM (e.g., Smith et al., 1998). Though the current review focuses on neural evidence, it is useful to start with a behavioral study that is informative about the issues of interest and that led to the neuroimaging studies of categorization.

The experiment of interest is by Allen and Brooks (1991), who were concerned with the difference between categorization based on rules versus similarity. The participants' task was to categorize artificial animals into two categories, referred to as BUILDERs and DIGGERS (we use caps to indicate categories). The animals were composed from five, binary attributes, and were cartoonish in appearance. There were two phases to the experiment: a training phase, during which participants learned to correctly categorize a set of ten animals, and a test phase, during which participants were tested on some novel animals as well as on ones that they had learned. In the training phase, one group of participants, the Rule group, was taught a rule to distinguish BUILDERs from DIGGERs, e.g., "If an animal has at least two of the following attribute values—long legs, angular body, spotted covering—it is a BUILDER; otherwise it is a DIGGER". A second group of participants, the Similarity group, was presented the same animals but not the rule. This group was instructed that the first time they saw an animal they would have to guess its category, but on subsequent trials they would be able to remember what it was. Thus, the Rule group was induced to use rule-based categorization, which is dependent on WM, whereas the Similarity group was induced to employ explicit LTM, and presumably use the similarity of novel items to memorized training items (exemplars) when making categorization decisions.

In addition to the difference in instructions, the major variation in this experiment concerned two types of novel items presented during the test phase, and it is convenient to illustrate them with respect to BUILDER. One kind of novel item was an instance of BUILDER according to the rule, and was also extremely similar to an old item that was a known exemplar of BUILDER. This kind of item is referred to as a “positive match”. The other kind of novel item, a “negative match”, was also a BUILDER according to the rule, but it was extremely similar to a known exemplar of DIGGER. If Rule participants do indeed categorize novel items by rule, their dominant categorization of both positive and negative matches should be the same: BUILDER. If Similarity participants categorize novel items by retrieving the stored exemplar most similar to it and selecting the category associated with that exemplar, they should categorize positive matches as BUILDERs and negative matches as DIGGER. Thus the Rule and Similarity groups should differ on their dominant categorization of negative matches.

This is exactly what happened. For negative matches, the dominant categorization in the Rule group (55%) was BUILDER, whereas the dominant categorization in the Similarity group (86%) was DIGGER. These results provide evidence for the existence of two distinct categorization systems, one that corresponds to rule application and is based on WM, and the other that corresponds to exemplar similarity and is based on explicit LTM. A further analysis of the data in this study provides evidence that the systems operate concurrently and interact. This analysis considers categorization only in the Rule group, and focuses on the contrast between positive and negative test items. If the Rule participants always applied their rule and never engaged an exemplar-similarity process—i.e., they were not using the two categorization systems concurrently-- they should have performed the same on positive and negative matches. But if the Rule participants sometimes used exemplar-similarity concurrently with rule application, their performance should have been poorer on negative than positive matches; this is because for negative matches, rule-application picks out one category and exemplar-similarity selects the other. The latter result obtained: The error rate (where an “error” means going against the rule) was about 20% for positive matches, and 45% for negative matches. Furthermore, reaction times for correct responses were longer for negative than positive matches.

The latter results imply that both categorization systems of interest were operating concurrently and interacting. There is a simple account of how they interacted. The fact that a correct response is slowed when the exemplar-similarity process points to a different category than that selected by rule-application suggests that both processes provide outputs to a response-selection process, and the latter process takes longer when it receives conflicting outputs than compatible ones.

Using Novel Categories: Neuroimaging Studies

One of the first attempts to use neuroimaging to study the contrast between the WM and explicit LTM categorization systems employed a variant of the above Allan and Brooks paradigm. Patalano, Smith, Jonides, & Koeppe (2001) required two groups of participants to learn a pair of contrasting categories of artificial animals, either by Rule or Similarity instructions. Then both the Rule and Similarity groups performed a test phase while their brains were being imaged by Positron Emission Tomography (PET). In an effort to increase the memorabilty of the animals, new sets of artificial animals were created that had more perceptually distinct features. The categories used were artificial animals that were cartoonish and that varied on ten perceptual attributes (e.g., tail shape), with the attributes being relatively analyzable or separable (e.g., Garner, 1976). Also, in an effort to make rule-based categorization more demanding (so it would produce a larger PET response), the rule was made more complex – it now required matches on at least 3 of 5 attributes.

The issue of interest was whether categorization based on a rule leads to a meaningfully different pattern of results than categorization based on explicit LTM. Ideally, one would like to find a double dissociation between the two groups, with each group producing a distinctive pattern of activations (or deactivations). The imaging results, however, provided only a single dissociation. Relative to a baseline condition (just looking at the animals), the Rule group activated the prefrontal cortex (Brodmann areas [BAs] 6 and 46), the posterior parietal cortex (BA 7), the occipital cortex (BAs 17, 18, and 19), and certain sub-cortical areas (thalamus and cerebellum); in contrast, relative to the same baseline, the Similarity group activated occipital cortex (BAs 17, 18, and 19), and a sub-cortical area (cerebellum). Although the areas activated in the Similarity group in visual association cortex (BAs 17, 18, and 19) may be important for an exemplar similarity process (Tracy et al., 2003), they appeared to be largely a subset of those activated in the Rule group.

In an effort to pinpoint the critical differences between the two groups, specific regions of interest (ROIs) were created. Significantly greater activation in the Rule than the Similarity group was found in three ROIs: left-hemisphere (hereafter “left”) superior, posterior, parietal cortex (BA 7/19); the right-hemisphere homologue of the preceding ROI; and right prefrontal cortex (BA 46). Each of these areas is known to be involved in WM processes. The posterior parietal areas are frequently activated in studies of WM, regardless of the kinds of materials involved (e.g., Wager & Smith, 2003), and they are assumed to mediate selective attention (keeping items active in WM requires that one attend to them). The right dorsolateral prefrontal cortex also has a history of involvement in studies of WM, both with humans (e.g., Smith & Jonides, 1999) and non-human primates (e.g., Goldman-Rakic, 1987; the fact that only the right prefrontal region was involved may be due to the use of visual items.

In a second experiment, Patalano et al. (2001) tested only a Rule group, but this time changed the rule on every block of test trials (so that participants could not associate a particular animal with a particular category). This change should have reduced the extent to which Rule participants might also have engaged an exemplar-similarity process (recall that Allen & Brooks, 1991, found behavioral evidence for such concurrent processing). A reduction in the use of exemplar similarity should have resulted in a reduction in activation in occipital areas. But no such reduction was found when the results of this experiment were compared to those of the first: The second experiment, like the first, showed that rule-based categorization was associated with activation in occipital cortex, prefrontal cortex, posterior parietal cortex, and the cerebellum. This null finding – no reduction in occipital activation – is difficult to interpret, given that it hinges on a comparison of different participants, and that the imaging modality, PET, may have lacked the spatial resolution to detect differences between the groups. The upshot is that this pair of studies provides some evidence for different categorization systems, but no evidence that the systems can operate concurrently.

As noted, the above results are limited by the lack of brain areas active in the Similarity group but not the Rule group. There are at least two methodological reasons why the above experiments may have failed to reveal a distinctive neural signature of an exemplar-similarity process. First, an important component of exemplar-similarity may be integrating the features in each representation, but Patalano et al.’s (2001) artificial animals were composed of relatively separable attributes. Second, previously experienced exemplars become less available as time passes, and in the Patalano et al. study the training session preceded the test session by a matter of days. These two factors were changed in a follow-up imaging study by Koenig, et al. (2005). Again, the study involved two different groups categorizing the same items by either Rule or Similarity instructions. Although the items were once more artificial animals, called CRUTTER, now they were created to be more naturalistic, with their six features appearing to be more integrated (see the bottom halves of Figures 1A and 1B). And now the training period preceded the test phase by just a matter of minutes.

There were other changes as well. During training, participants (again young normals) learned only a single category, and did so by a procedure in which on each trial they had to match one of two alternatives to a standard. In the Rule group, on each trial participants were presented two artificial animals (the alternatives), along with a brief description of the features defining the rule plus outline sketches of these features (see Figure 1A); the participant's task was to select the alternative that satisfied the rule. The rule required matches on 3 of 4 critical attributes (the 2 additional features in CRUTTER were distractors that did not contribute to category membership), and the correct alternative had 3 matches (a Member), whereas the other alternative had only 1 or 2 matches (Non-member). In the Similarity group, on each trial participants were presented two artificial animals (the alternatives) along with the prototypical animal of the category (see Figure 1B). The prototype contained all 4 critical features, the correct alternative had 3 of them (a Member), whereas the other alternative had only 1 or 2 of them (Non-members)[1]. Thus, for any given trial, the alternatives were identical for the Rule and Similarity groups, and the same alternative was the correct choice for both groups, though the correct choice was based on different reasons. The training session was followed immediately by the test phase, in which pictures of single animals were presented. The test items contained members with 3 or 4 critical features, ambiguous items that contained 2 critical features, and Non-members that contained only 1 or 0 of the critical features. Participants were imaged by functional Magnetic Resonance Imaging (fMRI) during both training and test.

Although our primary concern is with the imaging results, it is worthwhile to first consider the behavioral results obtained during the test session. Figure 2 presents the percentage of times that participants endorsed a test item as a category member as a function of the number of critical features the test item contained, separately for the Rule and Similarity groups. In the Rule group, endorsement was *categorical*, with nearly 100% endorsement for Members (3 or 4 critical features), and only 0–10% endorsement for Non-members (0 or 1 of the critical features). In the Similarity group, endorsement was *graded*, with substantially less than 100% endorsement for Members, and continuous decreases in endorsement rate for test items that were increasingly dissimilar from the prototype. These results are exactly what one would expect if the participants were using the two different categorization systems. However, the different endorsement profiles for the two groups can also be predicted by existent mathematical/ computational models of category learning that assume only a single categorization system (e.g., Love et al., 2004;Nosofsky & Zaki, 1998). We expand on this point later.

The imaging results for the test session are presented in Figure 3. These results are only for trials in which the test item was a clear-cut Member (3 or 4 critical features) or a clear-cut Non-member (1 or 0 of the critical features). Panel 3A contains activations obtained in the Rule group relative to the Similarity group (Rule minus Similarity contrast), whereas Panel B contains the activations in the Similarity group relative to the Rule group (Similarity minus Rule contrast). This time there is a double dissociation, as each group had its distinctive pattern of activations. The Rule group selectively activated three left-hemisphere regions: inferior parietal cortex (BA 40), parietal-occipital cortex (BA 19), and anterior cingulate (BA 24/32). In contrast, the Similarity group selectively activated two regions: the left and right temporal-parietal cortex (BA 39/22). The latter activation fits well with the cognitive operations involved in similarity-based categorization: bilateral temporal-parietal cortex has been previously

[1] In this type of training, the descriptions of the rule features and the prototype serve as reminders of the relevant features for category membership, and we needed to use such reminders in behavioral studies with neurological patients that will be described below.

associated with the integration of multiple features (Beauchamp, Lee, Argall, & Martin, 2004), and such integration was presumably involved in forming and comparing representations of test items to known CRUTTER. The areas specific to the Rule group are generally associated with selective attention.

The distinctive activations differ somewhat from those obtained in the Patalano et al (2001) PET study. Patalano et al. did not find the temporal-parietal activation associated with the Similarity group in the study currently under discussion, presumably because the artificial animals used by Patalano et al. required less integration than did CRUTTER. The Rule results for the current study also differ somewhat from those of Patalano et al. Although both studies reported parietal activations associated with attention, the area in the current study was distinctly inferior to that in Patalano et al. (BA 40 vs. 7, respectively), and was also left-lateralized rather than bilateral. The other activations associated with visual processing and attention – in extrastriate occipital cortex – were more comparable in the two studies. Lastly, it is unclear why the right frontal activation obtained in Patalano et al. (2001) did not occur in the study of current interest.

Why do the Rule-based activations in Koenig et al. (2005) differ from those obtained in the previous PET experiments? Again, one possible factor is the difference in materials. The artificial animals used by Patalano et al. (2001) varied on ten attributes, most of which appeared to be quite separable (e.g., Garner, 1976). Consequently, Rule participants may have had to sequentially scan each item, searching for the relevant attributes in the pictured animal; this kind of spatial search is associated with bilateral activity in superior parietal cortex (e.g., Grady et al., 1994; Horwitz et al., 1992; Wilkinson, Halligan, Henson, & Dolan, 2002). The artificial animals used by Koenig et al. (2005) varied on only six attributes, so there were fewer spatial locations to scan than in the items of Patalano et al. (2001). Also, the attributes used in Koenig et al. (2005) were more integral, which again may have further reduced the amount of spatial scanning needed. Lastly, the integral nature of the Koenig et al. animals may also have led to a collapse of two features into one (e.g. fang-like teeth and a pointed snout can be combined), which would have reduced the number of features that would have to be inspected; this could be why Koenig et al. (2005) failed to find the right frontal activation obtained in Patalano et al. (2001). All of this is highly speculative, and cries out for imaging studies that systematically vary the nature of the category items.

In sum, the double dissociation provided by Koenig et al (2005) provides strong evidence for two different categorization systems, one tied to WM, and the other to explicit LTM. But this study, like the preceding PET one, offers no evidence that the two categorization procedures can be used concurrently.

Using Novel Categories: Patient Studies

To corroborate the cognitive functions of the regions revealed in imaging studies, we need behavioral studies of patients with lesions in these same regions. In our research on categorization systems, we have followed the strategy of using the same or similar paradigm in (1) imaging studies with normals and (2) behavioral studies with patients who have brain disease in areas implicated in the imaging studies. The patient population that we have used most frequently are those with probable Alzheimer's disease (AD). AD is a progressive neurodegenerative disease that selectively involves specific brain regions early in the course of the disease, including the hippocampus, posterolateral temporal-parietal cortex, and dorsolateral prefrontal cortex. This distribution of disease results in difficulty in retrieving information from familiar categories (“semantic memory”) in 30% to 50% of AD patients (Grossman et al., 1996). Several imaging studies with AD patients have shown correlations between difficulty in semantic-mmory retrieval and functional cortical defects in dorsolateral prefrontal and posterolateral temporal-parietal cortices (Grossman et al., 1997; Desgranges et al., 1998). Thus there is abundant evidence that AD patients have difficulties in using familiar categories. Here we focus on whether these patients also have difficulties in using novel categories.

In a recent study (Koenig, Moore, Glosser, Grossman, & Smith, in press), we used the same behavioral paradigm as in the preceding imaging study, but our participants included AD patients and Corticobasal Degeneration (CBD) patients, as well as age-matched normal controls. The reason for including CBD patients is that they have disease in parietal and parietal-occipital cortex, and this localization of disease should result in deficits in similarity-based learning and categorization. The participants learned a single, novel category, CRUTTER, by either Rule or Similarity instructions. Again we used the standard-plus-alternatives learning procedure that is illustrated in Figure 1A and B; now the standard – a description of the rule or a picture of the prototype – was also a reminder of the instructions, which the AD patients needed given their explicit LTM deficits. Because of these deficits, we also used the same kind of reminder items during the test phase, which means there were some visual differences between the Rule and Similarity conditions during categorization.

Figure 4 presents the behavioral results for the test phase, with Panels A, B, and C containing the results for controls, AD patients, and CBD patients, respectively. Each panel presents the percentage of times that participants endorsed a test item as a category member as a function of the number of critical features that the test item contained, separately for the Rule and Similarity groups. The data for the controls (older normals) in Panel A are remarkably similar to the data for young normals in Figure 2: Endorsement is categorical in the Rule group – with nearly 100% endorsement for Members and only 0–10% endorsement for Non-members – but there is a more graded endorsement profile for the Similarity group. Panel B presents the data for AD patients. The most striking difference from the controls' data is that the AD patients show a graded profile for Rule instructions, comparable to that for Similarity instructions. The obvious interpretation is that the AD patients are unable to apply a rule even when they are reminded of it, presumably because rule-based categorization is neurally mediated by frontal and parietal areas (as shown by the imaging data), and these neural regions are damaged in AD. Additional evidence for this hypothesis comes from examination of categorization in patients with frontotemporal dementia (FTD) (Koenig, Smith, & Grossman, in press). FTD is another neurodegenerative disease that affects primarily the frontal and temporal regions of the brain, where a form of Primary Progressive Aphasia or Semantic Dementia is associated with predominantly left hemisphere disease, and a disorder of social comportment and executive functioning follows right hemisphere disease. Subgroups of FTD patients with frontal disease had difficulty judging CRUTTER on the basis of a rule, regardless of the presence or absence of an aphasia, although their performance with similarity-based judgments was essentially normal.

The results for CBD patients (Panel C) also show a graded profile in the Rule group, and a suggestion of a deficit in similarity-based categorization as well (compare the Similarity functions for the two patient groups). CBD patients, unlike AD patients, differed significantly from controls during their similarity judgments, particularly in their judgments of Members. Moreover, individual analyses of CBD patients revealed that in a majority of cases (63%) the patient judged the category membership of a test item in the Similarity condition on the basis of a single feature. In those CBD patients who did not focus on a single feature, the endorsement function was essentially flat.

In this experiment (as well as in the preceding one), we have taken a categorical endorsement profile as an indicator of rule-based categorization, and a graded endorsement profile as an indicator of similarity-based categorization. Two different profiles, hence two different systems. But our argument is hardly air-tight as different endorsement profiles for the two instructional conditions can be predicted by existent mathematical/computational models of category learning (e.g., Love et al., 2004: Nososky & Zaki, 1998). These models can accommodate different endorsement profiles by assuming that the different instructions led to differences in how the participants distributed their attention across the attributes (rather than to a difference in the memory system used). For example, Rule participants may have restricted their attention entirely to rule-relevant attributes, whereas Similarity participants may have distributed their attention more evenly among all attributes. However, further results from the experiment of interest provide independent evidence that two different categorization systems are indeed involved.

In addition to the categorization task, the patient participants were given some standard neuropsychological tests—including the digit span and Stroop tasks, which are assumed to tap WM and selective attention, respectively, as well as tests that reflect explicit LTM. WM and selective attention are involved in rule- but not similarity-based categorization, and thus performance on the neuropsychological tasks should correlate only with rule-based categorization. For AD patients, performance on the neuropsychological tasks tapping WM and attention did indeed correlate significantly with performance on rule-based categorization, but not with performance on similarity-based categorization. In contrast, performance on tasks reflecting explicit LTM correlated with neither kind of categorization (there was little variability on the LTM tasks, as all patients scored at low levels). This pattern of correlations fits nicely with dual-categorization systems, but single-system accounts may have some difficulties accommodating the findings.

There is still further evidence for the dual systems when we look at the data for training[2]. Figure 5 presents the number of trials that Rule and Similarity participants needed to reach a criterion level of learning, separately for controls, AD patients, and CBD patients. Although the two patient groups performed less well than controls on both kinds of categorization, there is a striking difference between the two kinds of patients. Compared to controls, AD patients were more impaired on rule- than similarity-based categorization, whereas CBD patients showed the reverse pattern of impairments. This reversal of the deficit pattern fits well with the dual-systems approach – prefrontal damage in AD particularly hurts rule-based categorization, whereas parietal and parietal-occipital damage in CBD particularly impairs similarity-based categorization.

The results of this experiment provide converging evidence for the claim that there are different categorization procedures, with one being based on WM and associated attentional processes, and the other being based on similarity. Again, though, the study does not speak to the issue of whether the two systems operate concurrently.

Using Natural Categories: Neuroimaging and Patient Studies

Researchers often use artificial rather than natural materials to study how people use categories because of the difficulty of controlling extraneous factors with natural categories. Still, natural categories are the ultimate target of the inquiry, and it is useful to establish that the distinction between rule-and similarity-based categorization applies to natural categories as well. We attempted to do this in a pair of related studies, both of which were based on prior behavioral studies.

In these behavioral experiments (Rips, 1989: Smith & Sloman, 1994), on each trial participants (young normals) were presented a description of of an object (e.g., “a round object two inches

[2] Our review emphasizes findings about category use, which is tapped by the test phase in experiments with artificial categories. We maintain this emphasis on category use rather than acquisition because some studies with artificial categories did not image the training phase (e.g., Patalano et al., 2001), and no study of natural categories can realistically image acquisition.

in diameter"), and were asked which of two categories fits best with the description (e.g., QUARTER or PIZZA). All of the descriptions included a quantity – e.g., diameter – that was between the relevant values of the category choices; moreover, one of the categories is "fixed" on the relevant attribute (e.g.,QUARTER is fixed on diameter), whereas the other is "variable" on the relevant attribute (e.g., PIZZA). The descriptions and two categories were presented to two groups who received different sets of instructions, one intended to induce similarity-based categorization (" make your judgment on the basis of overall similarity"), the other intended to induce rule-based categorization ("only one of the choices is correct"). In the latter case, participants need to identify the category that is fixed on the relevant attribute, note that the described object can not meet the criterion of this category, and choose the category that is variable on the relevant attribute. The behavioral studies showed that with rule-based instructions participants were more likely to choose the variable category (e.g., PIZZA), whereas with instructions promoting similarity-based categorization participants chose the two categories equally often. All of this provides behavioral evidence for two kinds of categorization.

We adapted the PIZZA-QUARTER paradigm for neuroimaging (Grossman, Koenig, et al., 2003). Young, normal participants were given either rule- or similarity-based instructions, and then made their categorization judgments on a series of items like the PIZZA-QUARTER example while having their brains scanned by fMRI. We temporally separated the presentation of the written description of the object from the choice between the two categories, and our analysis focused on the contrast between Choice and Description phases. Based on the neuroimaging studies with artificial categories, we expected that the rule-based instructions would lead to greater activation in certain prefrontal and parietal regions, whereas similarity-based instructions would lead to greater activation in temporal-parietal areas because of the role of feature integration.

The behavioral results obtained in this study replicated the prior behavioral studies: Rule-instructed participants were more likely to choose the variable category than similarity-instructed participants (roughly 80% vs. 60%). Of greater interest are the imaging results. The first relevant contrasts were the activation differences between the Choice and Description phases when the variable category was in fact selected, separately for the Rule and Similarity groups. Both groups showed numerous common activations, including left dorsolateral prefrontal cortex, left inferior parietal cortex, and the anterior cingulate. Apparently, some WM capacity and selective attention were required for both groups. The most important contrast, though, is between Rule (Choice minus Description) versus Similarity (Choice minus Description). Only the Rule group activated right (vental) prefrontal cortex, whereas only the Similarity group activated right parietal cortex. The Rule-specific prefrontal activation is similar to that obtained in the PET study by Patalano et al (2001), whereas the Similarity-specific activation is similar to that obtained in the fMRI study by Koenig et al. (2005). These findings provide a double dissociation between rule- and similarity-based categorization with natural categories, though the overall differences are less pronounced than those obtained with artificial categories.

To obtain converging evidence from patient data, we used the basic PIZZA-QUARTER paradigm with AD patients (Grossman et al., 2002). These patients' damage in prefrontal cortex should selectively impair their rule-based decisions, which means eliminating the preference for the variable category when instructions fostered the use of rules. This is exactly what we found. AD participants chose the variable category 58% and 55% with Rule and Similarity instructions, respectively, whereas the comparable numbers for age-matched controls were 75% and 55% (the latter finding again replicate the basic behavioral result). Thus AD patients were normal with Similarity instructions but impaired with Rule instructions. There is an interpretation problem with these findings, though. The Rule condition was more difficult than the Similarity condition, as shown by the fact that choice latencies were longer in the former than the latter condition. To deal with this problem, we included a second Similarity group in which the stimulus information was visually degraded, so much so that latencies in this Degraded Similarity group were as long as in the Rule group. Performance in this Degraded Similarity group was almost identical to that in the original Similarity group, implying that difficulty per se is not responsible for the AD's deficit with Rule instructions.

The last two studies provide more evidence for the existence of two categorization systems, one that relies on WM resources, and another that relies on explicit LTM and uses similarity as its guide to categorization. But neither experiment speaks to the issue of whether the two systems can operate concurrently.

Triggering Conditions

In the above studies, participants were always instructed how to categorize the items (in an effort to control the processing in each condition). This kind of paradigm is not representative of most categorization situations, in which no information is explicitly stated about how to categorize the items. For more natural situations, the question arises as to what aspects trigger the various categorization systems. Any proposal about multiple categorization systems needs to address this question.

There is more to say about this question for rule- than similarity-based categorization. We have noted the potential importance of the separability or analyzability of the items’ attributes. Items that can easily be analyzed into separable attributes may promote rule use, whereas those whose attributes form an integral whole may be more likely to be categorized by a similarity process. This proposal is related to the claim that the critical determinant of rule use is the verbalizability of the items involved (e.g., Ashby & Maddox, 2005), as more analyzable items can be more easily verbalized. Factors other than stimulus analyzability also may influence how easy it is to verbally describe category members. In studies reported by Ashby and Maddox (2005) the identical set of items were partitioned in two different ways, such that one partition could be captured by a simple, verbalizable rule, whereas the other partition could be expressed only by a very complex description; the two partitions fostered different kinds of categorization.

There may also be triggers for rule use that pertain to the procedure of an experiment. One such determinant is whether participants are provided feedback about their categorization decisions. It seems plausible that people are more likely to search for rules, and use them, when feedback is provided than when it is not. But we have to qualify this claim by adding that feedback may lead to rule use *only* to the extent that the items are analyzable. For Ashby, Maddox and their colleagues have reported numerous studies of categorization in which feedback is given but the materials are difficult to describe, and the outcome is that participants do not use rules (see Ashby & Maddox, 2005). Similarly, in studies of probabilistic categorization, feedback is given but its probabilistic nature seems to block rule use (e.g., Shohamy et al., 2004). Thus materials can trump procedure.

Aside from its obvious supervisory function, feedback may also induce in participants an intention to learn a category, and this intention alone may be a trigger for rule-based categorization. Thus category-learning studies in which there is a variation only in whether participants are informed that a category is present produce different results for those who know and those who don’t (e.g., Reber et al., 2003).

We know less about the triggering conditions for categorization based on explicit LTM, perhaps because this system may operate more automatically. Indeed, the very first study that we considered— the Allen and Brooks (1991) behavioral study—showed that participants instructed to categorize by a particular rule formed LTM representations of the items encountered, and then used the similarity of these representations to test items in making categorization decisions. These results have been interpreted as showing that categorization based on explicit LTM is more automatic (i.e., does not require intention) than rule-based categorization (Smith et al., 1998).

A CONTRAST OF LTM SYSTEMS: EXPLICIT VERSUS IMPLICIT CATEGORIZATION

The studies retviewed in this section all involve novel categories. Again, though, the studies may be divided into neuroimaging and patient experiments. Because the research began with patient studies and then moved to neuroimaging experiments, we follow this order in our presentation.

Using Novel Categories: Patient Studies

The contrast of interest is between categorization based on explicit LTM versus implicit LTM (or “explicit- versus implicit-categorization”, for short). The notion of implicit categorization was advanced in a seminal study by Knowlton and Squire (1993). Their concern was whether medial-temporal lobe amnesics, who have minimal explicit LTM, were capable of learning a novel category. To study this issue, they compared amnesics and age-matched controls on a *prototype-extraction task*. Both groups of participants were presented a series of dot patterns. All of the patterns had been created by starting with a prototype pattern, and then transforming it to some degree by moving some proportion of the dots (after Posner & Keele, 1968). During training, participants look at dot patterns and nothing was mentioned about a category to either the amnesics or control. After training, all participants were informed that the patterns they had just seen belonged to the same category, and that they now had to determine which of a sequence of test items also belonged to that category. Note that this procedure is a variant of standard implicit-memory methodology: the purpose of the training information was disguised so that participants did not intentionally try to remember it, and hence presumably had to rely on an implicit system during the test phase.

During test, a series of novel items was presented, which varied in how similar they were to the prototype that had spawned the training items. Both the amnesics and the controls performed the unexpected categorization task with above chance accuracy, and the amnesics performed as accurately as the controls. These results were in sharp contrast to findings obtained in a test of recognition memory for the same kind of dot patterns. In the latter task, the patients performed significantly worse than controls though the test was not particularly demanding. The obvious conclusion was that category learning and use could be based on implicit memory, which remains intact when the medial temporal lobe is damaged.

Further evidence for implicit category learning was provided by examining the extent to which a participant endorsed a test item as a category member as a function of its similarity to the underlying prototype of the category (which was not presented during training). Amnesics and controls showed comparable endorsement profiles, and for both groups endorsement decreased as the test item became increasingly dissimilar from the prototype (Knowlton & Squire, 1993; Eldridge, Masterman, & Knowlton, 2002). These endorsement profiles are like those reviewed earlier for categories that had been learned by intentional similarity instructions (see Figure 2 and Figure 4). Thus, similarity-based categorization can be based on either explicit or implicit memory (though the similarity metrics may differ).

Several subsequent experiments have used the prototype-extraction task and have replicated the above findings for patients—relatively intact categorization, but impaired recognition—using as patients either medial-temporal-lobe amnesics (Kolodny, 1994: Reed, Squire, Patalano, Smith, & Jonides, 1999), or ADs (Bozoki, Grossman, & Smith, 2006; Eldridge, Masterman, & Knowlton, 2002). It is worth noting that in Reed et al. (1999) and Bozoki et al. (2006) the items were not abstract dot patterns, but rather a set of artificial animals (different from any used in the studies discussed earlier). Whereas abstract dot patterns relatively non-analyzable and are difficult to describe verbally, the artificial animals of current interest were composed of salient and separable attributes that could easily be verbalized (e.g., “a creature with a round striped body”). The fact that studies using these items obtained the usual results suggests that the implicit categorization system of interest can operate on both easy- and difficult-to-describe materials[3].

Another important aspect of the above studies concerns the structure of the attributes that comprise the category. The training instances share multiple features on a substantial set of correlated attributes, with no single feature or pair of features being defining of category membership. This mirrors the structure of most natural categories (e.g., Rosch, 1975; Smith & Medin, 1981), and is unlike the structure of the categories used in the experiments considered in the previous section that contrasted categorization based on multi-feature rules versus similarity. A consequence of the natural structure used in the implicit categorization studies is that most instances of the category were highly similar to one another (because they share multiple features). These implicit studies also typically included a recognition task--to show the dissociation between recognition and categorization—and for this task used items that were relatively dissimilar. The difference in inter-item similarity between the two tasks has led some researchers in mathematical learning theory to the conclusion that the results obtained in these studies are in fact explainable by a single categorization system (e.g., Love & Gureckis, in press; Nosofsky & Zaki, 1998).

To illustrate the point, consider an application of the SUSTAIN model of category learning to the finding of a dissociation between categorization and recognition (Love & Gureckis, in press). SUSTAIN assumes that to the extent training items are similar, participants will group them together into a single cluster (which is like a prototype), whereas dissimilar items will require their own clusters (i.e., they are treated like exemplars). The consequence for the above-mentioned prototype-extraction studies is that the similar training items in any of the categorization tasks could be represented as a single cluster, whereas the more dissimilar items used in the recognition tasks would require multiple clusters. Assuming only that patients with medial-temporal lobe damage are less likely to form a new cluster than are controls (which is captured by a single parameter in SUSTAIN), the model correctly predicts that the patients will be impaired on recognition but not categorization[4]. The same instantiation of the model can also predict the graded endorsement profile that is obtained in prototype-extraction studies.

Clearly, we are in need of further neuropsychological experiments that use the same materials for categorization and recognition, and that employ comparable tasks. Even if such experiments are performed and continue to produce a dissociation between categorization and recognition, it may still be possible for some mathematical models of category learning to account for the dissociation because these models contain multiple parameters, any of which can take on different values for patients and controls. In view of this, and because of the ever-present need for converging evidence, it is useful to turn to neuroimaging studies that contrast explicit versus implicit categorization.

[3] The issue about verbalizability is by no means resolved, though, as there are some differences in performance on prototype-extraction for dot patterns versus artificial animals. See Smith (2007) for a discussion of the relevant findings.

[4] This implementation of the model does not imply that the recognition task is more difficult, because the comparison of test items to stored representations may result in closer matches with multiple clusters than with a single one.

Using Novel Categories: Neuroimaging Studies

Evidence for two systems—Using fMRI, researchers have asked if prototype-extraction and recognition tasks recruit different neural networks, even in normal controls (this kind of research has not yet been done with patients). Reber, Stark, and Squire (1998) performed exactly this kind of experiment, using the same dot-pattern tasks employed by Knowlton and Squire (1993) but with young, normal controls. In prototype-extraction, participants first were presented different distortions of a prototype pattern, and then categorized novel patterns while being imaged by fMRI; in recognition, participants first were presented a few patterns (different from those in prototype-extraction), and then made Old-New judgments while being imaged by fMRI. The behavioral results indicated that, if anything, categorization was the more difficult task (categorization accuracy was only 58%, whereas the hit rate in recognition was 81%).

Of greater importance are the imaging results. In categorization, the fMRI contrast of interest was the difference in activation during test between a category member and a nonmember (Member minus Non-member); in recognition, the contrast of interest was the difference in activation between an Old item and a New one (Old minus New). These contrasts produced striking differences between the two tasks. The recognition task resulted in numerous activations, including several areas in the prefrontal cortex and the medial temporal lobe, areas that have repeatedly been implicated in neuroimaging studies of explicit memory (see Squire, Clark, & Bayley, 2004 and Wagner, Bunge, & Badre, 2004). But none of these areas was activated in categorization. Even more dramatically, an area in posterior occipital cortex known to be involved in visual processing (Broadman Area 17/18) was activated in recognition but *deactivated* in categorization. At face value, these results provide strong evidence that different memory systems are involved in categorization and recognition.

For our purposes, though, this study has some limitations. For one, the difference in materials used in the two tasks again leaves open the possibility that the greater activations in recognition than categorization may be due to the larger number of clusters that have to be created in recognition. A second problem is that the critical contrasts for categorization and recognition may not be strictly comparable—the difference between a category member and a nonmember is somewhat arbitrary in this task (there is no sharp division between members and nonmembers), and may be different in kind from the non-arbitrary difference between an Old and a New item. A more recent fMRI study by Reber, Gitelman, Parrish, and Mesulam (2003) solves these two problems by using only the prototype-extraction task but varying whether the task is performed with the usual incidental instructions (participants don’t know a category is present at the beginning of training) or instead with intentional instructions (participants are told there's a category during training). Because the items are identical in both conditions, and because the only fMRI contrast is between members and non-members at test, Reber et al. (2003) eliminated the two problems mentioned above. The expectation was that intentional instructions should engage categorization based on explicit LTM, which is mediated by structures in the medial temporal lobe (particularly the hippocampus), whereas incidental instructions should engage categorization based on implicit LTM, which is subserved by different neural structures.

Two different groups of participants (young normals) were used. During training, the incidental group was told to point to the center of each dot pattern, whereas the intentional group was told that the patterns all came from the same category and that they were to learn it. Participants were imaged by fMRI during the test trials when both groups were instructed to discriminate category members from non-members. The behavioral results showed that the intentional instructions led to substantially better categorization performance than incidental instructions. Such an effect is hardly diagnostic about whether there are separate explicit and implicit category learning systems, as the difference in accuracy could merely reflect a difference in the amount of attention paid during training. The imaging results were more informative. In line with findings reviewed earlier, intentional instructions led to increased activations in a number of brain areas, including prefrontal (BA 10) and parietal areas (BA 7); in line with the Reber et al. (1998) study discussed before, incidental instructions led to a deactivation in visual cortex (BA 19). The fact that a difference in instructions determined whether the pattern was one of activations or deactivation implies that the different instructions are recruiting qualitatively different categorization systems.

In a contrast of particular interest, Reber at al. (2003) focused on neural activity in two target regions, the hippocampus and the posterior occipital region (BA 19) that was deactivated in the incidental condition (as well as in Reber et al., 1998). Substantial research indicates that, relative to an appropriate baseline, hippocampal activation is a marker of explicit memory, whereas deactivation of extrastriate occipital cortex is a marker of implicit memory (as reflected in perceptual priming, see Buckner, 2000 for a relatively recent review). The results of this contrast showed a double dissociation; in the hippocampus, intentional instructions led to greater activity than did incidental instructions; in the occipital area, incidental instructions led to a greater deactivation than did intentional instructions. These results cannot easily be accounted for by any single system model of category learning.

Although the Reber et al (2003) imaging data are informative, they may underestimate what can be learned in the prototype-extraction task. For one thing, accuracy in the incidental condition was not much above chance—59%—which is lower than that typically reported for this task. Second, Reber et al. chose to block their test trials by whether the items were mostly categorical (clear-cut Members) or mostly non-categorical (clear-cut Non-members). As the authors themselves note, such a blocking procedure could have induced the participants to employ a task-specific strategy – e.g., “if I see a couple of category members, keep endorsing the items until I see some non-members”. This kind of strategy would mask what has actually been learned.

Evidence for concurrent operation of the two systems—We have recently performed a follow-up to the Reber et al. (2003) experiment that focuses only on the incidental condition, which presumably taps implicit categorization, and that corrects the problems mentioned above (Koenig et al., 2007). We tested young, normal controls in prototype-extraction, with the category CRUTTER (see Figures 1A and 1B). During training, participants saw 8 different category members presented 5 times each, and were instructed to simply look at the animals. During the test period, participants were imaged while they categorized novel and old Members, and Non-members (some items of ambiguous category membership were also included, but they are not of critical concern). The different types of items were randomly intermixed in an event-related design.

The behavioral results indicated that categorization accuracy was over 80% for novel items, substantially above that obtained by Reber et al. (2003). Furthermore, we found the usual endorsement profile for the participants, with endorsement rates being highest for novel prototypical Members (92%), next highest for old Members (83%), next for novel Members (74%), and lowest for Non-members (38%). The difference in accuracy between old and new Members suggests that explicit memory might have contributed to the findings.

The imaging results support this suggestion. Selected results are presented in Figure 6. Unlike Reber et al.'s (2003) results for their incidental condition, we found activation in the medial temporal lobe in the area of the hippocampus (Figure 6, Panel A). This activation suggests that participants may have been explicitly retrieving category exemplars seen during training, or explicitly encoding novel test items, or both. This in turn implies that though the task supposedly taps implicit categorization, explicit categorization also played a role. There were other activations as well, including bilateral prefrontal cortex (BA 46/9), bilateral parietal cortex (BA 7), and bilateral anterior cingulate (BA32). These latter areas are often found in studies of working memory (e.g., Wager & Smith, 2003), suggesting that the WM system too may have played a role (perhaps in trying to learn more about the category during the test phase —see Bozoki et al., 2006).

Of particular interest are certain contrasts that were intended to isolate the contributions of implicit and explicit categorization. One such contrast was that used by Reber and colleagues, Members minus Non-members. Like Reber et al. (1998; 2003), we found a deactivation in left occipital cortex (BA 19), which is a signature of implicit memory (see Figure 6B). Unlike Reber et al., our Member minus Non-Member contrast also revealed activation in a left medial temporal region that includes the hippocampus, which is the signature of explicit memory (see Figure 6B). Furthermore, activation changes in these two regions were correlated with behavioral performance. For the hippocampal region, which may be involved in explicit encoding of novel test items, activation for non-members during the first half of the test session (when items of this type were still novel) was highly correlated with categorization accuracy for category members. For the left occipital region that showed a deactivation for members, the magnitude of this deactivation correlated positively with categorization accuracy of members. This result supports the claim that implicit memory contributed to categorization.

The above results indicate that both implicit and explicit categorization were operative during the test trials. This is the strongest evidence that we have found for the concurrent operation of two categorization systems. There was, however, no evidence that the two systems interacted, as deactivation in the occipital region did not correlate with activation in the hippocampal region.

SUMMARY

Our initial contrast was between two explicit systems: categorization based on WM versus categorization based on explicit LTM. Three sources of evidence were provided for the distinction. First, there was the behavioral evidence from controls in these studies, particularly the shape of the function relating endorsement of a test item to the similarity of that test item to a prototypical category member. This endorsement function had a categorical profile with Rule instructions, but a graded profile with Similarity instructions. These different patterns were found in two studies, one involving senior participants (Koenig et al., in press) and the other involving younger participants (Koenig et al., 2005). The different endorsement functions fit perfectly with standard assumptions about how rules and similarity can be used to categorize novel items (Patalano et al., 2001). But the impact of these results is qualified by the fact that existent single-system models of categorization may be able to accommodate the variation in endorsement profiles by assuming that the different instructions led to differences in how the participants distributed their attention across the attributes (e.g., Love et al., 2004; Nosofsky & Zaki, 1998).

The second source of evidence for two different explicit systems was the pattern of categorization deficits in two patient populations (AD and CBD patients). AD patients, who have selective damage in prefrontal cortex, were impaired in rule- but not similarity-based categorization. This dissociation was obtained in experiments with both novel, perceptual categories (Koenig et al., in press) and natural, semantic categories (Grossman et al., 2002). Again it may be possible for existent single-system models to account for this dissociation by variations in memory-comparison and attentional processes, but the basic dissociation was bolstered by the finding that AD performance on rule-based categorization was correlated with their performance on standard neuropsychological measures of WM and selective attention. Furthermore, an analysis of training effects showed that CBD patients and AD patients had reverse patterns of deficits. Whereas AD patients had a greater deficit in learning a rule-than a similarity-based category, CBD patients had a greater deficit in learning a similarity-based category. This reversal fits with the nature of damage in CBD, which includes certain parietal-occipital regions that are likely involved in similarity judgments. This rich set of data on deficits provides more of a challenge to single-system models than just a single dissociation.

The third source of evidence for two different explicit categorization systems came from neuroimaging studies. Two studies with novel categories (Patalano et al., 2001; Koenig et al., 2005), as well as one with natural categories (Grossman, Koenig, et al., 2003), found differences between the two kinds of explicit categorization. Although Patalano et al.'s (2001) PET study found just a single dissociation – only Rule instructions led to distinctive activations – the two more recent fMRI studies reported a double dissociation – there were both Rule-distinctive areas (including prefrontal regions), and Similarity-distinctive areas (including parietal regions). The problem with these imaging findings, though, is that the exact regions activated varied across the experiments. Although there are many differences among the three studies of interest, perhaps the most striking one is the materials used: the categories range from artificial animals composed of many separable attributes, to more realistic artificial animals composed of more integrated attributes, to descriptions of natural objects.

The other contrast of categorization systems that we reviewed involved categorization based on explicit LTM versus categorization based on implicit LTM. Two sources of evidence were used to support this distinction, neuropsychological studies with patients and neuroimaging studies with young normals. The major finding from the neuropsychological studies was that patients with medial-temporal-lobe damage performed normally on an implicit categorization task while showing substantial impairment on standard recognition tasks that tap explicit LTM (e.g., Knowlton & Squire, 1993; Eldridge et al., 2002). This dissociation has been obtained numerous times, and with materials that are easy or difficult to describe (e.g., Bozoki et al., 2006). Furthermore, these studies also showed graded endorsement profiles for implicit categorization. Again the impact of these results is compromised by the fact that formal models of category learning can explain the dissociation – categorization intact but recognition impaired – by assuming that the categorization tasks used make less demands on a component memory process shared by both kinds of tasks (Love & Gureckis, in press; Nosofsky & Zaki, 1998; Smith, 2007).

The second source of evidence for two different LTM categorization systems came from neuroimaging studies. Studies by Reber et al. (1998; 2003) showed that whereas recognition or categorization based on explicit memory led to a specific pattern of activations, categorization based on implicit memory led to a specific deactivation in posterior cortex. This deactivation in posterior cortex was recently replicated by Koenig et al. (2007), who showed that the magnitude of the deactivation was correlated with categorization performance. In addition, Koenig et al. (2007) found hippocampal activity, which presumably reflects a contribution of explicit memory to a task that previously had been thought to tap only implicit categorization. The occurrence of hippocampal activation *and* posterior deactivation provides evidence that both explicit- and implicit-LTM systems were active concurrently.

In closing, it is worth contrasting our tripartite system -- involving WM, explicit LTM and implicit LTM -- with the dual-categorization system proposed by Ashby and Maddox (e.g., 2005)-- involving rule-based WM and implicit LTM. A recent neuroimaging study of young normals investigated the neuroanatomic basis for the latter approach and provides a direct contrast of categorization based on WM versus that based on implicit LTM (Nomura et al., 2006). The study contrasted the information-integration task, described earlier, with a rule-based task. The two tasks used the identical stimuli, but the stimuli were divided into categories in different ways in the two tasks. The information-integration task required categorization based on the combination of two dimensions, and that combination was not easily named; the rule-based WM task used a single criterion that could easily be verbalized. Activations during performance of the two tasks overlaped considerably, but there were also noticeable differences between the two tasks. Comparisons of activations during correct and incorrect categorization trials during the rule-based task revealed activation of frontal, temporal and occipital neocortex bilaterally as well as activation of the hippocampus bilaterally and of the right caudate. A direct contrast of activation in regions of interest – the hippocampus and the body of the caudate – showed greater activation in the left hippocampus for the rule-based task than the information-integration task. This resembles the activations that we observed during rule-based categorization described earlier (Koenig et al., 2005). However, a contrast of activations during correct and incorrect judgments during the information-integration task showed bilateral activation of the body of the caudate, and the region of interest analysis showed greater activation in the body of the right caudate in the information-integration than the rule-based task.

Thus the implicit memory system associated with prototype-extraction appears to differ from that recruited by the information-integration task. While we too observed activation in the head of the right caudate during the implicit acquisition of CRUTTER (Koenig et al., 2007), the general pattern of recruitment that we observed during prototype-extraction otherwise differs considerably from those of Nomura et al (2006). We cannot compare these two implicit memory systems directly, however, because the tasks differ so markedly: Koenig et al (2007) used one category, for example, while Nomura et al (2006) used two categories; Koenig et al did not provide feedback while Nomura et al did; the category taught by Koenig et al is relatively natural and verbalizeable while the category used by Nomura et al is neither. Additional work is needed to provide a more sensitive test of whether there are indeed two different implicit, categorization system.

References

Allen SW, Brooks LR. Specializing the operation of an explicit rule. Journal of Experimental Psychology 1991;120:3–19.

Ashby FG, Maddox WT. Human category learning. Annual Review of Psychology 2005;56:149–178.

Baddeley, AD. Working memory. Oxford: Oxford University Press; 1986.

Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 2004;41:809–823. [PubMed: 15003179]

Bozoki A, Grossman M, Smith EE. Can patients with Alzheimer's disease learn a category implicitly? Neuropsychologia 2006;44:816–827. [PubMed: 16229868]

Buckner, RL. Neuroimaging of memory. In: Gazzaniga, M., editor. The New Cognitive Neurosciences. Vol. 2nd Ed.. Cambridge, MA: MIT Press; 2000. p. 1013-1022.

Desgranges B, Baron J-C, de la Sayette V, Petit-Taboue MC, Benali K, Landeau B, Lechevalier B, Eustache F. The neural substrates of memory systems impairment in Alzheimer's disease: A PET study of resting brain glucose utilization. Brain 1998;121:611–631. [PubMed: 9577389]

Eldridge LL, Masterman D, Knowlton BJ. Intact implicit habit learning in Alzheimer's disease. Behavioral Neuroscience 2002;116:722–726. [PubMed: 12148939]

Foerde, K.; Knowlton, BJ.; Poldrack, RA. Modulation of competing memory systems by distraction; Proceedings of the National Academy Sciences; 2006. p. 11778-11783.

Garner WR. Interaction of stimulus dimensions in concept and choice processes. Cognitive Psychology 1976;8:98–123.

Goldman-Rakic, PS. Circuitry of primate prefrontal cortex and regulation of behavior representational memory. In: Plum, F., editor. Handbook of physiology, Section 1. Vol. Vol. 7. Bethesda, MD: American Physiological Society; 1987. p. 373-417.

Grady CL, Maisog J, Horwitz B, Ungerleider LG, Mentis MJ, Salerno JA, Pietrini P, Wagner E, Haxby JV. Age-related changes in cortical blood flow activation during visual processing of faces and location. Journal of Neuroscience 1994;14:1450–1462. [PubMed: 8126548]

Grossman M, D′Esposito M, Hughes E, Onishi K, Biassou N, White-Devine T, Robinson KM. Language comprehension difficulty in Alzheimer's disease, vascular dementia, and fronto-temporal degeneration. Neurology 1996;47:183–189. [PubMed: 8710075]

Grossman M, Koenig P, Glosser G, DeVita C, Moore P, Rhee J, Detre J, Alsop D, Gee JC. Neural basis for semantic memory difficulty in Alzheimer's disease: An fMRI study. Brain 2003;126:293–311.

Grossman M, Smith EE, Koemg P, Glosser G, DeVita L, Moore P, McMillan C. The neural basis for categorization in semantic memory. Neuroimage 2002;17:1549–1561. [PubMed: 12414293]

Grossman M, Smith EE, Koenig P, Glosser G, Rhee J, Dennis K. Categorization of object descriptions in Alzheimer's disease and frontotemporal demential: limitation in rule-based processing. Cognitive, Affective, & Behavioral Neuroscience 2003;3:120–132.

Grossman M, White-Devine T, Payer F, Onishi K, D′Esposito M, Robinson KM, Alavi A. Constraints on the cerebral basis for semantic processing from neuroimaging studies of Alzheimer's disease. Journal of Neurology, Neurosurgery and Psychiatry 1997;63:152–158.

Horwitz B, Grady CL, Haxby JV, Schapiro MB, Ungerleider LG, Mishkin M. Functional associations among human posterior extrastriate brain regions during object and spatial vision. Journal of Cognitive Neuroscience 1992;4:311–322.

Hull CL. Quantitative aspects of the evolution of concepts. Psychological Monographs 1920;28:1–86.

Jonides, J. Working memory and thinking. In: Smith, E.; Osherson, D., editors. An invitation to cognitive science: Thinking. Vol. 2nd ed.. Vol. Vol. 3. Cambridge, MA: MIT; 1995. p. 215-265.

Knowlton BJ, Mangels JA, Squire LR. A neostriatal habit learning system in humans. Science 1996;273:1399–1402. [PubMed: 8703077]

Knowlton BJ, Squire LR. The learning of categories: Parallel brain systems for item memory and category knowledge. Science 1993;262:1747–1749. [PubMed: 8259522]

Koenig P, Moore P, Glosser G, Grossman M, Smith EE. Categorization of novel animals by patients with Alzheimer's disease and corticobasal degeneration. Neuropsychology. in press

Koenig P, Smith EE, Glosser G, DeVita C, Moore P, Mc Millain C, Gee J, Grossman M. The neural basis for novel semantic categorization. Neuorimage 2005;24:369–383.

Koenig P, Smith EE, Grossman M. Semantic categorization of novel objects in frontotemporal dementia. Cognitive Neuropsychology. in press

Koenig P, Smith EE, Troiani V, Antani W, McCawley G, Moore P, Cross K, Grossman M. Collaborating implicit and explicit memory mechanisms: Evidence from Alzheimer's disease and fMRI. Manuscript submitted for publication. 2007

Kolodny JA. Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. Psychological Science 1994;5:164–169.

Love BC, Gureckis TN. Bridging levels: A cognitive model of hippocampal mediated learning. Cognitive, Affective, & Behavioral Neuroscience. in press

Love BC, Medin DL, Gureckis TM. SUSTAIN: A network model of category learning. Psychological Review 2004;111:309–332. [PubMed: 15065912]

Murphy, GL. The big book of concepts. Cambridge, Massachusetts: MIT Press; 2002.

Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitetlman DR, Parrish TB, Mesulam M-M, Reber PJ. Neural correlates of rule-based and information-integration visual category learning. Cerebral Cortex 2006;17:37–43. [PubMed: 16436685]

Nosofsky RM, Palmeri TJ, McKinley SC. Rule-plus-exception model of classification learning. Psychological Review 1994;101:53–79. [PubMed: 8121960]

Nosofsky RM, Zaki SR. Dissociations between categorization and recognition in a amnesic and normal individuals: An exemplar-based interpretation. Psychological Science 1998;9:247–255.

Patalano AL, Smith EE, Jonides J, Koeppe RA. PET evidence for multiple strategies of categorization. Cognitive, Affective, & Behavioral Neuroscience 2001;1:360–370.

Posner MI, Keele SW. On the genesis of abstract ideas. Journal of Experimental Psychology 1968;77:353–363. [PubMed: 5665566]

Reber PJ, Gitelman DR, Parrish TB, Mesulam MM. Dissociating explicit and implicit category knowledge with fMRI. Journal of Cognitive Neuroscience 2003;15:574–583. [PubMed: 12803968]

Reber PJ, Stark CEL, Squire LR. Contrasting cortical activity associated with category memory and recognition memory. Learning & Memory 1998;5:420–428. [PubMed: 10489259]

Reed JM, Squire LR, Patalano AL, Smith EE, Jonides JJ. Learning about categories that are defined by object-like stimuli despite impaired declarative memory. Behavioral Neuroscience 1999;113:411–419. [PubMed: 10443769]

Rips, LJ. Similarity, typicality, and categorization. In: Vosniadou, S.; Ortony, A., editors. Similarity and analogical reasoning. Cambridge: Cambridge University Press; 1989. p. 21-59.

Roediger, HL.; McDermott, KB. Implicit memory in normal human subjects. In: Boiler, F.; Grafman, J., editors. Handbook of neuropsychology. Vol. Vol. 8. Amsterdam: Elsevier; 1993. p. 63-131.

Rosch E. Cognitive representations of semantic categories. Journal of Experimental Psychology: General 1975;104:192–233.

Schacter DL. Implicit memory: History and current status. Journal of Experimental Psychology: Learning, Memory, and Cognition 1987;13:501–518.

Schacter DL, Buckner RL. Priming and the brain. Neuron 1998;20:185–195. [PubMed: 9491981]

Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA. Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain 2004;127:851–859. [PubMed: 15013954]

Smith EE. The case for implicit category learning. 2007Manuscript submitted for publication.

Smith EE, Jonides J. Storage and executive processes in the frontal lobes. Science 1999;283:1657–1661. [PubMed: 10073923]

Smith, EE.; Medin, DL. Categories and concepts. Cambridge, MA: Harvard University Press; 1981.

Smith EE, Sloman SA. Similarity-versus rule-based categorization. Memory and Cognition 1994;22:377–386.

Smith EE, Patalano A, Jonides I. Alternative mechanisms of categorization. Cognition 1998;65:167–196. [PubMed: 9557382]

Squire, LR.; Clark, RE.; Bayley, PJ. Medial temporal lobe function and memory. In: Gazzaniga, M., editor. The cognitive neurosciences. Vol. 3rd Ed.. Cambridge: MIT Press; 2004. p. 691-708.

Squire, LR.; Knowlton, BJ. The medial temporal lobe, the hippocampus, and the memory systems of the brain. In: Gazzaniga, M., editor. The new cognitive neurosciences. Vol. 2nd Ed.. Cambridge: MIT Press; 2000. p. 756-776.

Tracy JI, Mohamed F, Faro S, Pinus A, Tiver R, Harvan J, Bloomer C, Pyrros A, Madi S. Differential brain responses when applying criterion attribute versus family resemblance learning. Brain and Cognition 2003;51:276–286. [PubMed: 12727182]

Wager T, Smith EE. Neuroimaging studies of working memory: A meta-analysis. Cognitive, Affective, & Behavioral Neuroscience 2003;3:255–274.

Wagner, AD.; Bunge, SA.; Badre, D. Cognitive control, semantic memory, and priming: Contributions from prefrontal cortex. In: Gazzaniga, M., editor. The cognitive neurosciences. Vol. 3rd Ed.. Cambridge: MIT Press; 2004. p. 709-725.

Wilkinson DT, Halligan PW, Henson RNA, Dolan RJ. The effects of interdistracter similarity on search processes in the superior parietal cortex. Neuroimage 2002;1:611–619. [PubMed: 11848704]

Wheeler ME, Buckner RL. Functional dissociation among components of remembering: Control, perceived oldness, and content. Journal of Neuroscience 2003;23:3869–3880. [PubMed: 12736357]

Wheeler, ME.; Peterson, SE.; Buckner, RL. Memory's echo: Vivid remembering reactivates sensory-specific cortex; Proceedings of the National Academy of Sciences; 2000. p. 11125-11129.

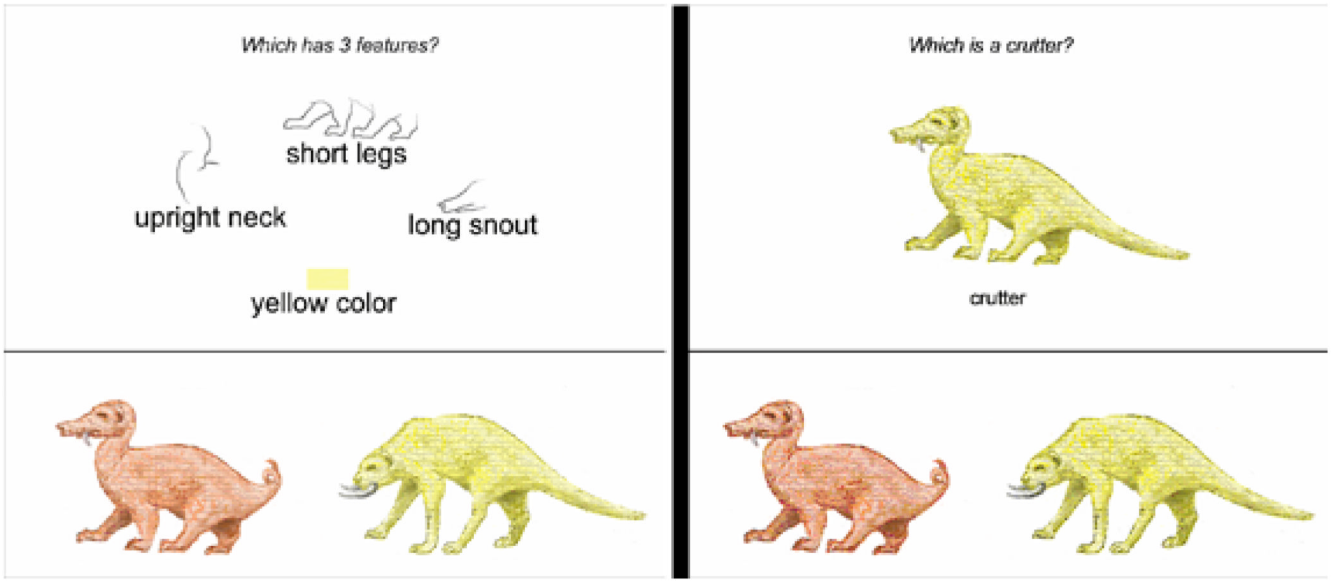


Figure 1. Koenig et al. (2005) experiment

Illustration of arrays seen during the training session. Panel A: rule-based training, showing the four criterial features in the upper half of the screen and two CRUTTER in the lower half (Member on left). Panel B: Similarity-based training, showing a prototype in the upper half of the screen and two CRUTTER in the lower half (Member on left).

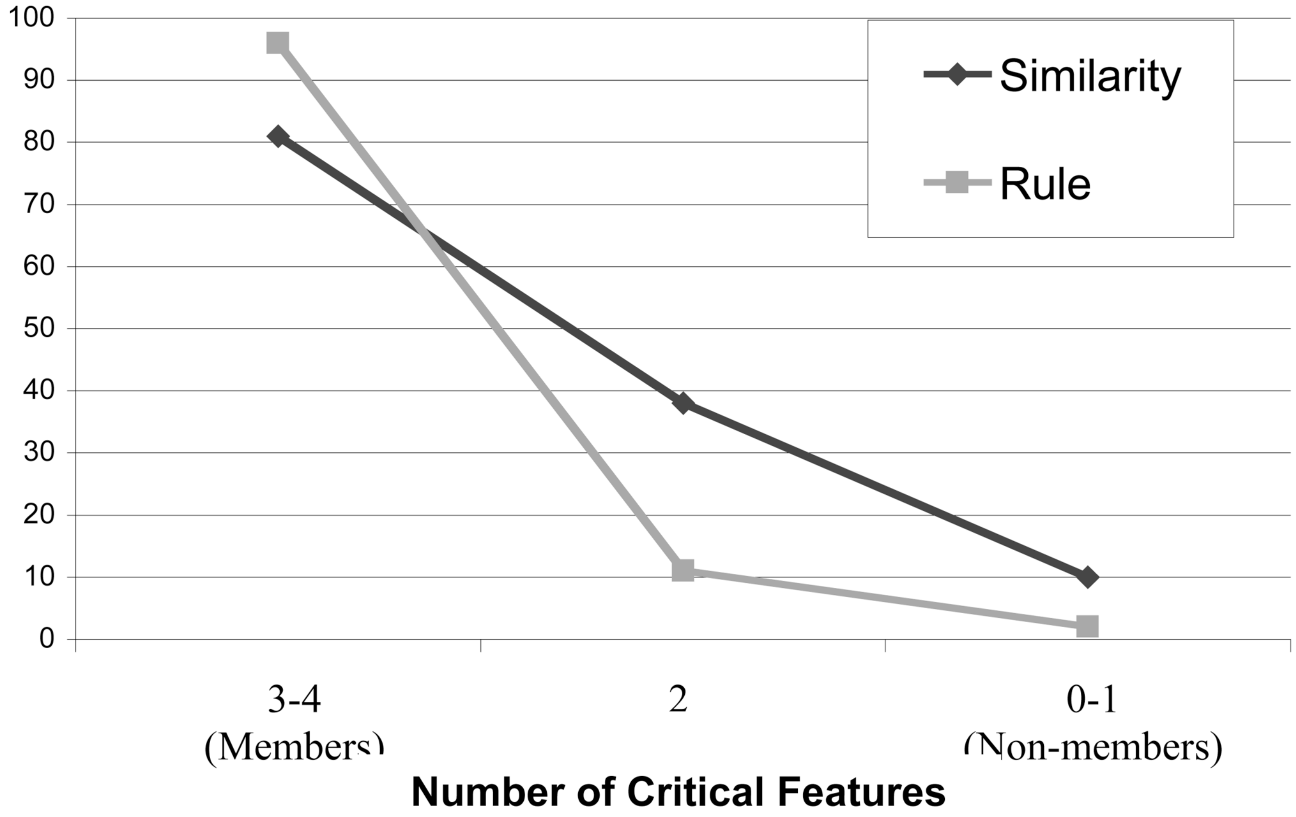
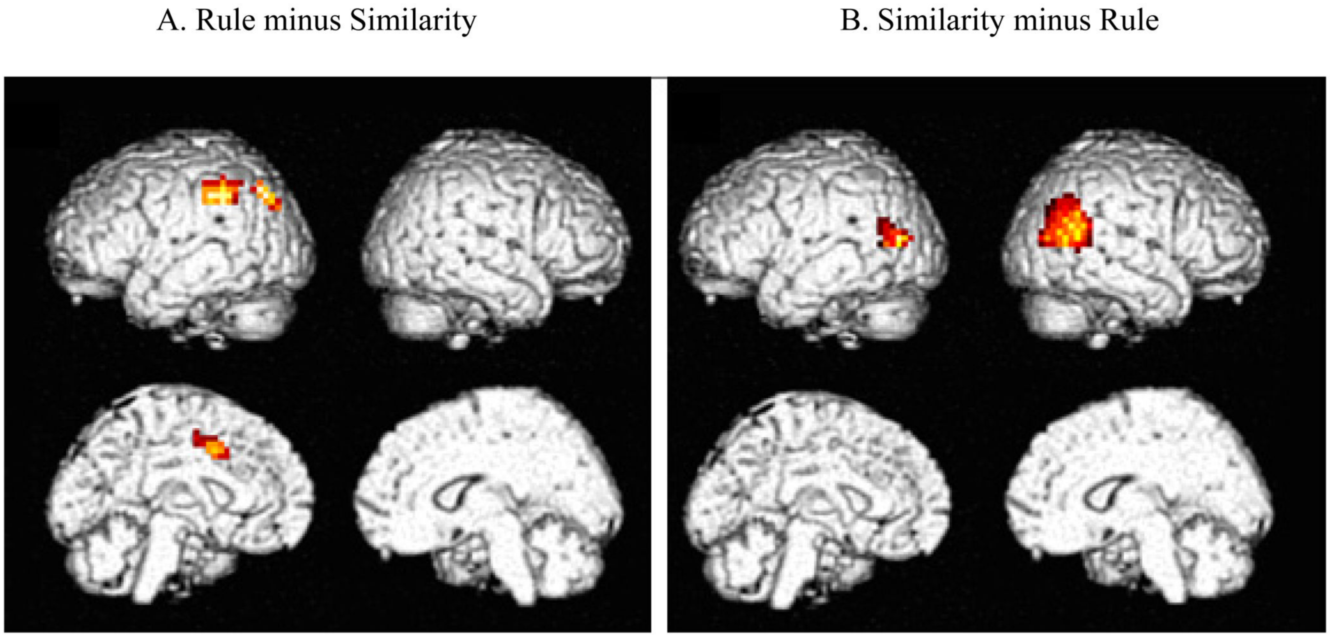


Figure 2. Koenig et al. (2005) experiment, Behavioral results

Category endorsement functions, which show the percentage of time that participants endorsed a test item as a function of the number of critical features it contained, separately for Rule and Similarity groups.



A. Rule minus Similarity

B. Similarity minus Rule

Figure 3. Koenig et al. (2005) experiment, Imaging results

Right panel: Rule- minus Similarity-based categorization for clear-cut Members and Non-members. Panel B: Similarity- minus Rule-based categorization for clear-cut Members and Non-members.

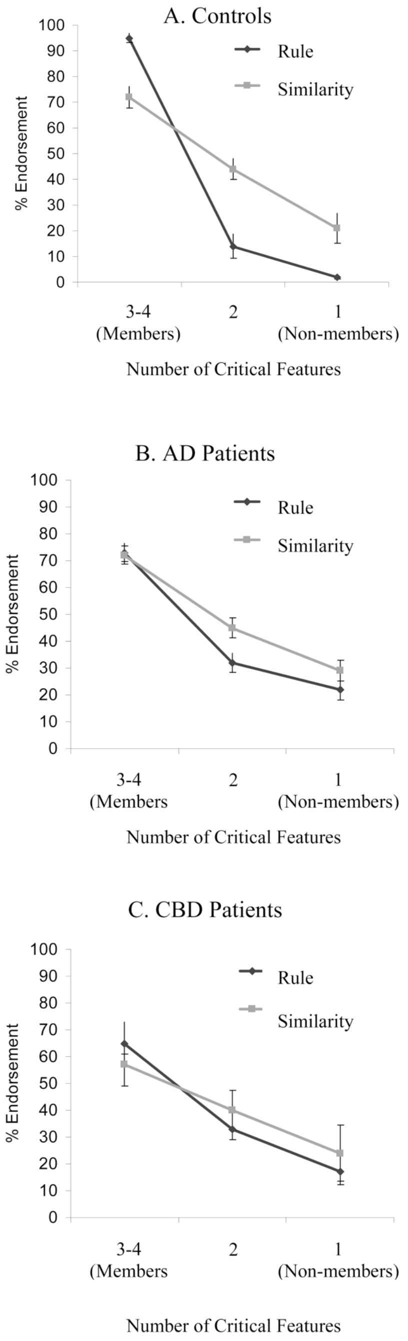


Figure 4. Koenig et al. (in press) experiment

Category endorsement functions for Rule and Similarity groups: Panel A: Controls; Panel B: AD patients; and Panel C: CBD patients.

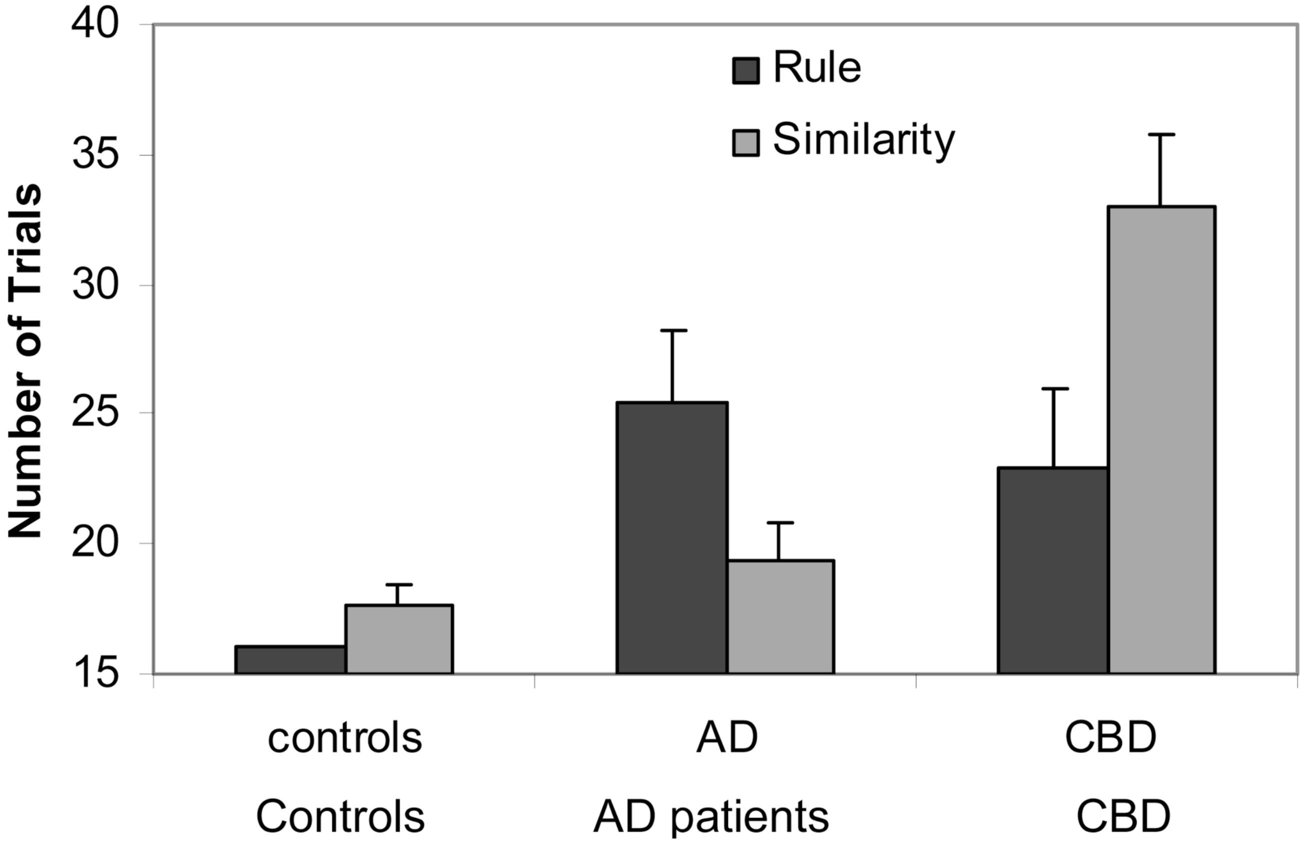
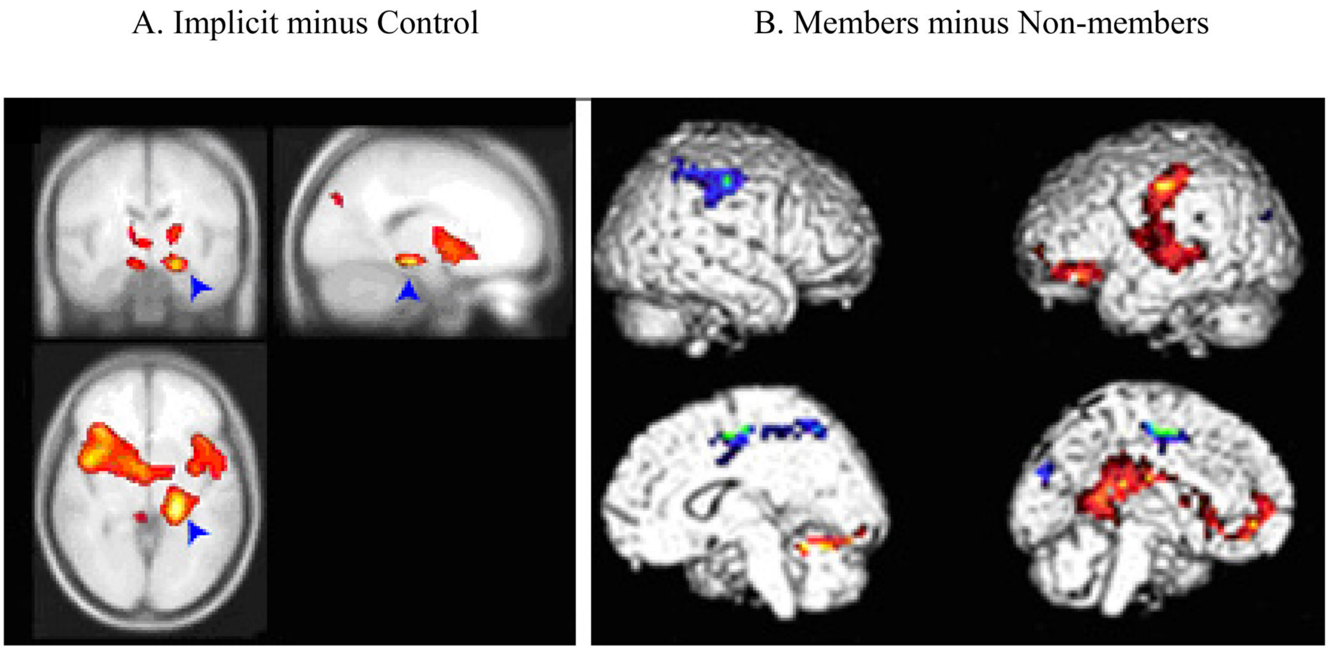


Figure 5. Koenig et al. (in press) experiment

Number of training trials needed to reach learning criteria as a function of type of participant, separately for Rule and Similarity groups.



A. Implicit minus Control

B. Members minus Non-members

Figure 6. Koenig et al. (2007) experiment, Imaging results

Panel A: Hippocampal activation during an implicit categorization task; Panel B: Posterior deactivation during the same implicit categorization task.