Can patients with Alzheimer’s disease learn a category implicitly?

Andrea Bozoki a, Murray Grossman b, Edward E. Smith c,∗

a Michigan State University, MI, USA
b University of Pennsylvania, Philadelphia, PA, USA
c Department of Psychology, 402B Schermerhorn Hall, Columbia University, 1190 Amsterdam Avenue, New York, NY 10027, USA

Accepted 1 August 2005
Available online 14 October 2005

Abstract

Can a person with a damaged medial-temporal lobe learn a category implicitly? To address this question, we compared the performance of participants with mild Alzheimer's disease (AD) to that of age-matched controls in a standard implicit learning task. In this task, participants were first presented a series of objects, then told the objects formed a category, and then had to categorize a long sequence of test items [Knowlton B. J., Squire L. R. (1993). The learning of categories: parallel brain systems for item memory and category knowledge. Science, 262, 1747–1749]. We tested the hypotheses that: (1) both Control and AD participants would show evidence for implicit learning after the unwanted contribution of learning during test is removed; (2) the degree of implicit learning is the same for AD and Control participants; (3) training with exemplars that are highly similar to an unseen prototype will lead to better implicit category learning than training with exemplars that are less similar to a prototype. With respect to the first hypothesis, we found that both AD and Control participants performed better on tests of implicit learning than could be attributed to just learning on test trials. We found no clear means for evaluating our second hypothesis, and argue that comparisons of the degree of implicit learning between patient and control groups in this paradigm are confounded by the contribution of other memory systems. In line with the third hypothesis, only training with similar exemplars resulted in significant implicit category learning for AD participants.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Implicit memory; Categorization; Aging; Alzheimer’s disease

1. Introduction

1.1. Implicit learning and the medial-temporal lobe

Learning to classify items into categories according to common or overlapping features is a fundamental cognitive task. Exposure to members of a category facilitates later categorization of similar but novel instances of that category. Presumably, this is one of the means by which children learn about the world, and how adults refine and add to their knowledge of previously learned categories. Starting in the late 1970s, many studies of category acquisition demonstrated that learning often relied on the explicit memorization of category exemplars (e.g., Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1991).

Such explicit memory however, may not be necessary for all forms of category learning. In a classic study, Knowlton and Squire (1993) presented both medial-temporal lobe amnesics and normal controls with a series of dot patterns; all the patterns were transformations of a prototype pattern, but during presentation nothing about a category was mentioned to either group of participants. After presentation, however, all participants were informed that the patterns they had just seen were instances of a category, and that they were to determine which of a sequence of test patterns also belonged to that category. Both amnesiac and Control participants performed this unexpected categorization task with above-chance accuracy, and remarkably the amnesiacs were as accurate as the controls. The conclusion drawn was that both groups of participants had learned the category implicitly rather than explicitly, where implicit memory is known not to rely on the medial-temporal lobe (e.g., Schacter, 1992). The idea of implicit category-learning was born, and added to the list of kinds of implicit learning. Since the initial experiment, sev-
eral studies have replicated the findings that medial-temporal lobe amnesics manifest likewise on implicit category learning (e.g., Kolody, 1994; Reed et al., 1999; Squire & Knowlton, 1995—see Keri, 2003 for a review).

1.2. Strengthening the case for implicit learning

But the existence of implicit category learning has been challenged by a number of related papers including Nosofsky and Zaki (1998), Palmeri and Flanery (1999) and Zaki (2004). All of these articles argue that results with medial-temporal lobe patients can be explained in terms of explicit and working-memory systems, without any appeal to an implicit system. While these papers offered a number of arguments, the most compelling is the Palmeri and Flanery (1999) demonstration that the results obtained in the implicit category learning paradigm may be entirely due to working memory. These authors did not present any training stimuli, but told participants that such items had been presented subliminally, and then gave the participants the same kind of test—a sequence of members and non-members— that is used in neuropsychological studies of implicit learning. Remarkably, the participants scored above chance. This finding indicates that category learning can occur on test trials alone, likely mediated by working-memory mechanisms that detect and amalgamate similar patterns into a category representation. This in turn implies that above-chance performance on the standard test of "implicit learning" cannot be unequivocally attributed to implicit learning, and hence that the data do not offer much evidence for category learning by implicit memory.

One main goal of the present paper is to test whether patients with medial-temporal lobe damage learn a category implicitly or by means of working memory. Toward this end, we introduced two stringent criteria of implicit category learning, criteria that assess learning when the contribution of learning-during-test has been subtracted or minimized. According to the first criterion, one needs to demonstrate that, in a standard implicit category-learning condition, performance on test trials exceeds that obtained in a condition in which no training stimuli are presented (see Keri, 2003, for a similar point). Accordingly, in the present experiment AD patients and normal controls were tested in both conditions. The main hypothesis of interest is that both Control and AD participants manifest implicit learning when a stringent criterion of learning is used.1 A stronger hypothesis is that AD patients manifest normal implicit learning even when a stringent criterion is used.1

In addition to comparing categorization performance with and without training items, another stringent test of implicit learning is to determine whether participants show above-chance performance on the categorization test before there has been sufficient chance for working-memory-based learning to occur on test trials. Accordingly, we measured categorization accuracy after the first 10 test trials (as well as at the end of test trials). Again the main hypothesis is that Control and AD participants will manifest implicit learning, and again a stronger hypothesis is that AD patients manifest normal implicit learning even when this stringent criterion is used. The obvious problem with this criterion, though, is that 10 trials may not provide sufficient statistical power to detect a smallish effect.

What precludes amnesics from using an explicit, exemplar strategy in categorization is damage to the medial-temporal lobe, particularly the hippocampal system. But surgery and encephalitis are not the only kinds of neurological damage leading to functional impairment of this system with consequent memory loss. Specifically, the defining feature of early Alzheimer’s disease (AD) is impairment of explicit, hippocampally-mediated memory due to neuronal loss from amyloid deposits and neurofibrillary inclusions. Thus, this disease provides an alternate model system for testing implicit categorization in the absence of an intact mechanism for explicit learning, and AD patients were employed in the present study. If patients with AD can implicitly learn a category it is almost certainly on the basis of extra-hippocampal circuitry, or at least involving the recruitment of additional brain regions to augment the working of the damaged medial-temporal lobe.2

1.3. Effects of similarity of training items

A second goal of the present paper is to test the hypothesis that implicit memory makes a contribution to category learning to the extent that the members of a category are similar to one another. There have been relatively few categorization studies with AD patients, and fewer still that examine dot-pattern learning or other forms of non-semantic categorization. But Keri et al. (1999, 2001) examined just this type of learning in two successive papers. In both papers, their task required participants to view sequences of dot patterns, all of which were created by systematic distortions of a prototype dot pattern; participants were then told that the patterns they had seen all belonged to a category, and that they were to indicate which of a sequence of new test patterns also belonged to this category. The results of Keri et al. (1999) suggested implicit category learning (as assessed by the standard lenient criterion) was impaired in AD patients; in particular, patients’ categorization of prototype items during test was notably impaired. Their follow-up study, in a larger cohort of 72 individuals with AD, showed "relatively spared" categorization (impairment was demonstrated only in the sub-group with moderately severe disease). Thus, although numerous researchers have found that amnesic patients perform

---

1 Knowlton and Squire (1993) and Reed et al. (1999) did try to assess the contribution of learning-during-test by instructing a separate group of control participants to "imagine" that a set of training trials had been presented, and then giving them a standard categorization test. Neither study found any evidence for category learning in the absence of training items. Presumably these failures to find learning during test were due to the fact that the "imagine" instructions did not sufficiently convince the participants that they could learn the category, in contrast to the Palmeri & Flanery cover story about subliminal presentation of training items (see Palmeri & Flanery, 1999).

2 Patients with early Alzheimer’s disease show some deficits in other memory systems as well, including working-memory and semantic memory (e.g., Percy & Hodges, 1999). Still the impairment of explicit memory is the most profound. But the impairment in working memory might reduce the ability of these patients to learn a category during test.
2. Methods

Morbidities, and CNS-acting medications. However, cholinesterase inhibitors not completing the task or the neuropsychological testing, leaving 40 AD and to the AD subjects, were recruited from the Ann Arbor, MI community via paid group of 42 healthy adult volunteers, “Controls,” who were age and sex matched with NINCDS-ADRDA criteria for the diagnosis (McKhann et al., 1984). A control with mild to moderate clinically probable AD (Folstein MMSE > 12), and were 1981, Chapter 6), but with these imaginary animals there is lit- it is unclear what exactly is learned (a prototypical pattern, a d to be matched participants according to age, there was a margin- namic animals that vary on 1 or more of 10 features (spotted versus striped body, snout versus trunk, etc.). With dot patterns, it is unclear what exactly is learned (a prototypical pattern, a description of parts of the pattern, etc.—see Smith & Medin, 1981, Chapter 6), but with these imaginary animals there is lit- tle doubt that what is learned are feature representations of the animals. Further, Reed et al. (1999) have replicated the basic neuropsychological findings with these imaginary animals, as they showed that anmiesias performed as well as normal par- ticipants on the categorization test, but substantially poorer than normals on a test of episodic memory of the training items.

2.1. Participant selection and characterization

A total of 86 participants were recruited; 44 participants were diagnosed with mild to moderate clinically probable AD (Folstein MMSE > 12), and were recruited through the Michigan Alzheimer’s Disease Research Center. Each met NINCDS-ADREA criteria for the diagnosis (McKhann et al., 1984). A control group of 42 healthy adult volunteers, “Controls,” who were age and sex matched to the AD subjects, were recruited from the Ann Arbor, MI community via paid advertisements. Four AD and two Control participants were later excluded for not completing the task or the neuropsychological testing, leaving 40 AD and 40 Control participants in the study.

All participants were screened for a history of psychos, neurological co- morbidities, and CNS-acting medications. However, cholinesterase inhibitors (memory-enhancing drugs for AD) were permitted for AD participants, and selective serotonin reuptake inhibitors (a popular and non-sedating class of anti-depressant medications) were permitted for all participants. All participants were screened for residual depression with the Geriatric Depression Scale and excluded for GDS > 11.

Each participant was administered a battery of neuropsychological tests: Barona Estimated Demographic IQ (Barona, Reynolds, & Chestain, 1984); Folstein Mini Mental Status Exam (MMSE) (Folstein, Folstein, & McHugh, 1975) to assess overall cognitive function; Non-verbal Continuous Performance Test (CPT- X and CPT-XOX) to assess simple attention and working memory (Glaser & Goodglass, 1993); nine-item California Verbal Learning Test (CVLT) to assess verbal explicit memory (Libon et al., 1996); Rey-Osterreith Complex Fig. (ROCF) to assess visuospatial explicit memory (Spreen & Strauss, 1998); Pyramid-Palm Trees test (PPT) to assess visual object processing (Howard & Patterson, 1992); Thames Valley Test Company, Bury St. Edmunds, UK; the Benton Visual Form Discrimination test (VFD) to assess visuospatial process- ing (Benton et al., 1994). Healthy older participants scoring below 25 on the MMSE, or below the 10th percentile (corrected for age and education) on any component test, were excluded from further consideration. We chose these tests in an effort to examine several areas of cognitive processing relevant to acquiring a novel category, including explicit long-term memory and working memory. Table 1 provides a comparison of the AD and Control participants on a num- ber of demographic variables. Despite our efforts to match participants according to age, there was a marginally significant difference in that variable, as well as in level of education, favoring Controls. As expected, there was a highly significant difference in mean MMSE scores, with the AD patients performing substantially lower than the Controls. All t-values are given, along with significance levels in Table 1.

Table 1: Characterization of AD and Control participants

<table>
<thead>
<tr>
<th></th>
<th>AD (n=40)</th>
<th>Controls (n=40)</th>
<th>t-value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex distribution</td>
<td>60% F</td>
<td>65% F</td>
<td>0.46</td>
<td>ns</td>
</tr>
<tr>
<td>Age</td>
<td>73.6 ± 7.9</td>
<td>70.5 ± 6.9</td>
<td>2.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Education</td>
<td>14.7 ± 3.6</td>
<td>16.1 ± 3.9</td>
<td>2.01</td>
<td>0.05</td>
</tr>
<tr>
<td>MMSE</td>
<td>26.4 ± 3.5</td>
<td>28.4 ± 1.5</td>
<td>12.96</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

2.2. Materials

AD participants were randomly assigned to receive High-Similarity training (HS), Low-Similarity training (LS), or No Training (NT). Controls were ran- domly assigned to one of these conditions after age and sex matching. A total of 10 AD and 10 Control participants were assigned to each of the HS and LS conditions, and twice that many were assigned to the NT condition (because two different test lists were used in this condition—see below). Stimuli for cat- egorization consisted of a set of novel cartoon animals used earlier by Reed et al. (1999) (see Fig. 1). To form the set, a prototype animal was created and then transformed on 1 or more of 10 attributes in a binary fashion – either the pro- totypic value (or feature) of the attribute was maintained or the non-prototypic value was substituted – for a total of 100 possible stimuli. The experiment con- sisted of three phases, each presented on a Macintosh PowerPC or Powerbook G3 computer: categorization training, categorization testing, and recognition memory. All cartoon animal images were approximately 2 in.², presented on a 15 in. monitor.

2.3. Conditions

2.3.1. High versus low similarity training conditions

During the training phase, 20 animals were each presented twice in pseudo- random order for a total of 40 trials. Each animal was presented for 3 s with a 500 ms inter-stimulus interval during which a fixation cross was presented. Total duration of training was 140 s. In the HS condition, each of the 40 animals presented had either 8 or 9 features in common with the (unseen) prototype. In the LS condition, each of the 40 animals had 6 or 7 features in common with...
Fig. 1. Examples of stimuli used in the categorization experiment. Prototype “Peggle,” example of a high-similarity (eight prototypic features) item, a low-similarity (six prototypic features) item, and a non-Peggle item (0 prototypic features).

2.3.3. Categorization test

All conditions culminated with a test phase. This phase consisted of 65 self-paced test item presentations, given 2–5 min after the completion of training (for HS and LS conditions). The test animals ran the format from 0 to 10 features as common with the prototype, presented in a pseudorandomized order that was the same for all participants within a training group. For the HS group, there were: 10 stimuli (5 new and 5 old) with 9 features in common with the prototype, 5 for each of the other animal types (e.g., 6 features in common with the prototype), including 5 presentations of the prototype itself and 5 of the anti-prototype (the animal that has zero features in common with the prototype). For the LS group, the pattern was the same, except that now there were 10 stimuli (5 new and 5 old) with items that had 6 and 7 features in common with the prototype, and 5 stimuli of each of the other animal types. The order of item presentation also differed between HS and LS groups. These two test lists were also used in the NT condition. We used different test lists for HS and LS conditions, and for the NT condition, because: (1) we wanted to include more tests of high-similarity items in the HS condition (namely old ones as well as new ones), and more tests of low-similarity items in the LS condition (old as well as new ones); (2) when comparing the HS (LS) and NT condition, we wanted the test lists to be identical. (As it turned out, though, there was no difference between the old and new test items.) Stimuli were self-paced, and participants in the HS and LS conditions were instructed as follows (the instructions for the NT condition have already been described):

All of the animals you just saw belong to the category “Peggles.” You will now see a new bunch of animals. About half will be Peggles and about half will not. Press the ‘Y’ button if you think the animal is a Peggle, and the ‘N’ button if you think it is not. Try to go with your first impression.3

Categorization data were scored according to how often each type of test animal was endorsed—judged to be a member of the category Peggle. Categorization accuracy (CatAcc) was calculated as the percentage of all trials in which the participant endorsed an item with six or more features in common with the prototype, or rejected an item with less than five prototype features. Items with exactly five prototype features were dropped from this particular calculation as it was felt to represent an indeterminate condition. Six features was chosen as the cutoff for category membership, to maintain consistency with our training, during which low-similarity items were defined as category members and had either six or seven features in common with the prototype. We also calculated an accuracy score based on just the first 10 trials (CatAcc10) for each participant, using the same criteria as those for CatAcc. Recall that this score could be used as a second stringent criterion for implicit category learning.

Another measure of category learning was the slope of the function relating endorsement rate to number of features in common with the prototype (typicality)—the steeper the slope, the better the category has been learned.

3 Because different AD patients were tested in the NT conditions, as well as in the HS and LS conditions, it is important to establish that there were no demographic or cognitive differences between these four groups of patients. Separate analysis of variance were performed on each of the five demographic-cognitive measures listed in Table 1. Not one of the five measure showed a significant difference across patient groups (the maximum F achieved was F(3, 36) = 1.44, ns).
We defined endorsement rate (CatEnd) as the percentage of trials at a given level of typicality – e.g., 6 or 7 features in common with the prototype – that an item was designated as a category member. In order to improve our power to detect differences between groups, we collapsed each pair of successive typicality levels – e.g., 6 or 7 features in common with the prototype – into a single level, leaving the prototype (all 10 features in common with the prototype) as its own level. We therefore ended up with six levels of typicality.

2.3.4. Recognition test
As a check on their ability to explicitly recognize individual exemplar animals, participants in the HS and LS conditions underwent a recognition task, immediately after concluding the categorization task. For expediency, the NT group did not perform this task. In the recognition task, a second set of cartoon animals, not previously seen by the participants, were presented during a study phase. These images looked qualitatively different from the items used during the categorization task (that is, they were not just different exemplars from the Puggle group of animals). However, other parameters were the same as in the earlier categorization task (i.e., 10 binary attributes with a prototype consisting of a particular value of each attribute). A set of 20 animals were presented twice each in a pseudo-random order for a total of 40 trials. Each animal was presented for 3 s with a 500 ms inter-stimulus interval during which a fixation cross was presented. Total duration of presentation was 140 s. In the HS condition all animals presented had either eight or nine features in common with the (unseen) prototype; in the LS condition all animals had six or seven features in common with the prototype.

Prior to presentation, participants were instructed:

You will see a new bunch of animal cartoons now. Look at each animal as it appears on the screen and try to memorize its appearance. You will be tested on your ability to recognize them soon.

Immediately following the presentations, participants were given an explicit recognition test. To keep procedures parallel to the categorization task, 15 new and 10 previously shown items were presented during the test phase. These animals were not given a category name. Participants were instructed:

You will now see some more animals. Some of these animals will be identical to the ones you just saw but most of them will be new. Press the ‘Y’ button if you think the animal is one you just saw, and ‘N’ if you think it is new.

Because there was a tendency on the part of all participants to make their memory judgments on the basis of features shared with the (unseen) prototype (that is, to respond as they did in the categorization task), recognition accuracy was assessed only by comparing new and old items with the same degree of typicality. That is, we compared accuracy of correct identification and correct rejection of stimuli with, say, eight features in common with the prototype (5 new versus 5 old).

3. Results and discussion
In what follows, we first discuss the results for overall categorization accuracy, then consider the endorsement-rate measure of category learning, and finally describe the results for the explicit memory task.

3.1. Categorization accuracy
Consider the categorization accuracy for the initial 10 test trials — CatAcc10 scores. These scores provide an indication of whether Control or AD participants in the HS and LS conditions manifested any implicit learning before learning during-test had much time to manifest itself. For the HS condition, CatAcc10 scores for Controls and ADs are 79% and 60%, respectively; only the result for Controls is significantly greater than a chance score of 50% (p < 0.05). Furthermore, the score for Controls

is significantly greater than that for AD patients (t (18) = 2.53, p = 0.02). For the LS condition, the CatAcc10 scores for ADs and Controls were 53% and 67%; the results for Controls is significantly greater than chance (t (9) = 3.26, p < 0.01). For both HS and LS conditions, then, we have stringent evidence for implicit category learning, but only in Controls. There is no evidence for implicit category learning in AD patients by this measure.4

Table 2 presents overall categorization accuracy – CatAcc scores – and leads to a different conclusion. The scores are presented for Control and AD participants in the HS, LS, and NT conditions (with scores for the NT condition being presented separately for the two different test lists). Starting with the results for the HS condition (top half of Table 2), note first that we have replicated the standard finding of comparable categorization accuracy for Control and Patient participants (76% and 73%, respectively). Of greater interest is our second stringent measure of implicit learning, the difference between performance with and without training items, HS-NT. There appears to be evidence for implicit learning, particularly in AD participants (HS-NT = 14%). These scores were submitted to a two-factor analysis of variance (diagnosis—AD versus Control; condition—HS versus NT). There was a main effect of diagnosis, Controls outperforming AD participants (F (1, 36) = 4.15, p = 0.05), but the training effect apparent in Table 2 failed to reach significance (F (1, 36) = 2.27, ns) as did the interaction effect (F (1, 36) = 1.34, ns). In view of this, and given that the contrast between HS and NT is central to the present study, we tested for the training effect by directly comparing the HS and NT conditions for each group of participants. For the AD patients there is a significant difference (t (18) = 2.7, p < 0.02); for the Control participants the difference is in the right direction but is not significant (t (18) = 1.0, ns).

The results for the LS condition are in the bottom half of Table 2. Our stringent measure of LS-NT offers no evidence for implicit category learning in either group of participants. This difference from the HS results is not due to a variation in the NT conditions (see Table 2), as the two different test lists led to comparable results (for the AD patients, t (18) = 1.1, ns; for the Controls, t (18) = 1.0, ns).

Table 2 provides evidence for implicit learning in AD patients.
for the Controls, \( t(18) = 0.00 \). Rather the difference between the two sets of results in Table 2 is that HS training resulted in substantially greater learning than LS training. A two-way ANOVA – AD versus Control X type of training – was applied to the CatAcc scores. There was a beneficial effect of type of training – HS better than LS (\( F(1, 36) = 7.13, p < 0.01 \)) – but no effect of group, as both AD and Control participants performed comparably. Based on our second stringent criteria, then, there is clear evidence for implicit category learning, but only when the training items are highly similar.

Thus, our first stringent test (CatAcc10 > 50%) provides evidence that Controls manifest implicit learning, whereas our second stringent test (CatAcc difference between HS and NT) provides evidence for implicit learning in AD participants. Why this asymmetry in outcomes? There are two questions here. (1) Why did the AD patients not pass our first stringent criterion (at least in the HS condition)? (2) Why did the Controls not show more evidence for implicit learning by our second stringent criterion (at least in the HS condition)? With regard to question (1), one possibility is that the test has insufficient power to obtain significance for the AD patients (note that their CatAcc10 score, 0.60, was in the right direction). A CatAcc10 score for a particular condition was based on a total of only 100 observations (10 trials \( \times \) 10 participants), whereas a CatAcc score for a condition was based on 650 observations (65 trials \( \times \) 10 participants). In addition, the CatAcc scores may have showed more evidence for implicit learning than the CatAcc10 scores because the AD patients continued to learn implicitly over the course of the 65 test trials (25 of the 65 test trials in the HS test list contained items with at least 8 features in common with the prototype). With regard to question (2), our Controls may not have shown more evidence for implicit learning by our second stringent test because of a “strategic” effect. Controls may have chosen to rely less on learning during the test in the HS than the NT condition because they had relatively little need of learning during the test in the HS condition—their implicit learning was sufficient. Yet we are assuming “equal need” by subtracting the NT score from the HS score. Hence, we may have “subtracted too much” from the HS scores for Controls.

### 3.2. Learning during test

Do both groups of participants show learning during test, and if so, in comparable amounts? Table 3 presents the relevant data. The data are drawn from only NT trials (collapsed over the HS and LS test lists, to increase power), and compare accuracy for the first 10 trials, CatAcc10, to accuracy for all trials, CatAcc, separately for Control and AD participants. It appears that both groups learned during test trials, as both groups show performance increases with trials, and these increases are comparable. An analysis of variance cannot be performed on these data because CatAcc and CatAcc10 scores are not independent. But a direct comparison of CatAcc to CatAcc10 shows significant effects for both AD participants (\( t(18) = 5.58, p = 0.001 \)), and Control participants (\( t(18) = 5.89, p = 0.001 \). (Although the performance increases are comparable for the two groups, there is suggestive evidence that overall performance is better for Controls than AD participants—\( t(18) = 1.82, p < 0.10 \)).

#### 3.3. Category endorsement

Our other categorization measure is category endorsement (CatEnd), the slope of the function relating the percentage of times an item is endorsed as a member of the category to the number of features it shares with the prototype. The functions for the AD and Control participants in the HS and NT conditions are shown in Fig. 2. Note that these functions provide more detailed information than our CatAcc and CatAcc10 measures; the functions tell us whether categorization accuracy increases monotonically with the similarity of the test item to the category prototype, and how sensitively tuned categorization is to similarity (the steepness of the function). Category endorsement functions are routinely treated as the most sensitive measures of implicit category learning (e.g., Knowlton & Squire, 1993; Palmeri & Flanery, 1999), and a comparison of the functions in Fig. 2 offers our most detailed stringent test for implicit category learning.

Consider first the contrast between HS and NT for AD participants in Fig. 2A. The results present support for implicit category learning in AD patients. At every one of the six points on the \( x \)-axis, training led to better performance than no training (remember: good performance for the points corresponding to 0–1 and 2–3 features requires not endorsing the item). We used regression analysis to calculate a best-fitting line for each group and condition, and then compared slopes with \( t \)-tests. The difference between the slopes of the trained and untrained AD groups was significant (\( t(8) = 2.63, p < 0.05 \)), though there was no such evidence in the equivalent comparison in the Controls (\( t(8) = 1.18, p = 0.37 \)). It is also worth noting that while the slope for the AD participants in the NT condition leaves little doubt that the patients are capable of learning during test, overall they learned less during test than did the Controls: the NT slope for the Controls was significantly greater than that for the AD participants (\( t(8) = 5.72, p < 0.01 \)). There was no comparable difference in slopes between Controls and AD participants for the HS condition (\( t(8) = 0.67, p = 0.51 \)), which indicates that there is no evidence for superiority of Controls over AD in implicit learning.

Fig. 3 compares endorsement rates in the HS and LS conditions, separately for AD (Panel A) and Control participants (Panel B). The slope was steeper in the HS than the LS condition for both AD and Control participants; the difference in slopes was significant for both AD patients (\( t(8) = 2.26, p < 0.05 \)) and Controls (\( t(8) = 3.5, p < 0.01 \)). However, there was no difference
Fig. 2. Endorsement ratings for test items ("Yes, it belongs to the category") as a function of the typicality of the test item (i.e., the number of features the test item shares with the prototype), separately for the HS and NT conditions. The functions in 2A are for AD patients, the functions in 2B are for Controls.

Fig. 3. Endorsement ratings for test items ("yes it belongs to the category") as a function of the typicality of the test item (i.e., the number of features the test item shares with the prototype), separately for the HS and LS conditions. The functions in 3A are for AD participants, the functions in 3B are for Control participants.

3.4. Accuracy for episodic memory (recognition task)

Neither AD nor Control participants were significantly different from chance (50%) in their ability to determine old from new test items. For the HS condition, recognition accuracies were 39% and 52% for AD and Control participants, respectively. For the LS condition, recognition accuracies were 36% and 44% for AD and Control participants, respectively. When adjusted for degree of similarity to the prototype, there was no difference between old and new items, for either HS or LS participants. That is, there was no difference in accuracy for new versus old items at each level of typicality in which both new and old items were presented at test (six and seven features for LS; eight and nine features for HS). Rather, both groups of participants seemed to base their responses on the similarity of the test item to the prototype (as in the categorization task). The resulting floor effect for explicit memory makes further analysis of these data of little use.

To determine whether there was a dissociation between explicit retrieval and the mechanisms subserving categorization, we examined the neuropsychological data for both verbal (CVLT total recall score) and visuospatial (Rey-Osterreith 5 min delay) episodic memory tasks, collected concurrently with the categorization experiment (see Table 4). These data served as our indicators of explicit memory. As might be expected, Controls far outperformed AD participants in both verbal and visuospatial episodic memory function. There were also significant differences in the other tested cognitive domains—simple attention...
4. General discussion

As outlined in the introduction, we set out to explore several questions. First, and most important, are AD patients and matched Controls able to implicitly learn novel categories, when stringent criteria for learning are used to rule out the contribution of working memory? Second, assuming the answer to the first question is positive, is implicit learning in AD participants equivalent to that in Control participants? Third, are AD participants able to learn a novel category solely through learning on the test trials, and if so, is this learning comparable to that in Control participants? Fourth, does the similarity of training instances matter? That is, do individuals learn novel categories better if the exemplars are closely related to one another than if they are more diverse in their features? We start by briefly discussing the last question, then move on to a more extended discussion of the first three questions.

4.1. What are the effects of similarity of training?

Consider first the effects of similarity of training with Control participants. When the measure of performance was accuracy on the first 10 test trials, Controls showed above-chance accuracy with both high- and low-similarity training items, but with somewhat greater accuracy in the high- than the low-similarity condition (79% versus 67%). When the measure of performance was accuracy on all test trials with versus without training (HS–NT or LS–NT), Controls showed no evidence for implicit category learning with either kind of training, but their accuracy with training was greater for high- than low-similarity items (see Table 2). Thus, even in the latter case, there is evidence that high-similarity training leads to better implicit learning. The safest conclusion seems to be that, though Control participants can learn a category implicitly even with low-similarity training item, they may learn more with high-similarity items.

The results are more clear-cut for AD participants. They showed evidence for implicit learning only when the training items were similar. This result raises the possibility that Keri et al. (1999) finding of impaired implicit-category-learning in AD participants may have been due to their use of relatively dissimilar training items, i.e., Keri et al. used only high-distortions of the prototype dot pattern. However, it is difficult to compare our variation in similarity among artificial animals with their variation in similarity of dot patterns. What is clear, though, is that in these kinds of paradigms implicit learning occurs to the extent the exemplars are similar; in our study we obtained evidence for implicit learning in AD participants only with high-similarity training items.

A cautionary note is in order, though, in interpreting any of these findings as showing less implicit learning with low-similarity training items. It is possible that any lack of evidence for learning – e.g., the fact that the difference between LS and NT was not significant for either group of participants – was due to a strategic effect. Specifically, in the LS condition, participants could have implicitly learned something about the category structure, but what they learned may have been sufficiently impoverished that they chose to ignore it and rely instead on what they learned during test. This possibility of strategic choice arises because the paradigm permits multiple ways for category learning to occur, which means that participants had the option of choosing to emphasize one way over another.

4.2. Do AD and Control participants show evidence for implicit learning with stringent criteria?

This is the major question motivating our research. In addressing the question, we focus on the high-similarity condition. Using the standard criterion for implicit category learning – overall categorization accuracy on the test trials following training – both groups of participants showed evidence for implicit category learning. Categorization accuracy was 76% for Controls and 73% for AD participants; both numbers are significantly greater than chance, and there is no significant difference between them. But, as noted at the outset, this standard criterion is too weak given the Palmeri and Flanery (1999) demonstration that category learning can occur during test trials.

One stringent criterion for implicit learning was that participants showed above-chance accuracy on the first 10 trials. Controls passed this criterion; but AD patients did not, presumably because of the insensitivity of the measure or because AD patients needed additional implicit learning during test to raise their categorization accuracy. Our second stringent criterion for learning was the overall difference between categorization performance when it was preceded by actual training trials and when it was not. Under this criterion, AD participants showed evidence for implicit category learning. To our knowledge, this is the first demonstration of implicit category learning using a demanding criterion like that routinely applied in studies of perceptual priming (e.g., in the standard fragment-completion task, implicit memory is assumed only when performance is better on primed items than on unprimed items, e.g., Schacter, 1992). Control participants did not pass this stringent criterion, presumably because they chose not to use learning-during-test when they had already learned something substantial from the training; this
strategic effect undermined our subtracting learning-during-test from their overall CatAcc scores. Hence the results based on our two stringent criteria are somewhat mixed. While we have offered explanations of why AD patients failed to pass the first stringent criterion and Controls the second, we have no independent support for these accounts. But there is a third source of evidence—the functions relating the probability of endorsing a test item as a category member to the typicality of that test item. The slope of such a function offers a detailed measure of the degree of learning, a measure that goes beyond our CatAcc scores. For AD patients, the slope of the function for the HS condition significantly exceeded that of the NT condition (see Fig. 2). This is the strongest evidence we have for implicit category learning in patients with medial-temporal lobe damage. Still, the fact that our results are somewhat mixed means that the issue is still open.

The preceding evidence has implications for the larger issue about whether all category learning can be accounted for in terms of just episodic and working memory (e.g., Nosofsky & Zaki, 1998; Palmeri & Flanery, 1999), or whether an implicit learning system needs to be posited as well. The categorization performance of normal controls in paradigms like the one we used cannot offer strong evidence about this larger issue. This is because control performance could be based on episodic memory of the training items (even though they were learned unintentionally). Stronger evidence about the issue is provided by the categorization performance of patients with medial-temporal lobe damage. Patient performance is much less likely to be due to explicit memory of the training items, and if working memory can also be ruled out then categorization performance must be due to another memory system—implicit memory. Thus, to the extent our results demonstrate accurate categorization by AD patients, such results offer evidence for the existence of implicit category learning.

4.3. Is AD implicit-category learning equal to that of the Controls?

This turns out to be difficult to answer, given that the two groups passed different criteria for learning. It appears that these two groups used a different mix of (at least) two different mechanisms to achieve comparable overall accuracy scores on the contrast between HS and NT. Specifically, in the HS condition AD participants may have relied on both implicit learning during training and some learning during test, whereas Controls seem to have relied mainly on implicit learning.

Even this view may be too simple, as the Controls may also have relied on their explicit memory of the training instances. Although studies of implicit category learning routinely assume that they have eliminated this possibility by disguising the training trials as something that would block intentional learning, and subsequent conscious retrieval, there is no direct evidence for this assumption. Indeed there is now evidence that the assumption of no-explicit-learning is mistaken. First, in a recent study of priming in a fragment-completion task in normals, May, Hasher, & Foong (2005) interviewed their participants after the study and found that a sizeable number of them were aware of the relation between the learning items and the test items, even though such a connection was obscured. This priming paradigm is typically assumed to be free of intentional learning. Second, in an fMRI study using the standard implicit-category-learning paradigm with normals, Reber, Stark, & Squire (1998) found increased activation in anterior frontal areas during the categorization test; these areas are routinely associated with intentional retrieval from episodic memory (Lepage, Ghaifar, Nyberg, & Tulving, 2000). In sum, for normal participants, the standard paradigm for assessing implicit category learning may involve all three major memory systems—implicit and explicit long-term memory, as well as working memory (for learning during test trials).

4.4. Are AD and Control participants comparable in learning during test?

The answer to this question is a qualified “yes.” Without training, AD participants performance on the categorization test improved from 54% on the first 10 trials to 62% on overall accuracy at the end of 65 trials. Similarly, the Controls improved from 60% to 71%. There was no statistical difference between these estimates of improvement (though the Controls started at a higher level).

This kind of learning has been attributed to a working-memory mechanism (Palmeri & Flanery, 1999). How might this mechanism operate? In addressing this question, it is useful to have an example at hand, and accordingly Fig. 4 provides the first 10 test items in the HS test list. One possibility is that after seeing the first few items, participants notice that successive items share features, and then select one or two as the ones defining of category membership (e.g., striped body); they then use this feature (or features) to determine their subsequent categorization decisions. This is an hypothesis-testing strategy, a strategy that is known to rely heavily on working memory (Ashby & O’Brien, 2005).

This interpretation has some drawbacks, though. There is no evidence that participants use a single feature to guide their categorization decisions in this task (Reed et al., 1999). And even young normal participants typically do not spontaneously generate hypotheses that involve more than one feature (e.g., Ashby & O’Brien, 2005). Furthermore, there is a more general problem with any hypothesis that attributes learning-during-test to working memory: we have shown that AD patients are capable of such learning, but AD patients are known to have some impairment in working memory (e.g., Perry & Hodges, 1999). Such an impairment would explain why Controls outperformed AD patients on the first 10 trials of test (61% versus 54%), but the impairment is not in line with the finding that AD and Control participants improved over test trials at the same rate.

If working memory is not the sole mechanism underlying learning-during-test, what other memory system is involved? It cannot be explicit long-term memory for both groups of participants, because this system is severely impaired in AD. This leaves us with implicit long-term memory, but this seems unlikely for Controls because during the test trials they would likely have attempted to learn the category explicitly as they knew there was a category present. We are left then with
the (unparsimonious) hypothesis that Control and AD participants may have used different mechanisms in learning during test—primarily working memory for Controls, and possibly a combination of working memory and implicit learning for AD patients. Note that this hypothesis also offers another reason why Control participants learned more than their AD counterparts on the first 10 test trials (the CatAcc10 scores). Only the Controls could make extensive use of working memory, and working memory alone could accomplish some learning during the first 10 test trials without training (see Fig. 4). If this hypothesis is correct, then AD and Control participants used different learning mechanisms, or combinations of mechanisms, during test as well as during training.

5. Study limitations

There are a number of limitations of our study that prevent us from drawing strong conclusions about certain questions. We note five such limitations in what follows.

5.1. Learning during test

We used the standard implicit category learning task, which includes a long series of uninterrupted test trials. Researchers have been aware from the start that there was a possibility of learning during test trials (see Knowlton & Squire, 1993), but Palmeri and Flanery (1999) were the first to demonstrate that,
and the difference in training – HS versus LS versus NT – resulted in there being only 10 participants in most of the critical comparisons. (The difference between participant groups and the difference in training – HS versus LS versus NT – are both necessarily between-subject variations.) Note that doubling the sample size would have required testing a total of 80 AD patients, a very time-consuming endeavor. Given our limited sample size, some of our statistical comparisons lacked power and resulted in only borderline significance. Furthermore, the lack of power limited our ability to treat the CatAcc10 scores as definitive. (Recall that the fact that such scores did not exceed chance for AD patients may have been due to a lack of power.) Further experiments with AD patients and matched Controls will be needed to increase one’s confidence in the present results.

5.4. Medications
The use of cholinesterase inhibitors by a large subset of our AD participants could conceivably be a confound in our findings, contributing to an artificial boost in the performance of AD participants. However, from concomitant neuropsychological profiling, it is clear that these AD participants retain a profound deficit of explicit episodic memory for both verbal and visuospatial stimuli. Although it is possible that cholinesterase inhibitors exert a selectively greater effect on implicit memory mechanisms, there is no evidence for this in the neuropsychological literature to date.

6. Conclusions
The most striking aspect of these results is that our AD participants showed some evidence for implicit-learning, and for learning-during-test. Individuals with AD can perform this type of categorization task, with or without training; but the disease burden may affect the mechanisms by which the task is preferentially accomplished—AD participants may use a different combination of implicit memory and working memory than Controls do. If true, this would be an example of an adaptive compensatory strategy occurring in the setting of a progressive and devastating disease, enabling patients to maintain some aspects of cognitive performance within the normal range for a longer period of time.

Acknowledgements
We would like to thank Nathalie James for her significant contributions to data collection and statistical analysis for this project. Her efforts were integral to the final paper. We also thank Phyllis Koenig for her helpful comments on a recent version of the manuscript. We are particularly indebted to the anonymous reviewers of an earlier version of this manuscript, whose insightful and incisive comments led us to reconceptualize our research, which in turn led us to different conclusions than we had initially drawn. This research was supported by NIH grants: AG08671 (Michigan Alzheimer’s Disease Research Center), AG08808 (Pepper Geriatrics Research and Training Grant), and P32-AG00114 (U. of M. Institute of Gerontology Training Grant) and R01-036827, and AG15116. This research was carried out at the University of Michigan.
References


