Category Learning in Alzheimer's Disease and Normal Cognitive Aging Depends on Initial Experience of Feature Variability

Jeffrey S. Phillips, Ph.D.[a,*], **Corey T. McMillan, Ph.D.**[a], **Edward E. Smith, Ph.D.**[b], and **Murray Grossman, M.D., Ed.D.**[a]

[a]Penn Frontotemporal Degeneration Center, Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania

[b]Department of Psychology (deceased), Columbia University, New York, New York

Abstract

Semantic category learning is dependent upon several factors, including the nature of the learning task, as well as individual differences in the quality and heterogeneity of exemplars that an individual encounters during learning. We trained healthy older adults (n=39) and individuals with a diagnosis of Alzheimer's disease or Mild Cognitive Impairment (n=44) to recognize instances of a fictitious animal, a "crutter". Each stimulus item contained 10 visual features (e.g., color, tail shape) which took one of two values for each feature (e.g., yellow/red, curly/straight tails). Participants were presented with a series of items (learning phase) and were either told the items belonged to a semantic category (explicit condition) or were told to think about the appearance of the items (implicit condition). Half of participants saw learning items with higher similarity to an unseen prototype (high typicality learning set), and thus lower between-item variability in their constituent features; the other half learned from items with lower typicality (low typicality learning set) and higher between-item feature variability. After the learning phase, participants were presented with test items one at a time that varied in the number of typical features from 0 (antitype) to 10 (prototype). We examined between-subjects factors of learning set (lower or higher typicality), instruction type (explicit or implicit), and group (patients vs. elderly control). Learning in controls was aided by higher learning set typicality: while controls in both learning set groups demonstrated significant learning, those exposed to a high-typicality learning set appeared to develop a prototype that helped guide their category membership judgments. Overall, patients demonstrated more difficulty with category learning than elderly controls. Patients exposed to the higher-typicality learning set were sensitive to the typical features of the category and discriminated between the most and least typical test items, although less reliably than controls. In contrast, patients exposed to the low-typicality learning set showed no evidence of learning. Analysis of structural imaging data indicated a positive association between left hippocampal grey matter density in elderly controls but a negative association in the patient group, suggesting differential reliance on hippocampally-mediated learning. Contrary to hypotheses, learning did not

---

*Corresponding author: jefphi@mail.med.upenn.edu.

differ between explicit and implicit conditions for either group. Results demonstrate that category learning is improved when learning materials are highly similar to the prototype.

Graphical Abstract
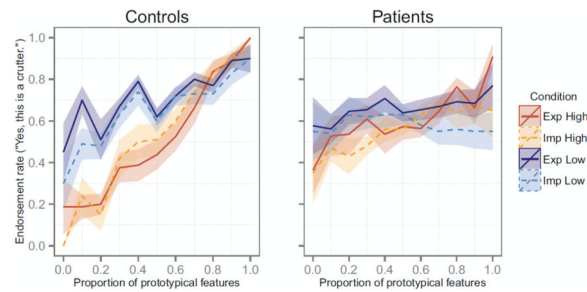
Keywords

category learning; prototype extraction; Alzheimer's disease; semantic memory; episodic memory; hippocampus; medial temporal lobes; neurodegenerative disease

1. Introduction

Category knowledge is fundamental to the way that humans perceive, think about, and interact with the world. Furthermore, the ability to learn new and ad hoc categories throughout the lifespan is one indicator of healthy cognitive aging. In age-related diseases such as Mild Cognitive Impairment (MCI) and mild Alzheimer's disease (AD), patients exhibit declarative memory difficulty despite minimal deficits in other cognitive domains (DuBois et al., 2007; Mickes et al., 2007). Patients with age-related disease nevertheless retain a limited capacity for novel category learning (Bozoki et al., 2006; Koenig et al., 2008). These findings hint at the potential to leverage patients' residual category knowledge and learning capacity in novel coping strategies or cognitive interventions. Clarifying the mechanisms behind older adults' successful category learning is critical to developing efficient learning protocols, both in neurodegenerative disease and cognitively normal aging.

Several task- and stimulus-centered factors may influence the speed, robustness, and transferability of category learning (Koenig et al., 2007; Maddox et al., 2010; Zeithamova et al., 2008). Such factors may include the quality of the exemplars that an individual encounters when initially learning a category. In real-world situations, for example, individuals' experiences with a given class of items (for example, animals or fruits) may vary according to their background—in particular, some participants may encounter greater variability among the features of category members than others. We hypothesized that the similarity of learning items to an unseen prototype (and thus, to other learning items) would affect learners' ability to extract the central tendency of a novel category. Although several studies have reported successful category learning in AD (Bozoki et al., 2006; Heindel et al., 2013) and MCI (Nosofsky et al., 2012), the effect of experiential factors, such as feature variability, on patients' learning has received little attention.

Furthermore, evidence from functional magnetic resonance imaging (fMRI; Reber et al., 2003; Smith and Grossman, 2008; Zeithamova et al., 2008), computational modeling (Ashby et al., 2011; Kéri et al., 2002; Nomura and Reber, 2008, 2012), and patient-based studies (Heindel et al., 2013) suggests that successful category learning may be supported by multiple neurocognitive mechanisms (Poldrack and Foerde, 2008; but see Nosofsky et al., 2012). Explicit mechanisms include rule-based learning, which taxes working memory and executive function (Grossman et al., 2013a; Nomura et al., 2007), as well as similarity-based learning, which involves episodic memory for previously studied exemplars (Koenig et al., 2005, 2007). Explicit learning is rapid and transferable to novel contexts (Reber et al., 1996), but it may also be sub-optimal for learning categories defined by complex or non-verbal features (Ell and Ashby, 2006). Patients with AD and MCI exhibit marked atrophy in the hippocampus, which may prevent them from using declarative learning approaches (Gifford et al., 2015; Libon et al., 1998). Implicit mechanisms include perceptual learning of stimulus features, where modality-specific cortical representations may be modified (Folstein et al., 2012, 2013; Xu et al., 2013), as well as procedural learning of stimulus-response associations through striatally-mediated dopaminergic signaling (Ashby and Maddox, 2011; de Vries et al., 2010; Shohamy et al., 2008). Implicit learning may be relatively spared in MCI and AD (Knowlton and Squire, 1993). Individuals with AD and MCI exhibit relatively intact learning on tasks that are thought to involve procedural learning, such as perceptual-motor sequence learning (Gobel et al., 2013) and probabilistic classification (Eldridge et al., 2002). This procedural learning may be related to relatively preserved striatal anatomy in AD. Perceptual representations in temporo-occipital cortex (TOC) are also spared relative to the hippocampus and MTL (Frisoni et al., 2007) and are thought to support visual semantic representations (Grossman et al., 2013b, Peelle et al., 2014), suggesting modification of TOC representations as a possible mechanism for preserved learning of perceptual feature knowledge in AD (Kéri et al., 2001).

In the current study, we used a prototype extraction task (Smith and Grossman, 2008) to investigate category learning differences between older adults with normal cognition and those with a clinical diagnosis of AD or MCI. Participants were trained to identify members of a novel visual category, the "crutter" (Koenig et al., 2005), while rejecting non-crutters. Learning conditions varied in learning items' similarity to an unseen category prototype (higher or lower) and the nature of instructions given to participants (promoting either explicit or implicit learning). The target category was probabilistically defined by the combination of 10 discrete visual features, and learning was modeled as a process of assigning subjective decision weights to features based on their perceived association with the target category. Overt categorization decisions were assumed to represent a mathematical integration of decision weights (i.e., by averaging) to produce an overall estimate of a test item's likelihood of belonging to the target category. We note that such a feature-integration framework underlies several computational models of category learning, including those which capture individual differences in learning mechanisms (e.g., Ashby et al., 2011; Love et al., 2004).

We predicted that cognitively normal controls and patients with AD or MCI would demonstrate significant category learning. Controls were expected to demonstrate superior learning, endorsing more typical items and rejecting less typical items at higher rates than

patients. Learning was expected to be more efficient for controls and patients alike when they learned from items that were highly similar to the category prototype, with little between-item variability, allowing them to more easily discover the prototype. In contrast, we predicted that individuals who encountered learning items with lower typicality (and thus greater between-item feature variability) were likely to learn less efficiently, even when learning items were derived from the same prototype as the set of higher-typicality exemplars. We further predicted that controls would learn under both explicit and implicit instruction conditions, and that learning set typicality effects would be reduced among controls who used an explicit learning approach, as their intact declarative memory would help them to overcome feature variability and learn the category prototype. In contrast, we predicted that patients would exhibit impaired learning in the explicit condition but relatively preserved implicit learning.

2. Methods

2.1. Participants

A total of 93 older adults with normal cognition ("controls") or a clinical diagnosis of Alzheimer's disease (AD) spectrum disorder participated after giving informed consent according to the guidelines outlined by the Institutional Review Board of the University of Pennsylvania. Exclusion criteria included any history of psychiatric illness, including major depression, or other neurological conditions such as stroke or hydrocephalus. Patients were recruited through the University of Pennsylvania Cognitive Neurology Clinic and were diagnosed in clinical consensus meetings with either MCI or AD. AD patients met McKhann et al.'s (2011) criteria for probable AD, while MCI patients met the core clinical criteria of Albert et al. (2011). Data from 10 participants (2 control, 3 MCI, 5 AD) were excluded due to non-compliance with instructions on the behavioral task: 9 of these participants endorsed all test items at a rate of 95% or more, indicating a likely misunderstanding of task instructions, while 1 AD participant had a high non-response rate coupled with very long response times. These exclusions left a total of 39 controls and 44 patients. The patient group comprised 37 individuals with AD and 7 individuals with MCI; patients with MCI were distributed evenly across experimental conditions. Post-hoc analyses of behavioral effects indicated no reliable difference in category learning performance between patients with AD and MCI, and no reliable association between patients' learning and measures of global cognition (Section 3.1). Neuroimaging analyses involved the subset of participants (control = 14; AD = 13; MCI = 2) that had anatomical image data collected within 6 months of the behavioral experiment. Post-hoc analyses confirmed that participants in the neuroimaging sample displayed the same pattern of behavioral effects as in the full behavioral dataset (Section 3.5).

Table 1 provides means and standard deviations of several participant characteristics by experimental condition. Mean MMSE score in the patient group (mean=20.3, s.d.=4.9) was significantly lower than in the control group (mean=28.8, s.d.=0.8) [$t(69)$=-9.1, $p<0.001$***], reflecting patients' global cognitive impairment. Controls were screened to verify their negative neurological history (e.g., MMSE > 27 or self-report) and negative psychiatric history (e.g., no depression or substance abuse). Patients were administered the

MMSE at a mean interval of 115.5 (123.6) days from the experimental session; MMSE data were available for 29 of 39 controls, at a mean interval of 42.5 (108.5) days. MMSE did not differ by learning set or instruction type [both F(1,63) < 1.3, p>0.2], and in direct contrasts patients' MMSE scores did not differ by learning set [t(40)=-1.2, p<0.23] or instruction type [t(40)=0.8, p<0.44]. Education level did not differ between groups, learning sets, or instruction types [F(1,75) < 0.8, p>0.3 for all main effects and interactions]. No effects of age were found [all F(1,75) < 2.8, p>0.09], and male/female ratios did not differ across conditions (all χ^2 < 1.5, p>0.2). Disease duration did not differ for patients across conditions [F(1,40) < 0.7, p>0.4 for all effects]. Additionally, effects of learning set and instruction type among patients were investigated to rule out possible experimental confounds. No differences among patient groups were found on measures of episodic memory, executive function and attention, or visuospatial function; a summary of these neuropsychological results is available in Supplementary Table 1.

2.2. Materials

Stimuli were 92 images of fictitious animals created by manipulating 10 discrete visual features, each with 2 possible values (Table 2). The target category (referred to as a "crutter") was exemplified by the prototype animal (i.e., the image with all 10 of the typical feature values from Table 2). In contrast, the antitype possessed all 10 of the antitypical feature values. The prototype and antitype are illustrated in Figure 1. The remaining 90 stimulus items each had from 1 to 9 typical features. The category boundary between crutters and anti-crutters was non-deterministic: participants were not given a rule to distinguish crutters from non-crutters, and feedback was not given during any phase of the experiment. From this stimulus set, we created two learning conditions. In the low-typicality condition, 10 items with 6 or 7 typical features were presented; in the high-typicality condition, each image shared 8 or 9 of the prototype's features. Within each of these learning sets, all typical features were presented with approximately equal frequency.

2.3. Procedure

The experiment used a 2 × 2 design with between-subjects factors of group (patient or control), instruction type (explicit or implicit), and learning set typicality (high or low). Within each group, participants were randomly assigned to experimental conditions. The experiment comprised 3 phases: a learning phase in which participants were exposed to category exemplars, a test phase in which they judged whether or not items belonged to the target category, and a recognition memory post-test. All 3 phases of the experiment were completed in a single session.

2.3.1. Learning Phase—Each learning set comprised 10 unique items, presented 6 times each; a single random stimulus order was used for each learning set across participants. Neither the antitype nor the prototype was shown during learning. Images were presented for 6 s each, and trials immediately followed one another. Prior to the learning phase, participants were given one of two instruction types. In the explicit condition, participants were told:

You're going to see animals on the computer screen one by one. They are all the same kind of animal, and they're called crutters. As you look at each one, keep in mind that they all belong to the same category.

In the implicit condition, participants were instructed:

You're going to see animals on the computer screen one at a time. I want you to look at each animal as it appears on the screen and think about its appearance.

Every 15 trials, a crosshair fixation screen appeared in the center of the screen. During these pauses, participants in the explicit condition were reminded:

Remember, these are all the same kind of animal.

In the implicit condition, participants were reminded:

Remember to think about the animals' appearance.

2.3.2. Test Phase—After learning, participants proceeded to the test phase, in which they viewed images on 94 trials and judged whether each was a crutter or not. Test items were shown for up to 12 s, and controls responded by pressing one of 2 buttons on a keyboard. Patients responded in the same manner as able; those who could not make a keypress responded verbally, and a research assistant pressed the indicated key for them. (This method was used to reduce the likelihood of erroneous responses due to participants' forgetting the response-key mappings, as we noted in our preliminary studies.) The subsequent trial began immediately after each button press. Test items comprised 2 repetitions each of the antitype and prototype images, as well as 90 additional exemplars presented 1 time each. These 90 exemplars included the 10 training items, as well as 80 novel items, drawn evenly from a range of 1–9 typical features (i.e., 10 items from each level of typicality). A single pseudo-random stimulus order was used for all participants, with learning items from the 2 sets distributed in interleaved fashion throughout. Participants in the explicit condition were instructed:

You will now see a bunch of animals on the screen one at a time. Some of them will be crutters, that is, the same kind of animal you saw before, and some of them won't. Your job is to decide, for each one, whether it is a crutter or not. If you think the animal is a crutter, press ["yes" key] or say "yes." If you think an animal is not a crutter, press ["no" key] or say "no."

In the implicit condition, participants were instructed:

All of the animals you just saw are all the same kind of animal, and they're called crutters. You will now see another bunch of animals one at a time. Some of them will be crutters and some of them won't be. For each one, if you think the animal is a crutter, press ["yes" key] or say "yes." If you think an animal is not a crutter, press ["no" key] or say "no."

Every 15 trials, participants were presented with a fixation point for 5 seconds; during this interval, the experimenter gave a verbal reminder:

Remember, some of the animals are crutters and some of them are not.

2.3.3. Recognition Memory Post-Test—A 2 m delay followed the end of the test phase. Participants next completed a yes/no recognition memory test comprising 10 previously shown and 10 novel images, presented in the same randomly-determined order for all participants. None of the previously shown images had served as learning items, but rather had been first shown during the test phase. They ranged from 1 to 9 in the number of typical features (2 items with 5 typical features and 1 at every other level). The novel images, in turn, all had 8 typical features. Participants were instructed:

> Now you're going to see a few more animals. Some of them will be the exact ones that you've already seen, and some of them will be ones you haven't seen before. Your job is to say for each one whether you've seen that exact animal before or not. For each one, if you think you saw that exact animal before, press ["yes" key] or say "yes." If you think you're seeing an animal for the first time, press ["no" key] or say "no."

As in the test phase, AD participants responded either manually or verbally, and the experimenter pressed the indicated button as needed. Each image appeared for up to 20 s and disappeared as soon as participants responded. After the post-test, participants were asked how they knew whether an item had been previously seen or not. These responses were recorded for qualitative evaluation.

2.4. Behavioral Data Analysis

Between-condition differences in age, education level, and MMSE were investigated using 3-way ANOVAs with factors of group, learning set, and instruction type. Analysis of disease duration was necessarily limited to patients, so a 2-way ANOVA omitting the factor of group was used. An adjusted significance criterion of $p<0.0125$ (corresponding to $p<0.05$ with a Bonferroni correction for 4 ANOVAs) was used for all main effects and interactions. Pairwise follow-up tests used two-tailed, two-sample t-tests assuming equal variance between groups. These t-tests, performed only as indicated by ANOVA results, used a significance threshold of $p<0.05$.

Because crutters and non-crutters were not separated by a deterministic category boundary, responses from the test phase could not be labeled as correct or erroneous. However, we reasoned that effective category learning should be marked by sensitivity to an item's typicality: the more typical features an item possessed, the higher its expected likelihood of being endorsed (Bozoki et al., 2006). Thus, the primary outcome of interest was the slope of the line relating items' proportion of prototypical features to their likelihood of endorsement. This slope, hereafter referred to as a learning score, was calculated for each participant by fitting a linear regression model with single-trial responses as the outcome (1 = endorsed, 0 = not endorsed) and the number of prototypical features on each trial as a single predictor. Category learning differences among experimental conditions were tested by applying an analysis of variance (ANOVA) to a multiple regression model, with learning score as the outcome and factors of group, instruction type, and learning set typicality. Age was included in the model as a nuisance covariate to address the variability in patients' ages.

As an alternative assessment of participants' learning, we also contrasted endorsement rates for the least (0–2 features) and most (8–10 features) typical items. This analysis examined participants' ability to discriminate between the category prototype and antitype, independent of their performance for the most ambiguous items (i.e., those with a more even ratio of typical and antitypical features). Additionally, this alternative analysis allowed us to assess whether experimental factors differentially affected categorization of highly typical and antitypical test items.

The chief outcomes of interest in the recognition memory post-test were individuals' d' scores, representing the ability to discriminate between familiar and novel items. These outcomes were analyzed using ANOVAs on multiple linear regression models with predictors of group, instruction type, and learning set. We additionally examined associations between episodic memory for test items and participants' category learning performance. The effects of item learning history were examined using an ANOVA on a mixed-effects linear model with a factor of learning history (seen in learning phase or new at test), as well as two-way interactions of learning history with group, learning set, and instruction type. The outcome was participants' mean endorsement rate for trained and new stimuli at the trained level of typicality (6–7 or 8–9 features). Paired t-tests were used to test the difference in endorsement rates for trained and new items within each experimental condition; a Bonferroni-adjusted significance threshold of $p<0.00625$ (0.05/8) was used.

2.5. Neuroimaging methods and analysis

Axial T1-weighted anatomical magnetic resonance (MR) images were collected on a Siemens Trio 3.0 Tesla scanner at the Hospital of the University of Pennsylvania using a repetition time of 1620 ms, echo time of 3 ms, 1 mm thick slices, in-plane resolution of 0.9766×0.9766 mm, and field of view of 256×256 x 192 voxels. Only data collected within 6 months of the behavioral testing date were used, resulting in a reduced data set of 31 participants. Raw T1-weighted images were visually inspected as an initial quality check. The PipeDream framework (http://sourceforge.net/projects/neuropipedream/) was used to further assess data quality, estimate grey matter density, and spatially normalize images to a common analysis space. PipeDream is an automated MR processing pipeline based on the publicly available, open-source Advanced Normalization Tools (ANTS, http://www.picsl.upenn.edu/ANTS/; Avants et al., 2014; Tustison et al., 2014). Images were corrected for intensity inhomogeneities using Tustison and Gee's (2009) implementation of the N3 algorithm (Sled et al., 2008). Next, images were aligned to a labeled template brain using a symmetric diffeomorphic normalization method available in ANTs (Avants et al., 2011; Klein et al., 2009). This template was constructed from images of healthy and diseased elderly brains collected using the same scanner and imaging parameters as the current dataset. Next, images in the original acquisition space were subjected to an automated 6-class tissue segmentation (cortical and subcortical grey matter, brainstem, cerebellum, white matter, and cerebrospinal fluid) using a Markov random-field approach (Zhang et al., 2001) supervised by probability maps derived from the template brain. Individual GM density maps were created by combining the brainstem, cortical grey matter (GM), and subcortical GM segmentations; these maps were then warped to MNI template space for analysis. Maps were smoothed with a Gaussian kernel of 4 mm full-width half-

maximum and resampled to 2 mm isotropic voxels. A mean GM density threshold of 0.2 was used to exclude voxels that represented variable anatomy across participants.

The imaging analysis targeted regions of interest (ROIs) implicated in A/not-A category learning tasks (Figures 3 and 4). We hypothesized that explicit learning would engage a declarative system based around the hippocampus (Nomura and Reber, 2008); and that implicit learning would involve either a procedural system based around the striatum (Shohamy et al., 2008) or perceptual learning through modification of visual feature representations in temporo-occipital cortex (TOC; Kéri et al., 2001; Xu et al., 2013). We thus used anatomical labels from the OASIS-30 dataset (Klein and Tourville, 2012) to extract the mean GM probability for each participant in each of 6 ROIs: left and right hippocampus, striatum, and TOC. The striatum ROI comprised the caudate, nucleus accumbens, and putamen, while TOC comprised fusiform, occipital fusiform, and inferior temporal gyri. Mean GM density in each ROI was calculated for each participant. We then computed a regression model for each ROI with learning score as the outcome, and predictors of group, learning set, mean GM density, and their 2- and 3-way interaction terms. A threshold of $p < 0.0083$ (equivalent to $p < 0.05$ after Bonferroni correction) was used to assess the significance of model F-tests.

3. Results

3.1. Overall learning performance

Category learning was measured in terms of participants' sensitivity to item typicality: that is, their tendency to endorse test items with many typical features and to reject those with many atypical features. Collapsed across experimental conditions, learning scores for both patients [mean=0.16 (0.52); $t(43)=2.1$, $p < 0.05$] and controls [mean=0.61 (0.38); $t(38)=10.1$, $p < 0.001$***] exceeded the null hypothesis of zero (i.e., equivalent endorsement rates regardless of test item typicality), indicating a pattern of successful learning across participants. An ANOVA on these learning scores indicated a significant effect of group [$F(1, 74)=23.4$, $p < 0.001$***], reflecting greater learning for controls than for patients [$t(78)=4.6$, $p < 0.001$***].

Similarly, both patients and controls endorsed the most typical test items more frequently than the least typical items, although this difference was greater for controls: an ANOVA on endorsement rates for test items with high (8–10 features) or low (0–2 features) typicality yielded both a main effect of typicality [$F(1,158)=79.4$, $p < 0.001$***] and an interaction of group with typicality [$F(1,158)=29.1$, $p < 0.001$***]. Collapsing across experimental conditions, patients endorsed high-typicality items at a rate of 0.66 (0.22) and low-typicality items at a rate of 0.54 (0.25), a difference which was statistically significant [$t(43)=2.1$, $p < 0.05$]. Controls endorsed high-typicality items at a rate of 0.84 (0.12), and low-typicality items at a rate of 0.37 (0.27). In both groups, the endorsement rate for high-typicality items was significantly greater than a chance level of 0.5 [patients: $t(43)=5.0$, $p < 0.001$*** ; controls: $t(38)=17.6$, $p < 0.001$***]. In contrast, controls were more effective at rejecting less typical items: endorsement of low-typicality items was significantly below chance for controls [$t(38)=-3.0$, $p < 0.01$**] but not for patients [$t(39)=0.5$, $p > 0.5$].

Age was not associated with learning score, either across all participants [$F(1,74)=-0.05$, $p>0.9$] or among patients alone [$F(1,38)=0.1$, $p>0.7$]. Among patients, learning did not reliably differ between those with MCI (mean=0.39, s.d.=0.35) and those with AD (mean=0.15, s.d.=0.55) [$t(10)=1.4$, $p>0.19$]. Patients' MMSE score was also uncorrelated with category learning [n=42; $F(1,35)=1.3$, $p>0.2$] but was negatively associated with disease duration [$F(1,35)=4.2$, $p<0.05$]. Because low- and high-typicality items were arbitrarily defined as those comprising 0–2 and 8–10 typical features, we conducted supplementary analyses using broader definitions (0– 3 and 7–10 features, respectively) and found the same pattern of statistical effects.

3.2. Effects of Instruction Type

To evaluate if learning approach (explicit or implicit) influenced category learning, we investigated the influence of instruction type. Contrary to hypotheses, learning scores were similar for participants trained with explicit and implicit instructions: averaging across learning set conditions, mean learning scores for patients were 0.18 (0.53) in the explicit condition and 0.13 (0.51) in the implicit condition. For controls, learning scores averaged 0.56 (0.39) in the explicit condition and 0.66 (0.37) in the implicit condition. The main effect of instruction type was non-significant [$F(1, 74)=0.0$, $p>0.9$], as were interactions between instruction type and other factors [all $F<0.7$, $p>0.4$]. The observed learning patterns (Figure 2) were not consistent with our hypothesis that controls would learn equally well under explicit and implicit instructions, while patients would learn better in the implicit condition. Subsequent analyses thus omitted the factor of instruction type.

3.3. Effects of Learning Set Typicality

We examined how the typicality of learning exemplars influenced categorization of test items. Learning exemplars in the high-typicality condition each comprised 8–9 typical features, while low-typicality exemplars had 6–7 typical features. Thus, participants in the high-typicality learning set condition encountered stimuli that were highly similar to the unseen category prototype and thus more similar to one another, resulting in lower feature variability than for stimuli in the low-typicality condition. As predicted, learning from high-typicality exemplars led to better categorization than learning from low-typicality exemplars [main effect of learning set: $F(1,74)=16.0$, $p<0.001$; $t(77)=3.6$, $p<0.001$], suggesting that participants were able to learn from exemplars with low feature variability and high similarity to the category prototype. Controls exposed to the high-typicality learning set had a mean learning score of 0.88 (0.23), which was significantly better than the mean of 0.36 (0.32) for controls in the low-typicality learning condition [$t(35)=5.8$, $p<0.001$]; these means significantly exceeded the null hypothesis for both groups [high-typicality learning: $t(18)=16.6$, $p<0.001$***; low-typicality learning: $t(19)=5.1$, $p<0.001$***]. Patients showed a similar trend: patients in the high-typicality learning condition had a mean learning score of 0.29 (0.53), which significantly exceeded the null hypothesis [$t(20)=2.5$, $p<0.03$], while patients in the low-typicality learning condition had a mean score of 0.04 (0.48), which was not reliably different from zero [$t(22)=0.4$, $p<0.7$]. However, in a direct contrast, patients' scores did not reliably differ between learning conditions [$t(40)=1.6$, $p = 0.12$]. The interaction of group and learning set typicality was also non-significant [$F(1, 74)=2.1$, $p>0.15$].

Categorization of high (8–10 features) and low (0–2 features) typicality test items also differed according to learning set condition, as supported by an interactive effect of learning set and high/low test item typicality on endorsement rates [F(1,158)=17.9, p<0.001***]. The three-way interaction of group, learning set, and test item typicality approached but did not reach significance [F(1,158)=2.8, p<0.1]. Controls in both learning set conditions endorsed high-typicality items significantly more often than low-typicality items (Table 3), although this effect was greater for controls who learned from the high-typicality learning set than those who learned from the low-typicality learning set (Table 3: t=17.0 and t=5.2, respectively). Like controls, patients who learned from the high-typicality learning set endorsed high-typicality test items significantly more often than low-typicality items [Table 3; t(20)=2.5, p<0.03], providing further evidence of patient learning. In contrast, patients in the low-typicality learning condition showed no reliable difference in their treatment of low- and high-typicality test items [t(22)=0.6, p<0.6].

Follow-up analyses addressed whether the above learning set effects resulted from participants' treatment of low-typicality items, high-typicality items, or both. A two-way ANOVA on endorsement rates for low-typicality test items yielded main effects of group [F(1,79)=10.9, p<0.01] and learning set [F(1,79)=17.4, p<0.001], as well as an interaction of group and learning set [F(1,79)=4.4, p<0.05]. Controls in the high-typicality learning condition had lower endorsement rates for low-typicality items (mean=0.20, s.d.=0.19) than their counterparts in the low-typicality learning condition [mean=0.53, s.d.=0.24; t(36)=4.8, p<0.001]. Their endorsement rates for the antitype alone were 0.38 (0.43) and 0.08 (0.25) in the low- and high-typicality learning conditions, respectively. In contrast, the two patient groups did not differ in their endorsement rates for low-typicality items. Patients endorsed low-typicality items at a rate of 0.59 (0.24) in the low-typicality learning condition and 0.48 (0.26) in the high-typicality learning condition [t(41)=1.5, p<0.15]. Patients' specific endorsement rates for the antitype were 0.57 (0.46) and 0.36 (0.42) in the low- and high-typicality learning conditions. Controls in the high-typicality learning condition were less likely to endorse low-typicality items than patients in either the high-typicality [mean=0.48 (0.26); t(37)=-3.9, p<0.001] or low-typicality learning conditions [mean=0.59 (0.24); t(40)=-5.9, p<0.001]. Collectively, these results suggest that only controls exposed to the high similarity learning set were able to reject less representative stimuli.

Interestingly, learning set typicality did not appear to influence endorsement of the most typical items, including the category prototype. An ANOVA on endorsement rates for test items with 8–10 typical features yielded only a main effect of group [F(1,79)=20.8, p<0.001]; the effect of learning set [F(1,79)=2.3, p<0.14] and its interaction with group [F(1,79)=0.0, p<0.99] were both non-significant. Endorsement rates for high-typicality test items did not differ for the two control groups [t(37)=1.6, p<0.13]: controls endorsed these items at a rate of 0.81 (0.12) in the low-typicality condition and 0.87 (0.12) in the high-typicality condition. Their specific endorsement rates for the prototype were 0.90 (0.21) in the low-typicality learning condition and 0.97 (0.11) in the high-typicality learning condition. Similarly, patients' endorsement rates for high-typicality items were 0.63 (0.23) in the low-typicality learning condition and 0.69 (0.20) in the high-typicality learning condition [t(42)=-0.9, p<0.4]. Patients endorsed the prototype alone at rates of 0.67 (0.36) and 0.79 (0.34) in the low- and high-typicality condition.

3.5. Associations with Neuroanatomy

In order to explain individual variability in learning scores that was not accounted for by behavioral analyses, we additionally examined associations with grey matter integrity in brain areas of interest. We conducted behavioral analyses on the subset of participants included in the imaging analysis, in order to assess their comparability to the full dataset. Imaged participants showed the same pattern of behavioral effects, including higher learning scores for controls than patients [t(21)=2.6, p < 0.05*] and for the high- relative to the low-typicality learning condition [t(27)=2.1, p < 0.05*]. As in the full dataset, learning did not differ for participants in the explicit and implicit learning conditions, and no significant interactions were observed among experimental factors [F(1,21) < 0.5, p>0.49 for all effects].

We sought to investigate associations between learning scores and 3 ROIs in each hemisphere, for a total of 6 tests. Hippocampus and TOC were atrophied bilaterally in the patient group relative to controls (Figure 4) [all t > 3.4, p < 0.01]; in contrast, striatal grey matter density did not differ in either hemisphere [left: t(27)=0.2, p < 0.9; right: t(26)=-0.9, p < 0.4]. Associations with behavior were investigated using multiple regression models for each ROI. The baseline model for these tests included learning score as the outcome, with predictors of group, learning set typicality, and the 2-way interaction of those factors. For each ROI, we added mean grey matter density, its interactions with group and learning set typicality, and the 3-way interaction of these factors as predictors to this baseline model. Of the 6 ROIs tested, only left hippocampus was a significant predictor of category learning performance, explaining 55% of the variance in learning scores (adjusted R2). An ANOVA on the left hippocampal model indicated a significant main effect of GM density [F(1,21)=9.9, p < 0.01**] and an interaction with group [F(1,21)=11.6, p < 0.01**]. Hippocampal GM was negatively associated with learning score in the patients [t=-3.2, p < 0.01**] but positively associated with controls' performance [t=2.8, p < 0.05*]. The association between left hippocampus and learning did not appear to depend on which learning set participants received (non-significant 2-way interaction of learning set x hippocampus and 3-way interaction of group x learning set x hippocampus). Figure 5 shows the best-fit regression line for each combination of the 2 groups and 2 learning set conditions in the left hippocampus.

3.6. Episodic Memory Effects

We additionally investigated potential episodic memory effects that would support the use of an explicit, declarative approach to the category learning task. First, we evaluated differences in categorization for items that had been learning exemplars, relative to those that were new at test; we reasoned that participants using an episodic approach would be more likely to endorse familiar than novel items. Overall, the endorsement rate of previously-seen items ("old") did not differ from that of items newly encountered in the test phase ("new") [non-significant effect of item history, F(1,78)=2.4, p<0.13]. However, the interaction of group and item history was significant [F(2,78)=9.7, p<0.001]; this result reflected the tendency for controls to show a greater learning history effect (i.e., higher endorsement rates for learning items) than patients (Figure 6). The interaction of learning set typicality with item history was also significant [F(2, 78)=3.8, p<0.05]. Controls in the high-

typicality learning condition endorsed old items at a rate of 0.91 (std. dev. = 0.12), and new items at a rate of 0.82 (std. dev. = 0.16), a difference which was significant [t(18)=3.0, p<0.01]. Learning history effects were not statistically reliable for the remaining conditions (all t<0.7, p>0.4). The pattern of results suggests that item familiarity factored into the categorization decisions of controls in the high-typicality learning condition, consistent with the use of an episodic learning approach.

Second, we examined performance on the recognition-memory post-test as a function of experimental condition. We reasoned that participants who employed explicit, declarative learning approaches were likely to exhibit better memory for test items. On the recognition memory post-test, d' scores differed significantly by group [F(1,74)=23.1, p<0.001], reflecting the episodic memory impairment of patients [controls: mean=2.2 (1.2); patients: mean=1.0 (1.1)]. Controls' d' scores were better than patients' within both the low- [t(40)=3.1, p<0.01] and high-typicality [t(33)=3.7, p<0.001] learning conditions.

After the recognition memory post-test, a subset of participants (n=44) were asked how they knew whether items were old or new. (Responses were unavailable in other cases due to non-response on the participant's part, omission of the question due to time constraints, or experimenter oversight.) Their responses were summarized by grouping them into 5 response categories (Table 4). All 19 controls interviewed described decision criteria that suggested an explicit response strategy, including mention of single features, mention of multiple features, or some other intentional episodic strategy (e.g., "If it looked similar to one that I saw before."). Notably, this control sample included 7 participants who learned under the implicit instruction set. In the patient group, 12 of 25 participants gave responses that suggested explicit strategies; of these, only 6 had participated in the explicit instruction condition. Twelve of the remaining 13 patients reported a strategy inconsistent with explicit memory, e.g. that they made arbitrary old/new decisions (Table 4, row 1: didn't remember, guessed) or used a non-recollective strategy (Table 4, row 2: relied on a feeling of intuition, familiarity, or novelty). These participants are likely to reflect a combination of individuals whose old/new decisions were truly random and perhaps some who responded on the basis of implicit memory for familiar stimuli. A single patient gave an irrelevant response (i.e., "because you made me look at the animals the first time [in the learning phase]"). Of these patients, 7 had received explicit instructions, and 5 had received implicit instructions. The likelihood of responses suggesting an implicit strategy significantly differed by group [$\chi^2 = 14$, p<p<0.001], but not by learning set typicality [$\chi^2 = 0$, p>0.9] or instruction type [$\chi^2 = 0$, p>0.9]. Collectively, these results suggest that experimental instructions did not affect the nature (explicit or implicit) of participants' memory for test items. Rather, group (patients or controls) was the best predictor of strategy on the recognition memory test, with controls exclusively adopting an explicit approach, and patients alternately reporting implicit and explicit approaches.

4. Discussion

Previous studies have shown that patients with Alzheimer's disease retain some capacity to learn novel semantic categories (Bozoki et al., 2006; Grossman et al., 2013a; Heindel et al., 2013; Kéri et al., 2001; Koenig et al., 2007, 2010). The current study suggests important

qualifications to the magnitude of and necessary circumstances for patients' learning. First, controls' learning effects were larger and more reliable than those of patients. Second, early exposure to a high-typicality learning set significantly enhanced later categorization performance and appeared to be a necessary condition for patients' learning. Third, the manipulation of learning set typicality appeared mainly to influence participants' responses to low-typicality test items. Finally, we found that associations between learning score and left hippocampal grey matter differed in direction between controls and patients, a finding that warrants further investigation using both functional and anatomical imaging.

Both controls and patients learned features of the novel crutter category, as above-chance learning scores in both groups indicated an association between endorsement rate and stimulus typicality. Moreover, controls and patients alike endorsed high-typicality items more often than low-typicality test items, indicating their ability to discriminate between the category prototype and antitype, independently of their responses to more ambiguous test items. However, learning strongly depended on the quality of stimuli that participants encountered during learning. Controls exposed to the high-typicality learning set demonstrated stronger sensitivity to test item typicality than controls in the low-typicality learning condition. Among patients, only those exposed to the high-typicality learning set demonstrated learning, both in their learning scores and through their ability to discriminate high- and low-typicality items. At the same time, patients' learning scores did not differ in a direct contrast of low- and high-typicality learning conditions. The current results thus do not establish a clear benefit for the high-typicality learning set in AD patients; however, qualitatively similar effects in the control and patient groups, coupled with the prominence of learning set effects in the omnibus analysis, suggest category knowledge is influenced by participants' experience of feature variability among category exemplars. In naturalistic learning situations, learners whose initial exposure to a category involves highly typical exemplars with low feature variability may have a distinct advantage over those who encounter a more heterogeneous set of exemplars, a possibility that is corroborated by previous research in visual (Bozoki et al., 2006) and spatial (Jee et al., 2013) concept learning. Moreover, experiential factors in category learning are relevant to prior research demonstrating that the organization of semantic knowledge can differ systematically between cultural groups (Love and Gureckis, 2005; Medin et al., 2006; Ross et al., 2003).

We consider the above results with respect to participants' ability or inability to develop an internal category prototype to guide their discrimination between category members and non-members. This prototype development can be the result of either explicit (i.e., episodic or semantic) or implicit (i.e., procedural or perceptual) learning (Smith, 2008). In either case, we propose that ideal categorization performance is likely to require an explicit strategy, implemented either during the learning or test phases. Because typical and atypical features alike were presented during the learning phase, both were associated with a certain degree of familiarity (or perceptual fluency, as argued by Smith, 2008). For atypical features, this familiarity trace may provide a signal that works counter to optimal categorization decisions, just as lure familiarity or habitual responses can increase error rates in episodic memory tests (Kelley and Jacoby, 2000). Participants who learn a category explicitly have the opportunity during learning to selectively encode the more typical features and to suppress encoding of atypical features. Encoded features can then be consolidated in

memory and recalled during the categorization test as part of an explicit prototype representation. In contrast, participants who learn implicitly are unaware during learning that typical and atypical features have differential associations with a target category. In the test phase, these participants would need to explicitly distinguish between more and less typical features based on recall of feature frequencies during the learning phase. Participants who do not engage in such recollection, but rather respond based on implicit memory or feature familiarity, would not be expected to distinguish between familiar features of the target and distractor categories (i.e., the prototype and its antitype).

Controls in the high-typicality condition had a nearly ideal response profile, endorsing the prototype 97% of the time, and the antitype only 8% of the time. We thus propose that controls in the high-typicality condition likely developed an explicit representation of the category prototype that guided their responses. It is less clear what approach controls in the low-typicality condition adopted. Some low-typicality controls may have used an explicit prototyping strategy like the one outlined above, but performed less efficiently due to the greater dispersion of features in the low-typicality learning set. Other controls may have relied on implicit memory or intuition, responding to atypical and typical features in proportion to their frequency in the learning set. In either case, it is clear that the learning experience of low-typicality controls did not promote an ideal internal prototype. The variability of patients' results likewise admits multiple possible explanations for their categorization performance. Some patients may have formulated a limited prototype model based on a subset of typical features (Zaki and Nosofsky, 2001) or may have responded based on implicit memory, driven by feature frequency during the learning phase.

Interestingly, effects of group and learning set typicality did not affect performance for all test items equally. Although all groups reliably endorsed highly typical test items at a rate above chance, only controls exposed to the high-typicality learning set were able to reliably reject low-typicality items (those possessing 0–2 typical features). Patients and controls in the low-typicality learning condition may thus have had difficulty in rejecting non-member lures that shared features with category members. Because these items could contain 1–2 typical features, participants who attended to those features exclusively may have endorsed them as category members. Patients may have been particularly disadvantaged in processing low-typicality items due to their episodic memory impairments: even antitypical features were present in the learning set (albeit with low frequency), and patients may have experienced difficulty in discriminating weak familiarity traces from multiple antitypical features from those of highly-typical items. Relatedly, participants relying on implicit learning may have resorted to guessing for low-typicality items in the absence of a clear familiarity signal driven by multiple typical stimulus features. We cannot rule out the possibility that confusion about task instructions may explain the endorsement of low-typicality items. For example, some participants may have assumed that any feature seen during the learning phase, however infrequently, was characteristic of the crutter category. These participants may have been reserving "no" responses in anticipation of a more obviously different foil set (e.g., a series of fish-like animals). However, the inclusion of periodic verbal prompts during the categorization test ("Remember, some of the animals are crutters, and some of them are not") makes this interpretation less likely.

Some evidence suggests that elderly controls were able to adopt an explicit learning strategy during acquisition of the novel crutter category, while patients were less likely to do so. The recognition memory results confirm elderly controls' superior explicit memory capacity and are consistent with episodic encoding of stimulus items. Moreover, elderly controls exposed to the high typicality condition had higher endorsement rates for previously seen stimuli than novel stimuli. Controls' reflections on their own recognition memory also suggest a group difference in their explicit awareness of stimulus features. In reflecting on the recognition memory post-test, nearly all of the controls that we interviewed said they had considered multiple stimulus features in their old/new decisions. Additionally, controls' grey matter density in left hippocampus was positively associated with learning score. Prior work has linked both structural and functional imaging measurements of hippocampus to the use of explicit strategies (Nomura and Reber, 2008), suggesting that controls may have used such a strategy in the current study. In contrast, learning scores for patients were negatively associated with hippocampal grey matter density. It is possible that this negative association reflects differential reliance on declarative and non-declarative learning mechanisms: For example, Ashby & Maddox (2011) propose that the implicit learning system is chronically active, but that input from this system only influences categorization when the declarative system releases control over behavior. Indeed, a number of functional neuroimaging studies have found functional connectivity or complementary response patterns in brain areas associated with declarative and non-declarative learning (Davis et al., 2012; DeGutis et al., 2007; Seger et al., 2011). Although the specific interpretation of these results is unclear, they collectively suggest that category learning involves the dynamic interaction of multiple memory systems. Patients who exhibited the greatest learning may have relied on implicit, non-declarative learning mechanisms, such as perceptual or procedural learning. However, this interpretation was not supported by positive associations between patients' performance and the visual-perceptual system in TOC, nor between performance and the procedural system in the striatum. Future studies can address these hypotheses more directly by functional neuroimaging of patients during task performance. Additionally, high-resolution anatomical imaging of hippocampal subfields (e.g., Travis et al., 2014; Yassa et al., 2010; Yushkevich et al., 2015) would be valuable for determining the anatomical basis of hippocampal dysfunction in patients.

Contrary to hypothesis, no reliable differences between implicit and explicit instruction conditions were observed. We predicted that the patient group would be selectively impaired in the explicit condition, due to disease-related pathology in brain areas supporting episodic memory (Bozoki et al., 2006; Knowlton and Squire, 1993) and executive function (Grossman et al., 2013a; Koenig et al., 2007); in contrast, the patient group was expected to demonstrate relatively preserved performance in the implicit condition, due to the relative sparing of TOC (Frisoni et al., 2007), which is associated with perceptual learning of novel categories (Grossman et al., 2013b; Peelle et al., 2014). We further predicted that the effect of learning set typicality would be attenuated among controls in the explicit learning condition, as their intact declarative memory would allow them to acquire the category regardless of the typicality and variability of the features in the learning set. However, neither group's learning reliably differed according to instruction type. While it is possible that the null effect of instruction type was a simple result of insufficient statistical power,

this explanation is not compelling. Explicit and implicit response profiles (Figure 2) overlapped heavily, and it is unlikely that these profiles would become more distinct with the inclusion of more participants' data. Moreover, in one prominent contrast of explicit and implicit learning paradigms, Reber et al. (2003) reported an accuracy difference of nearly 9% between explicitly (N=8) and implicitly (n=12) trained participants, corresponding to an effect size of Cohen's d=2.2. We may assume that this effect is an outlier and instead propose a more modest effect size of 0.6; however, even with this reduced estimate the current study should have 77% power to detect an effect of instruction type in the omnibus analysis. The simplest explanation for this null effect is that participants did not reliably comply with different instruction sets promoting explicit or implicit processing strategies. Some participants in the explicit condition—particularly patients—may have disregarded the instruction to attend to and learn features of the novel category. Conversely, some participants in the implicit condition may have anticipated a categorization test and explicitly attempted to learn the common features of the learning exemplars. It is also possible that the implicit learning instruction to think about each animal's appearance may have encouraged explicit processing of learning exemplars.

Alternatively, it is possible that manipulation of task instructions does not significantly influence category learning, due to the independence of explicit and implicit learning. Prevailing theory proposes that explicit and implicit systems operate in parallel during most learning tasks, with some form of interaction between them (e.g., Ashby and Maddox, 2011; Sadeh et al., 2011). However, this interaction is limited: for example, Sanchez and Reber (2013) found that explicit instruction on an implicit perceptual-motor learning task raised explicit awareness of task objectives but did not affect implicit learning. Furthermore, Gureckis et al. (2011) provided evidence that the result of Reber et al. (2003) could have been driven by differences in learning task demands in the explicit and implicit conditions. When Gureckis et al. tested participants on explicit and implicit learning tasks with similar stimulus encoding methods, categorization accuracy was highly similar between the two conditions. Despite a well-developed category learning literature, there remain relatively few studies (including the current one) which have directly contrasted explicit and implicit learning approaches while holding task-and stimulus-related factors constant. The apparent failure of a simple instruction manipulation for promoting engagement of different learning systems is thus informative, and it highlights the need for further comparisons of explicit and implicit learning that avoid potential confounds such as those discussed by Gureckis et al. (2011). One interesting approach to differentiating explicit and implicit learning involves manipulating the delay at which feedback is presented to participants, with longer delays apparently promoting the engagement of a declarative learning system involving the hippocampus (Foerde and Shohamy, 2011; Smith et al., 2014).

Several aspects of the current study constrain our interpretation of its findings and highlight the need for further research on category learning in neurodegenerative disease. First, effects of learning set typicality and instruction type were tested in a between-subjects design, possibly resulting in reduced sensitivity for detecting true effects. Although within-subjects manipulation of these factors would be preferable from a statistical standpoint, such a study would be complicated by interference or facilitation effects and the loss of participants' naivety between the first and second sessions. Second, some participants' data were

excluded due to an endorsement rate of close to 100 percent, suggesting a failure to understand task instructions. Follow-up experiments should thus include an extended instruction phase in which the experimenter asks the participant to recapitulate the task instructions in his/her own words; additionally, we will include automated response checks that alert the experimenter to repeated use of a single response key. Third, the current experiment uses an "A/not-A" task (Heindel et al., 2013), in which participants are trained on a single category and then asked to endorse or reject items as members of this category. It is thus unclear how results from the current study, including learning set typicality effects, apply to "A/B" category learning tasks in which participants simultaneously learn 2 distinct categories. Fourth, patients with AD and MCI were grouped together to maximize statistical power, potentially obscuring performance differences between these 2 groups. Post-hoc analyses showed that the inclusion of MCI participants did not qualitatively change the observed effects of group or learning set. Furthermore, it can be argued that the distinction between a state of mild cognitive impairment and mild dementia is largely one of degree, based on neuropsychological convention and clinical judgment. Nevertheless, a direct contrast of category learning performance between individuals with AD and MCI potentially would be informative and should be pursued in future work. Fifth, the current study investigated a limited set of hypotheses regarding the anatomical correlates of category learning. Prefrontal cortex is notably implicated in explicit category learning (Ashby and Maddox, 2011), although it has also been linked to implicit learning (Kincses et al., 2004); the role of specific prefrontal subregions in both explicit and implicit category learning merits targeted investigation using functional imaging techniques. Sixth, it is possible that participants' categorization decisions evolved throughout the test phase as a result of learning during test (Bozoki et al., 2006); however, typical and atypical feature values were presented in equal proportion during the test phase, making a net change in participants' estimates of feature probability unlikely. Finally, future work should contrast the learning deficits of AD and MCI patients with those of other neurodegenerative disease patients. For example, individuals with frontotemporal dementia clinical syndromes (including semantic-variant primary progressive aphasia and behavioral-variant frontotemporal dementia) may be differentially sensitive to the typicality of learning items.

Overall, the results confirm and extend prior findings that patients with AD and other neurodegenerative diseases are capable of new category learning (Bozoki et al., 2006; Koenig et al., 2007, 2010), albeit at a reduced level relative to age-matched controls. To our knowledge, the feature space used in the current experiment (10 discrete visual features, as in Bozoki et al., 2006) is among the most complex used to date in studies of category learning in neurodegenerative disease, underscoring the significance of patients' learning. A strong effect of learning typicality in controls suggests that reducing feature variability and providing learners with early exposure to high-typicality exemplars is a general approach for improving category learning. At present, it is unclear to what extent such manipulations may benefit category learning in AD. Only patients exposed to the high-typicality learning set demonstrated significant learning; however, these patients' learning effects were not significantly greater than those of patients in the low-typicality learning condition. Close analysis of response profiles for the most typical and least typical test items indicates that patients and healthy seniors alike will benefit from learning manipulations that target their

ability to reject low-typicality non-member stimuli. Further research is needed to clarify the neuroanatomical basis of preserved category learning in AD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Phelps CH. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia. 2011; 7(3):270–279. http://doi.org/10.1016/j.jalz.2011.03.008.

Ashby FG, Maddox WT. Human Category Learning 2.0. Annals of the New York Academy of Sciences. 2011; 1224:147–161. http://doi.org/10.1111/j.1749-6632.2010.05874.x. [PubMed: 21182535]

Ashby, FG., Paul, E., Maddox, WT. COVIS 2.0. Formal Approaches in Categorization. New York, NY: Cambridge University Press; 2011.

Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage. 2011; 54(3):2033–2044. [PubMed: 20851191]

Avants BB, Tustison NJ, Stauffer M, Song G, Wu B, Gee JC. The Insight ToolKit image registration framework. Frontiers in Neuroinformatics. 2014; 8 http://doi.org/10.3389/fninf.2014.00044.

Bozoki A, Grossman M, Smith EE. Can patients with Alzheimer's disease learn a category implicitly? Neuropsychologia. 2006; 44(5):816–827. http://doi.org/10.1016/j.neuropsychologia.2005.08.001. [PubMed: 16229868]

Davis T, Love BC, Preston AR. Striatal and hippocampal entropy and recognition signals in category learning: Simultaneous processes revealed by model-based fMRI. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2012; 38(4):821–839. http://doi.org/http://dx.doi.org/10.1037/a0027865.

DeGutis J, D'Esposito M. Distinct mechanisms in visual category learning. Cognitive, Affective & Behavioral Neuroscience. 2007; 7(3):251–259. http://doi.org/http://dx.doi.org/10.3758/CABN.7.3.251.

de Vries MH, Ulte C, Zwitserlood P, Szymanski B, Knecht S. Increasing dopamine levels in the brain improves feedback-based procedural learning in healthy participants: An artificial-grammar-learning experiment. Neuropsychologia. 2010; 48(11):3193–3197. http://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2010.06.024. [PubMed: 20600185]

Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, Scheltens P. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. The Lancet Neurology. 2007; 6(8):734–746. http://doi.org/10.1016/S1474-4422(07)70178-3. [PubMed: 17616482]

Ell SW, Ashby FG. The effects of category overlap on information-integration and rule-based category learning. Perception & Psychophysics. 2006; 68(6):1013–1026. http://doi.org/http://dx.doi.org/10.3758/BF03193362. [PubMed: 17153195]

Foerde K, Shohamy D. Feedback Timing Modulates Brain Systems for Learning in Humans. The Journal of Neuroscience. 2011; 31(37):13157–13167. http://doi.org/10.1523/JNEUROSCI.2701-11.2011. [PubMed: 21917799]

Folstein J, Newton A, Gulick ABV, Palmeri T, Gauthier I. Category learning causes long-term changes to similarity gradients in the ventral stream: A multivoxel pattern analysis at 7T. Journal of Vision. 2012; 12(9):1106–1106. http://doi.org/10.1167/12.9.1106.

Folstein JR, Palmeri TJ, Gauthier I. Category learning increases discriminability of relevant object dimensions in visual cortex. Cerebral Cortex. 2013; 23(4):814–823. http://doi.org/http://dx.doi.org/10.1093/cercor/bhs067. [PubMed: 22490547]

Frisoni GB, Pievani M, Testa C, Sabattoli F, Bresciani L, Bonetti M, Thompson PM. The topography of grey matter involvement in early and late onset Alzheimer's disease. Brain: A Journal of Neurology. 2007; 130(3):720–730. [PubMed: 17293358]

Gifford KA, Phillips JS, Samuels LR, Lane EM, Bell SP, Liu D, Jefferson AL. Associations between Verbal Learning Slope and Neuroimaging Markers across the Cognitive Aging Spectrum. Journal of the International Neuropsychological Society. 2015; 21(06):455–467. http://doi.org/10.1017/S1355617715000430. [PubMed: 26219209]

Grossman M, Peelle JE, Smith EE, McMillan CT, Cook P, Powers J, Burkholder L. Category-specific semantic memory: Converging evidence from bold fMRI and Alzheimer's disease. NeuroImage. 2013; 68:263–274. http://doi.org/10.1016/j.neuroimage.2012.11.057. [PubMed: 23220494]

Heindel WC, Festa EK, Ott BR, Landy KM, Salmon DP. Prototype learning and dissociable categorization systems in Alzheimer's disease. Neuropsychologia. 2013; 51(9):1699–1708. http://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2013.06.001. [PubMed: 23751172]

Jee BD, Uttal DH, Gentner D, Manduca C, Shipley TF, Sageman B. Finding faults: Analogical comparison supports spatial concept learning in geoscience. Cognitive Processing. 2013; 14(2):175–187. http://doi.org/http://dx.doi.org/10.1007/s10339-013-0551-7. [PubMed: 23436210]

Kelley, CM., Jacoby, LL. Recollection and familiarity: Process-dissociation. In: Tulving, E., Craik, FIM., editors. The Oxford handbook of memory. New York, NY, US: Oxford University Press; 2000. p. 215-228.

Kéri S, Janka Z, Benedek G, Aszalós P, Szatmáry B, Szirtes G, Lörincz A. Categories, prototypes and memory systems in Alzheimer's disease. Trends in Cognitive Sciences. 2002; 6(3):132–136. http://doi.org/http://dx.doi.org/10.1016/S1364-6613(00)01859-3. [PubMed: 11861191]

Kéri S, Kálmán J, Kelemen O, Benedek G, Janka Z. Are Alzheimer's disease patients able to learn visual prototype? Neuropsychologia. 2001; 39(11):1218–1223. http://doi.org/http://dx.doi.org/10.1016/S0028-3932(01)00046-X. [PubMed: 11527559]

Kincses TZ, Antal A, Nitsche MA, Bártfai O, Paulus W. Facilitation of probabilistic classification learning by transcranial direct current stimulation of the prefrontal cortex in the human. Neuropsychologia. 2004; 42(1):113–117. http://doi.org/10.1016/S0028-3932(03)00124-6. [PubMed: 14615081]

Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, Parsey RV. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage. 2009; 46(3):786–802. http://doi.org/10.1016/j.neuroimage.2008.12.037. [PubMed: 19195496]

Klein A, Tourville J. 101 labeled brain images and a consistent human cortical labeling protocol. Frontiers in Neuroscience. 2012; 6:171. http://doi.org/10.3389/fnins.2012.00171. [PubMed: 23227001]

Knowlton BJ, Squire LR. The learning of categories: parallel brain systems for item memory and category knowledge. Science (New York, N.Y.). 1993; 262(5140):1747–1749.

Koenig P, Smith EE, Glosser G, DeVita C, Moore P, McMillan C, Grossman M. The neural basis for novel semantic categorization. NeuroImage. 2005; 24(2):369–383. http://doi.org/10.1016/j.neuroimage.2004.08.045. [PubMed: 15627580]

Koenig P, Smith EE, Grossman M. Categorization of novel tools by patients with Alzheimer's disease: Category-specific content and process. Neuropsychologia. 2010; 48(7):1877–1885. http://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2009.07.023. [PubMed: 19664644]

Koenig P, Smith EE, Moore P, Glosser G, Grossman M. Categorization of novel animals by patients with Alzheimer's disease and corticobasal degeneration. Neuropsychology. 2007; 21(2):193–206. http://doi.org/http://dx.doi.org/10.1037/0894-4105.21.2.193. [PubMed: 17402819]

Koenig P, Smith EE, Troiani V, Anderson C, Moore P, Grossman M. Medial Temporal Lobe Involvement in an Implicit Memory Task: Evidence of Collaborating Implicit and Explicit

Memory Systems from fMRI and Alzheimer's Disease. Cerebral Cortex. 2008; 18(12):2831–2843. [PubMed: 18400793]

Libon DJ, Bogdanoff B, Cloud BS, Skalina S, Giovannetti T, Gitlin HL, Bonavita J. Declarative and procedural learning, quantitative measures of hippocampus, and subcortical white alterations in Alzheimer's disease and ischaemic vascular dementia. Journal of Clinical and Experimental Neuropsychology. 1998; 20(1):30–41. http://doi.org/http://dx.doi.org/10.1076/jcen.20.1.30.1490. [PubMed: 9672817]

Love, BC., Gureckis, TM. Modeling Learning Under the Influence of Culture. In: -k, W.Goldstone, RL.Love, BC.Markman, AB., Wolff, P., editors. Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin. Washington, DC, US: American Psychological Association; 2005. p. 229-248.

Love BC, Medin DL, Gureckis TM. SUSTAIN: A Network Model of Category Learning. Psychological Review. 2004; 111(2):309–332. http://doi.org/10.1037/0033-295X.111.2.309. [PubMed: 15065912]

Maddox WT, Glass BD, O'Brien JB, Filoteo JV, Ashby FG. Category label and response location shifts in category learning. Psychological Research. 2010; 74(2):219–236. http://doi.org/10.1007/s00426-009-0245-z. [PubMed: 19471959]

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Phelps CH. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia. 2011; 7(3):263–269. http://doi.org/10.1016/j.jalz.2011.03.005.

Medin DL, Ross NO, Atran S, Cox D, Coley J, Proffitt JB, Blok S. Folkbiology of freshwater fish. Cognition. 2006; 99(3):237–273. http://doi.org/10.1016/j.cognition.2003.12.005. [PubMed: 16055109]

Mickes L, Wixted JT, Fennema-Notestine C, Galasko D, Bondi MW, Thal LJ, Salmon DP. Progressive impairment on neuropsychological tasks in a longitudinal study of preclinical Alzheimer's disease. Neuropsychology. 2007; 21(6):696–705. http://doi.org/10.1037/0894-4105.21.6.696. [PubMed: 17983283]

Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitelman DR, Parrish TB, Reber PJ. Neural Correlates of Rule-Based and Information-Integration Visual Category Learning. Cerebral Cortex. 2007; 17(1):37–43. http://doi.org/10.1093/cercor/bhj122. [PubMed: 16436685]

Nomura EM, Reber PJ. A review of medial temporal lobe and caudate contributions to visual category learning. Neuroscience & Biobehavioral Reviews. 2008; 32(2):279–291. http://doi.org/10.1016/j.neubiorev.2007.07.006. [PubMed: 17868867]

Nomura EM, Reber PJ. Combining Computational Modeling and Neuroimaging to Examine Multiple Category Learning Systems in the Brain. Brain Sciences. 2012; 2(2):176–202. http://doi.org/10.3390/brainsci2020176. [PubMed: 24962771]

Nosofsky RM, Denton SE, Zaki SR, Murphy-Knudsen AF, Unverzagt FW. Studies of implicit prototype extraction in patients with mild cognitive impairment and early Alzheimer's disease. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2012; 38(4):860–880. http://doi.org/http://dx.doi.org/10.1037/a0028064.

Peelle JE, Powers J, Cook PA, Smith EE, Grossman M. Frontotemporal neural systems supporting semantic processing in Alzheimer's disease. Cognitive, Affective, & Behavioral Neuroscience. 2014; 14(1):37–48. http://doi.org/10.3758/s13415-013-0239-6.

Poldrack RA, Foerde K. Category learning and the memory systems debate. Neuroscience & Biobehavioral Reviews. 2008; 32(2):197–205. http://doi.org/10.1016/j.neubiorev.2007.07.007. [PubMed: 17869339]

Reber PJ, Gitelman DR, Parrish TB, Mesulam MM. Dissociating Explicit and Implicit Category Knowledge with fMRI. Journal of Cognitive Neuroscience. 2003; 15(4):574–583. http://doi.org/10.1162/089892903321662958. [PubMed: 12803968]

Reber PJ, Knowlton BJ, Squire LR. Dissociable properties of memory systems: differences in the flexibility of declarative and nondeclarative knowledge. Behavioral Neuroscience. 1996; 110(5):861–871. [PubMed: 8918990]

Ross N, Medin D, Coley JD, Atran S. Cultural and experiential differences in the development of folkbiological induction. Cognitive Development. 2003; 18(1):25–47. http://doi.org/10.1016/S0885-2014(02)00142-9.

Sadeh T, Shohamy D, Levy DR, Reggev N, Maril A. Cooperation between the hippocampus and the striatum during episodic encoding. Journal of Cognitive Neuroscience. 2011; 23(7):1597–1608. http://doi.org/http://dx.doi.org/10.1162/jocn.2010.21549. [PubMed: 20666593]

Sanchez DJ, Reber PJ. Explicit pre-training instruction does not improve implicit perceptual-motor sequence learning. Cognition. 2013; 126(3):341–351. http://doi.org/10.1016/j.cognition.2012.11.006. [PubMed: 23280147]

Seger CA, Dennison CS, Lopez-Paniagua D, Peterson EJ, Roark AA. Dissociating hippocampal and basal ganglia contributions to category learning using stimulus novelty and subjective judgments. NeuroImage. 2011; 55(4):1739–1753. http://doi.org/http://dx.doi.org/10.1016/j.neuroimage.2011.01.026. [PubMed: 21255655]

Shohamy D, Myers CE, Kalanithi J, Gluck MA. Basal ganglia and Dopamine Contributions to Probabilistic Category Learning. Neuroscience and Biobehavioral Reviews. 2008; 32(2):219–236. http://doi.org/10.1016/j.neubiorev.2007.07.008. [PubMed: 18061261]

Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging. 1998; 17(1):87–97. http://doi.org/10.1109/42.668698. [PubMed: 9617910]

Smith EE. The case for implicit category learning. Cognitive, Affective, & Behavioral Neuroscience. 2008; 8(1):3–16. http://doi.org/10.3758/CABN.8.1.3.

Smith EE, Grossman M. Multiple systems of category learning. Neuroscience & Biobehavioral Reviews. 2008; 32(2):249–264. http://doi.org/10.1016/j.neubiorev.2007.07.009. [PubMed: 17904637]

Smith JD, Boomer J, Zakrzewski A, Roeder J, Church BA, Ashby FG. Deferred Feedback Sharply Dissociates Implicit and Explicit Category Learning. Psychological Science. 2014; 25(2):447–457. http://doi.org/10.1177/0956797613509112. [PubMed: 24335605]

Travis SG, Huang Y, Fujiwara E, Radomski A, Olsen F, Carter R, Malykhin NV. High field structural MRI reveals specific episodic memory correlates in the subfields of the hippocampus. Neuropsychologia. 2014; 53:233–245. http://doi.org/10.1016/j.neuropsychologia.2013.11.016. [PubMed: 24296251]

Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, Avants BB. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. NeuroImage. 2014; 99:166–179. http://doi.org/10.1016/j.neuroimage.2014.05.044. [PubMed: 24879923]

Tustison NJ, Gee JC. N4ITK: Nick's N3 ITK implementation for MRI bias field correction. Insight Journal. 2009

Xu Y, D'Lauro C, Pyles JA, Kass RE, Tarr MJ. Fine-grained temporal coding of visually-similar categories in the ventral visual pathway and prefrontal cortex. Frontiers in Psychology. 2013; 4 http://doi.org/http://dx.doi.org/10.3389/fpsyg.2013.00684.

Yassa MA, Stark SM, Bakker A, Albert MS, Gallagher M, Stark CEL. High-resolution structural and functional MRI of hippocampal CA3 and dentate gyrus in patients with amnestic mild cognitive impairment. NeuroImage. 2010; 51(3):1242–1252. http://doi.org/http://dx.doi.org/10.1016/j.neuroimage.2010.03.040. [PubMed: 20338246]

Yushkevich PA, Pluta JB, Wang H, Xie L, Ding S-L, Gertje EC, Wolk DA. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. Human Brain Mapping. 2015; 36(1):258–287. http://doi.org/10.1002/hbm.22627. [PubMed: 25181316]

Zaki SR, Nosofsky RM. A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. Cognitive, Affective & Behavioral Neuroscience. 2001; 1(4):344–359.

Zeithamova D, Maddox WT, Schnyer DM. Dissociable Prototype Learning Systems: Evidence from Brain Imaging and Behavior. The Journal of Neuroscience. 2008; 28(49):13194–13201. http://doi.org/10.1523/JNEUROSCI.2915-08.2008. [PubMed: 19052210]

Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging. 2001; 20(1):45–57. http://doi.org/10.1109/42.906424. [PubMed: 11293691]

Highlights

- Elderly controls and Alzheimer's disease patients learned a novel visual category.

- Patients and healthy seniors alike learned features of the target category.

- Controls who learned from high-typicality exemplars had superior learning.

- Controls' and patients' learning was differently associated with left hippocampus.

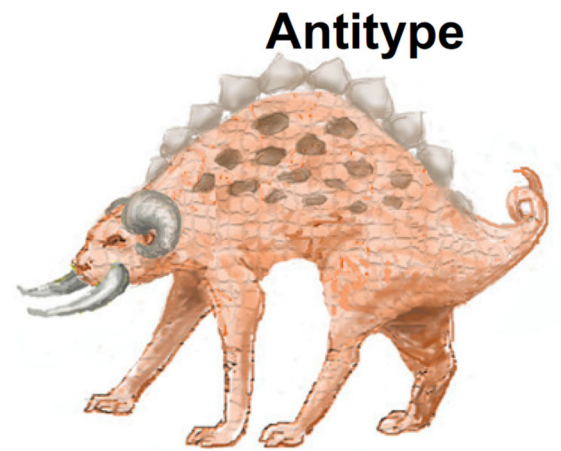- Learning was not affected by instruction type (explicit or implicit).
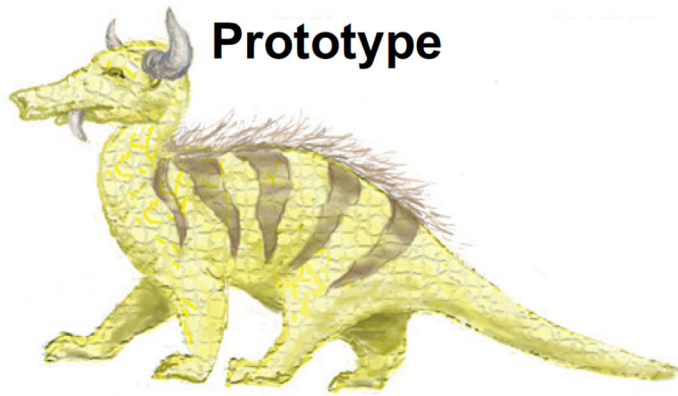
Figure 1.
Visual category exemplars: the crutter prototype and its antitype.
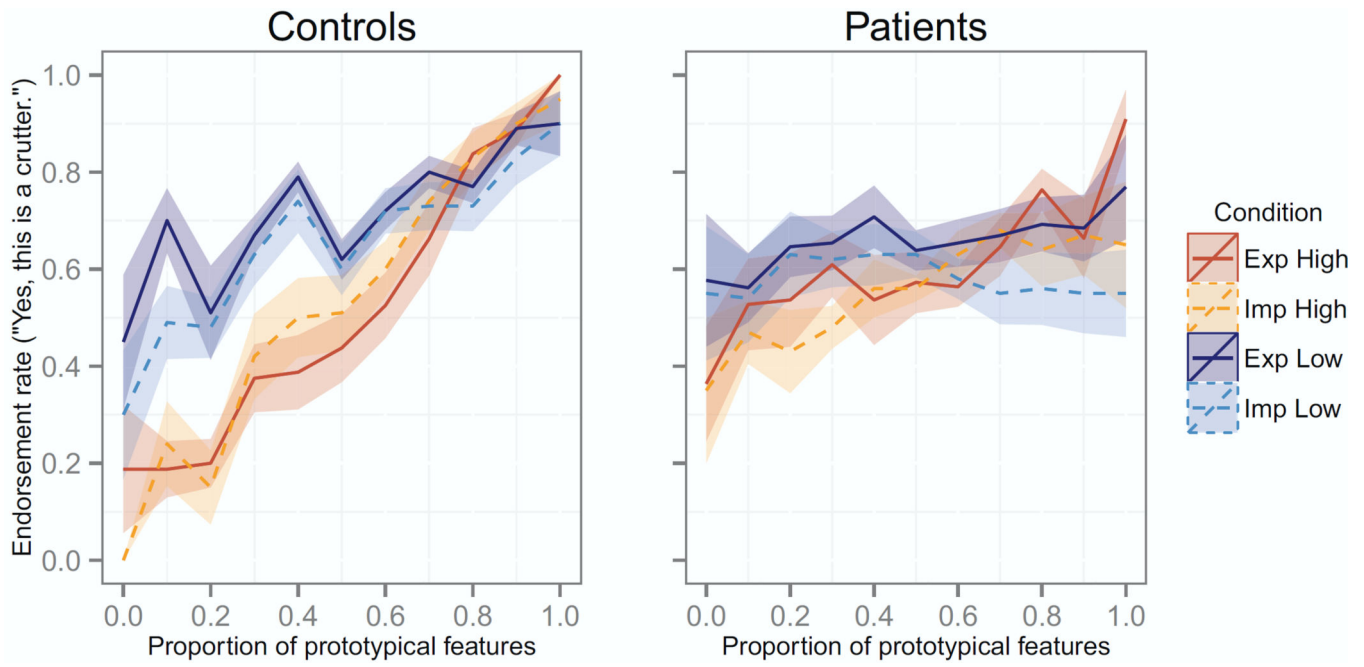
Figure 2.
Test item endorsement rates as a function of item typicality. Lines indicate the mean endorsement rate per condition; shaded regions indicate the standard error of the mean for each condition.
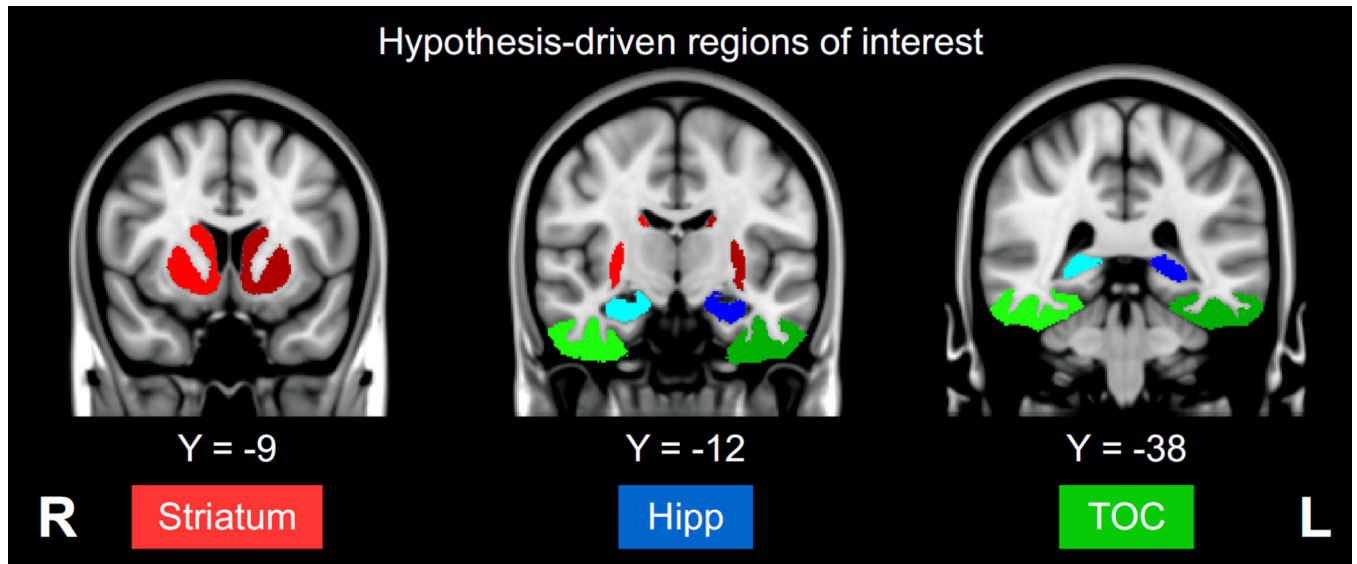
Figure 3.

Anatomical regions of interest (ROIs) from the OASIS-30 atlas, overlaid on the MNI-152 template brain. Mean grey matter density was calculated for each ROI and participant; these variables were used to predict individual learning scores in a multiple regression model, along with factors of group and training set. TOC (temporo-occipital cortex) includes fusiform and inferior temporal gyri; PFC (prefrontal cortex) includes middle and inferior frontal gyri. L=left, R=right.
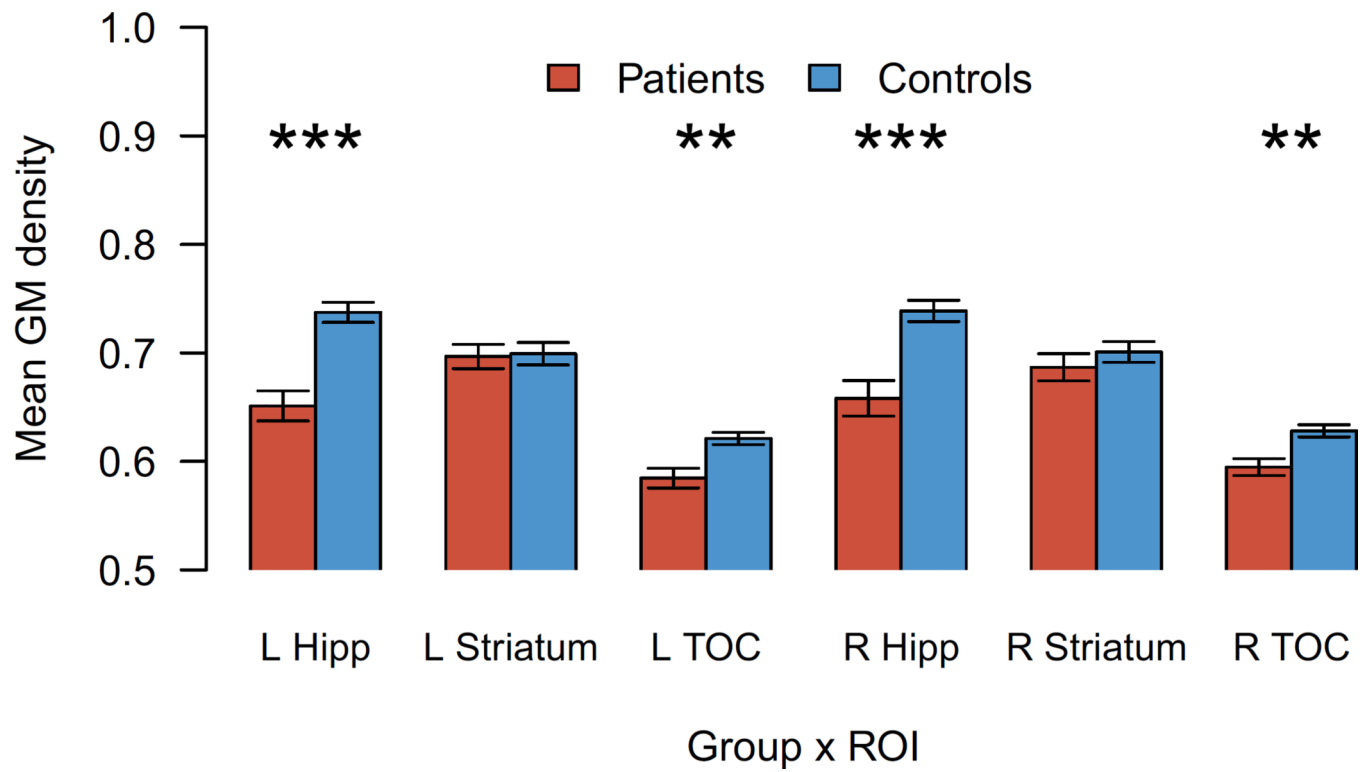
Figure 4.
Grey matter probability by group in anatomical regions of interest. Error bars represent the standard error of the mean. Hipp=hippocampus; TOC=temporo-occipital cortex; L=left; R=right. Two-sample, two-tailed t-test results, uncorrected for multiple comparisons: p<0.05*, p<0.01**, p<0.001***.
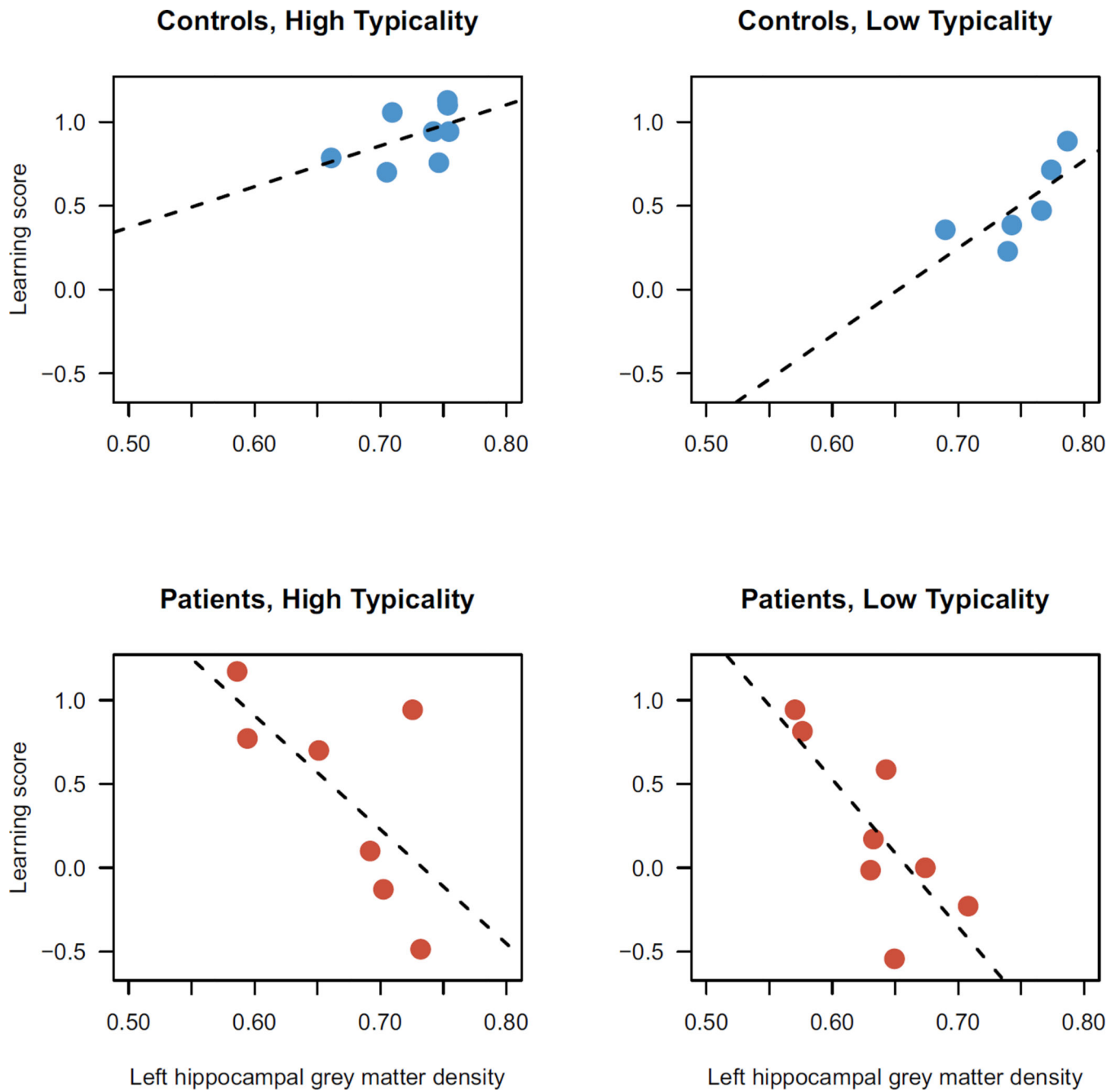
Figure 5.
Associations between learning score and grey matter density in left hippocampus. Each data point represents 1 participant. X-axis: mean grey matter density in anatomically-defined left hippocampus ROI. Y-axis: category learning score. The best-fit linear regression line is shown for each group.

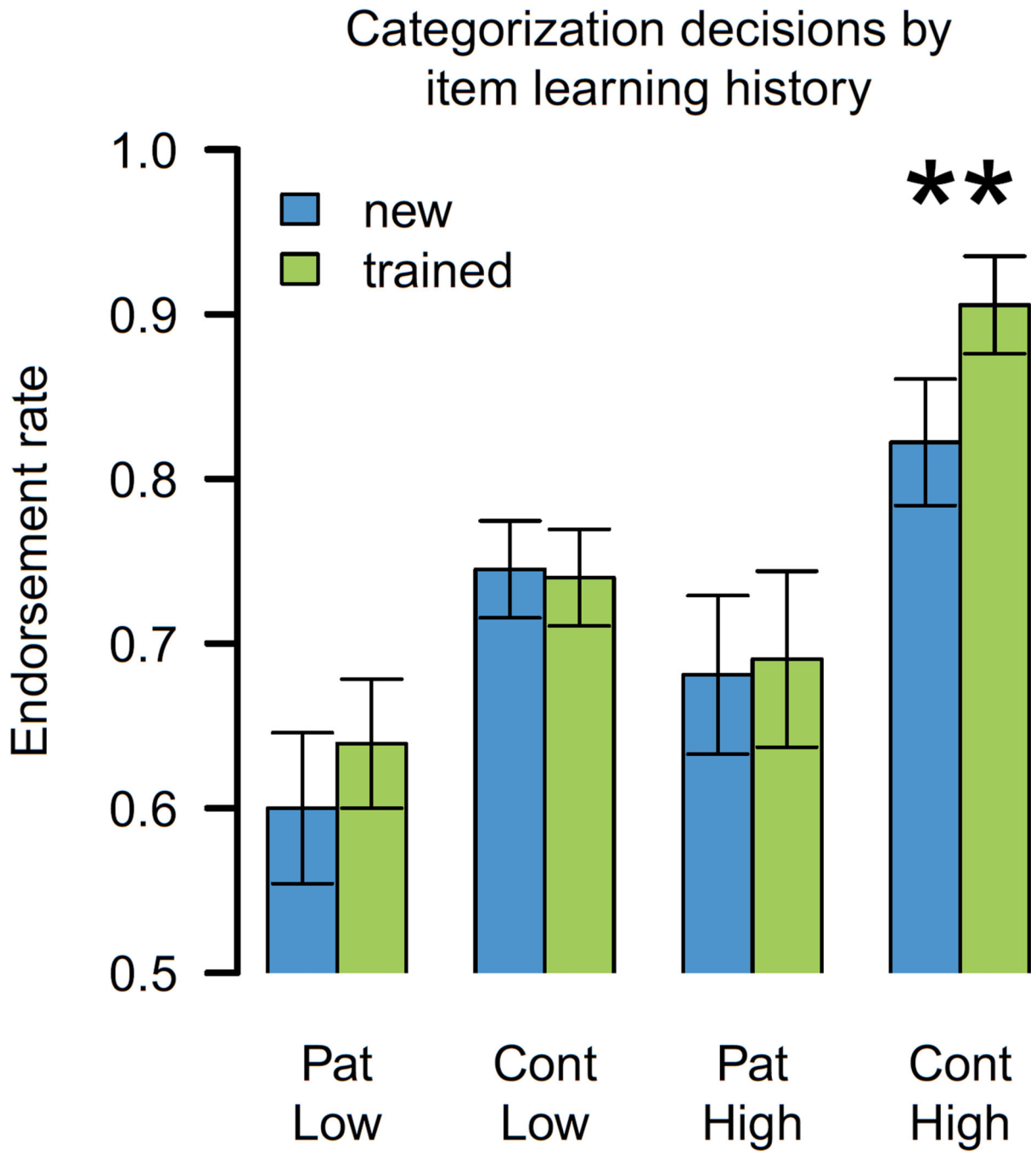## Categorization decisions by item learning history

Figure 6.
Learning history effects on test item categorization. Blue: items seen for the first time in the test phase; green: items previously encountered in the learning phase. Pat = patients; Cont=elderly controls; Low = low typicality learning set; High = high typicality learning set.

Table 1

Participant characteristics by condition, based on between-subjects design with factors of group (patients or controls), instruction (explicit or implicit), and learning set typicality (high or low).

Diagnosis	Learning set	Instruction	N	Disease duration	Age	Education	MMSE *	% female
Patients	Low	Exp	13	5.6 (2.0)	69.0 (9.3)	15.1 (2.7)	20.5 (5.6)	38
		Imp	10	6.0 (4.2)	65.7 (11.1)	15.3 (2.4)	18.0 (5.8)	30
	High	Exp	11	6.0 (3.6)	74.4 (7.8)	14.3 (2.2)	21.3 (3.8)	64
		Imp	10	4.7 (3.1)	70.9 (15.0)	14.8 (2.5)	21.3 (3.9)	40
Controls	Low	Exp	10	—	68.2 (12.8)	15.4 (2.6)	28.9 (0.8)	60
		Imp	10	—	64.1 (12.3)	15.4 (2.5)	28.9 (0.6)	60
	High	Exp	9	—	71.4 (9.0)	14.8 (2.9)	28.7 (1.2)	67
		Imp	10	—	65.5 (11.6)	15.9 (3.2)	28.7 (0.8)	40

*
Main effect of group (patients < controls).

Table 2

Stimulus feature values for category prototype and antitype.

Feature	Prototype	Antitype
Back	Flat	hump
Color	Yellow	red
Dorsal	Hair	plates
Face	Snout	pug
Horns	Steer	ram
Legs	Short	long
Neck	Vertical	downward
Pattern	Stripes	spots
Tail	Straight	curly
Teeth	Fang	tusk

Table 3

Mean category learning scores and endorsement rates for high- and low-typicality items. Learning scores were calculated for each individual as the regression coefficient relating stimulus items' proportion of typical features to their endorsement rate. Sample standard deviations for each condition are given in parentheses. T-statistics indicate contrast of observed learning score with model predictions. Asterisks indicate statistically significant differences between endorsement rates for high- and low-typicality test items.

Group	Learning set	Mean learning score (s.d.)	Endorsement of low-typicality items	Endorsement of high-typicality items	High–Low Difference
Patients	Low	0.04 (0.48)	0.59 (0.24)	0.63 (0.23)	t(22)=0.6, p<0.6
	High	0.29 (0.53)	0.48 (0.26)	0.69 (0.20)	t(20)=2.5, p<0.03*
Controls	Low	0.36 (0.32)	0.53 (0.24)	0.81 (0.12)	t(19)=5.2, p<0.001***
	High	0.88 (0.23)	0.20 (0.19)	0.87 (0.12)	t(18)=17.0, p<0.001***

Table 4

Self-reported strategies from recognition memory post-test. A subset of participants were asked how they knew whether items on the yes/no memory test had been previously seen. Responses were grouped according to whether they represented implicit or random approaches, such as guessing or intuitive responding, or explicit approaches, such as deliberate encoding of one or more stimulus features.

		Patient High (7 Exp, 4 Imp)	Patient Low (7 Exp, 7 Imp)	Control High (8 Exp, 4 Imp)	Control Low (4 Exp, 3 Imp)
Implicit or random approaches	Didn't remember, guessed	5	4	0	0
	Intuition, familiarity, novelty	1	2	0	0
	Single-feature strategy	0	4	1	1
Explicit approaches	Multi-feature strategy	3	3	11	6
	Other episodic strategy	1	1	0	0