Behavioral/Cognitive

Abstract Memory Representations in the Ventromedial Prefrontal Cortex and Hippocampus Support Concept Generalization

Caitlin R. Bowman and Dagmar Zeithamova
Department of Psychology, University of Oregon, Eugene, Oregon 97403

Memory function involves both the ability to remember details of individual experiences and the ability to link information across events to create new knowledge. Prior research has identified the ventromedial prefrontal cortex (VMPFC) and the hippocampus as important for integrating across events in the service of generalization in episodic memory. The degree to which these memory integration mechanisms contribute to other forms of generalization, such as concept learning, is unclear. The present study used a concept-learning task in humans (both sexes) coupled with model-based fMRI to test whether VMPFC and hippocampus contribute to concept generalization, and whether they do so by maintaining specific category exemplars or abstract category representations. Two formal categorization models were fit to individual subject data: a prototype model that posits abstract category representations and an exemplar model that posits category representations based on individual category members. Latent variables from each of these models were entered into neuroimaging analyses to determine whether VMPFC and the hippocampus track prototype or exemplar information during concept generalization. Behavioral model fits indicated that almost three-quarters of the subjects relied on prototype information when making judgments about new category members. Paralleling prototype dominance in behavior, correlates of the prototype model were identified in VMPFC and the anterior hippocampus with no significant exemplar correlates. These results indicate that the VMPFC and portions of the hippocampus play a broad role in memory generalization and that they do so by representing abstract information integrated from multiple events.

Key words: category learning; functional MRI; hippocampus; long-term memory

Significance Statement

Whether people represent concepts as a set of individual category members or by deriving generalized concept representations abstracted across exemplars has been debated. In episodic memory, generalized memory representations have been shown to arise through integration across events supported by the ventromedial prefrontal cortex (VMPFC) and hippocampus. The current study combined formal categorization models with fMRI data analysis to show that the VMPFC and anterior hippocampus represent abstract prototype information during concept generalization, contributing novel evidence of generalized concept representations in the brain. Results indicate that VMPFC–hippocampal memory integration mechanisms contribute to knowledge generalization across multiple cognitive domains, with the degree of abstraction of memory representations varying along the long axis of the hippocampus.

Introduction

Healthy memory function involves both the ability to remember details of individual experiences and the ability to generalize across experiences to create new knowledge. Memory for specific instances is known to be dependent on the hippocampus (Scoville and Milner, 1957; Tulving and Markowitsch, 1998; Eichenbaum, 2000). Generalization across experiences has traditionally been ascribed to other memory systems, such as the striatum (Knowlton et al., 1996), but recent research has shown hippocampal contributions to some forms of generalization, includ-

Received Sept. 25, 2017; revised Jan. 11, 2018; accepted Jan. 25, 2018.

Author contributions: C.R.B. and D.Z. designed research; C.R.B. performed research; C.R.B. analyzed data; C.R.B. and D.Z. wrote the paper.

Funding for this work was provided in part by the Lewis Family Endowment, which supports the Robert and Beverly Lewis Center for Neuroimaging at the University of Oregon (D.Z.); and by National Institute on Aging Grant F32-AG-054204 (C.R.B.).

The authors declare no competing financial interests.

Correspondence should be addressed to Dagmar Zeithamova, Department of Psychology, University of Oregon, Eugene, OR 97403. E-mail: dasa@uoregon.edu.

DOI:10.1523/JNEUROSCI.2811-17.2018

 0270-6474/18/382605-10$15.00/0

ing acquired equivalence (Shohamy and Wagner, 2008), concept learning (Zeithamova et al., 2008; Kumaran et al., 2009), and episodic inference (Preston et al., 2004; Zeithamova et al., 2012a; Pajkert et al., 2017). In episodic inference, the hippocampus contributes to generalization by interacting with the ventromedial prefrontal cortex (VMPFC) to integrate and encode current events in relation to prior knowledge (Schlichting and Preston, 2015). Such integrated memories then support novel inferential judgments (Zeithamova et al., 2012b), such as inferring that two children are siblings after seeing each of them with the same parent on separate occasions.

The degree to which hippocampal and VMPFC integration mechanisms identified in episodic inference contribute to other forms of generalization, such as concept learning, is unknown. Some evidence indicates that the hippocampus may not be critical for concept generalization, as individuals with episodic memory impairments can nonetheless learn some category structures and generalize to new examples (Knowlton and Squire, 1993; Filoteo et al., 2001). However, this evidence is not universally accepted (Zaki et al., 2003a; Zaki, 2004) and does not preclude hippocampal contributions to concept generalization in healthy brains. Indeed, VMPFC and hippocampal activation have been shown to track successful categorization judgments in healthy individuals (Zeithamova et al., 2008). While hippocampal activation during concept learning and generalization has been sometimes interpreted as evidence for the retrieval of specific exemplars (Koenig et al., 2008), recent evidence of generalized representations within the hippocampus (Collin et al., 2015; Schlichting et al., 2015) suggests that this interpretation of hippocampal activation may not always be accurate.

Integrating quantitative cognitive models with fMRI (O'Doherty et al., 2007) may help to resolve the nature of VMPFC and hippocampal contributions to concept generalization. Exemplar models posit that concepts are represented by individual instances and that classifying new stimuli involves retrieval and joint consideration of members of all relevant categories (Nosofsky, 1986; Kruschke, 1992). Prototype models posit that concepts are represented by their prototypes—generalized representations of the central tendencies abstracted across category exemplars (Posner and Keele, 1968; Reed, 1972; Homa et al., 1973). Fitting quantitative predictors derived from these models to neuroimaging data may clarify whether the hippocampus and VMPFC contribute to generalization by maintaining item-specific representations or by forming abstract concept representations.

Few studies to date have used model-based fMRI to specify representations underlying concept generalization (Nomura and Reber, 2012; Mack et al., 2013; Davis et al., 2017). Of particular relevance, Mack et al. (2013) found exemplar model correlates in lateral occipital and posterior parietal cortices, indicating that these regions contribute to categorization by representing individual category exemplars. Neither the hippocampus nor the VMPFC tracked either model, and no prototype correlates were identified. However, the lack of prototype correlates in that study could be driven by the category structure used: some exemplars were more similar to the prototype of the opposing category than to their own, making extraction of category prototypes difficult and less useful for task performance (Medin and Schaffer, 1978; Lamberts, 1995). Studies using training exemplars with more features and/or stronger coherence around prototypes have shown better prototype model fits (Smith and Minda, 1998; Minda and Smith, 2001), making them better suited for probing potential generalized representations in the hippocampus and VMPFC. In the present study, we aimed to elucidate the contributions of the VMPFC and hippocampus to concept generalization, using model-based fMRI in conjunction with a concept generalization task that engages these regions (Zeithamova et al., 2008) and where behavior indicated prototype formation.

Materials and Methods

Participants

Forty-two volunteers were recruited from the University of Oregon and the surrounding community and participated for financial compensation. Thirteen subjects were excluded from analyses for failing to complete the task (2 subjects), below-chance performance at the end of training and/or at categorization (5 subjects), structural abnormality (1 subject), and movement in excess of 2 mm within a run (5 subjects). This left 29 subjects (18 females; age, 18–28 years; mean age, 20.8 years; SD, 3.2 years) reported in all analyses. All participants provided written informed consent, were right handed, had learned English before 7 years of age, and were screened for neurological conditions and medications known to affect brain function. All experimental procedures were approved by Research Compliance Services at the University of Oregon.

Materials

Stimuli consisted of cartoon animals (Bozoki et al., 2006) that differed along the following eight binary dimensions: color (yellow/gray), shape of feet (clawed/webbed), shape of body (squared/circular), shape of tail (devil tail/feather tail), orientation of dots on body (vertical/horizontal), pattern on neck (stripes/thorns), head shape (with beak/with horn), and orientation of the head (forward/up; Fig. 1*A*). One stimulus was chosen randomly for each subject from a set of four possible prototypes to be the prototype of category A. The stimulus that shared no features with the category A prototype served as the category B prototype. The two possible versions of each feature can be seen on the two prototypes shown in Figure 1*A*. Physical distance between all stimuli was defined based on the number of differing features. Category A stimuli were those that share more features with prototype A than with prototype B. Category B stimuli were those that share more features with prototype B than with prototype A. Stimuli equidistant from the two prototypes were not used in the study.

Training set. The training set included four stimuli per category, each differing from the category prototype by two features (Table 1, training set structure; prototypes themselves were not presented during training). The general structure of the training set with respect to the category prototypes was the same across subjects, but, because the prototypes varied across participants, the specific stimuli and feature–category label associations also differed across participants.

Generalization set. Stimuli in the generalization phase included 58 unique stimuli. Eight new items were selected randomly at each distance from the category prototypes (e.g., eight items that were one feature from prototype A and thus seven features from prototype B, eight items that were two features from prototype A and thus six features from prototype B), excluding those equidistant from the two prototypes (four features from each). The generalization set also included the previously unseen prototypes and all training items. All training items were two features from their respective prototypes. Training items and category prototypes were each presented twice during generalization, whereas all other items were presented once. We repeated training items and prototypes to ensure dissociable predictions for the prototype and exemplar models for neuroimaging analyses as trials with old stimuli are particularly important under the assumptions of the exemplar model and the prototypes are particularly important under the assumptions of the prototype model. Given that there were only two prototypes, repeating prototypes also reduces noise in accuracy estimates for prototype classification (Kéri et al., 2001; Smith et al., 2008).

Experimental design

Participants first completed a feedback-based training session outside the scanner. Participants were told that they would learn to categorize members of two families of cartoon animals by sorting them and receiving feedback (Fig. 1*B*). Participants viewed individual stimuli on a computer

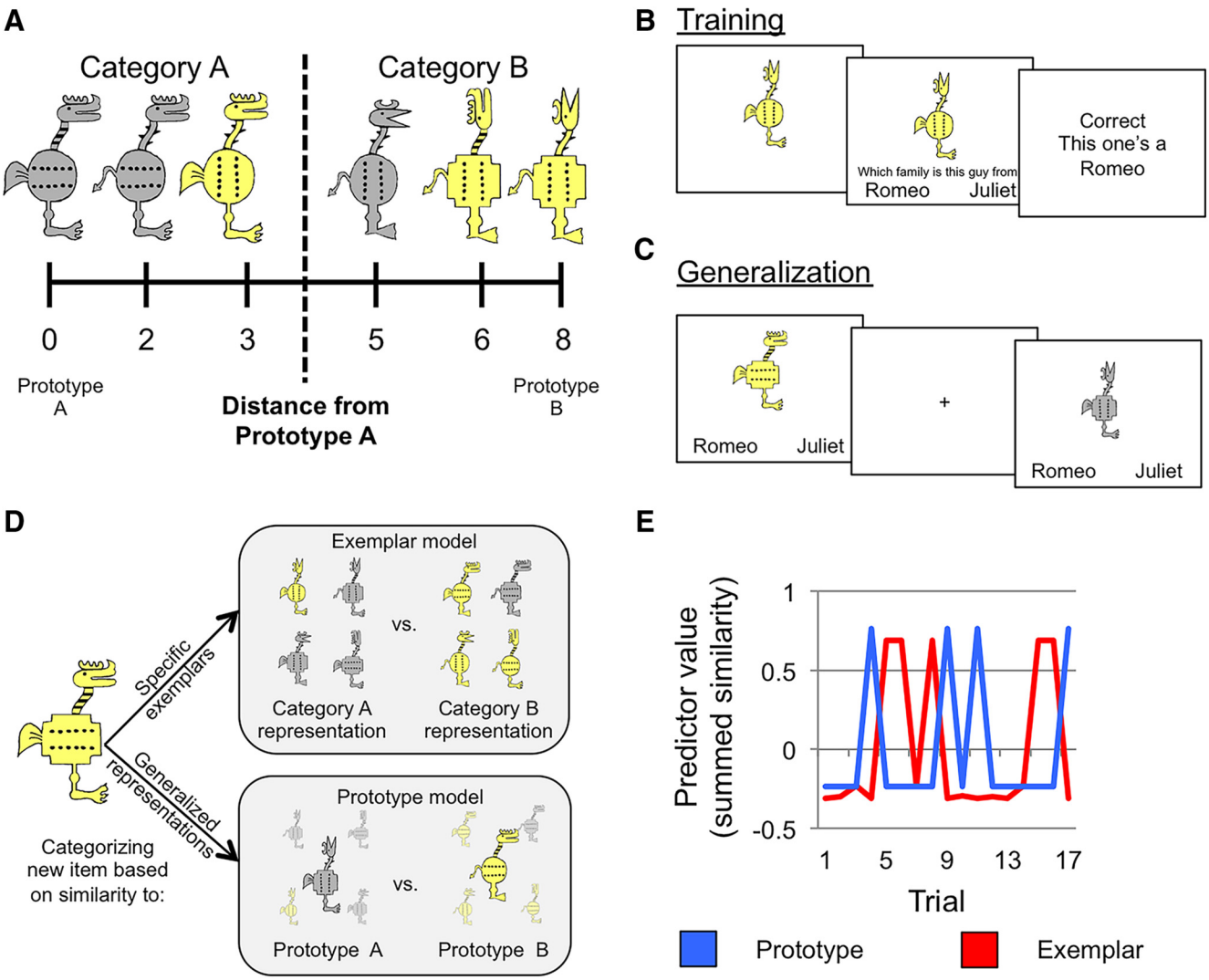

Figure 1. Prototype-learning task. ***A***, Example stimuli. The leftmost stimulus is the prototype of category A and the rightmost stimulus is the prototype of category B, sharing no features with prototype A. All stimuli in category A share more features with prototype A than the prototype B and vice versa. ***B***, Participants underwent feedback-based training outside of the scanner. ***C***, During a scanned generalization test, participants were asked to categorize old and new items without feedback. ***D***, Category representations and generalization to new items under assumptions of the prototype and exemplar models. Exemplar: categories are represented as individual exemplars. New exemplars are classified into the category with the most similar members. Prototype: categories are represented by their central tendencies (the category prototypes), and new exemplars are classified into the category with the most similar prototype. ***E***, Trial-by-trial summed similarity (mean centered) as predicted by the prototype (blue) and exemplar (red) models for one run in a representative subject. These values were entered as regressors into neuroimaging models as parametric modulators of the BOLD signal.

Table 1. Dimension values for example prototypes and training stimuli from each category

	Dimension values							
Stimulus	1	2	3	4	5	6	7	8
Prototype A	1	1	1	1	1	1	1	1
A1	1	1	0	1	1	0	1	1
A2	1	0	1	0	1	1	1	1
A3	0	1	1	1	1	1	1	0
A4	1	1	1	1	0	1	0	1
Prototype B	0	0	0	0	0	0	0	0
B1	0	0	1	0	0	1	0	0
B2	0	1	0	1	0	0	0	0
B3	1	0	0	0	0	0	0	1
B4	0	0	0	0	1	0	1	0

A1–A4 are individual training items from category A, and B1–B4 are individual training items from category B. Dimension values from each category prototype are presented with their corresponding category members. The version of each feature used for each stimulus is indicated by the feature dimension values (columns 1–8).

screen for 1 s, after which the prompt ("Guess which family this guy is from") and response options ("Romeo's" or "Juliet's") appeared on the screen under the stimulus, and participants made a self-paced judgment. Participants were then given feedback as to whether they were correct or wrong, and the correct family was displayed (e.g., "Correct", "This one was from Romeo's family"). There were five blocks of training, each containing six repetitions of all training items with self-paced breaks between the blocks. The order of stimuli was randomized within each block with the constraint that no more than three items of the same category could be presented consecutively.

Participants entered the scanner shortly after training. They first completed a resting scan lasting 5 min followed by two runs of passive viewing of training items and new items (data from these scans are not reported in this manuscript). Four runs of the concept generalization task followed (Fig. 1*C*). Each run consisted of 17 trials with a 5 s stimulus presentation and 7 s intertrial interval (ITI). Participants classified each item into one of the two categories (labeled Romeo's and Juliet's) using a button press while the stimulus was on the screen. Anatomical images were collected following categorization. Following the scan, participants were asked

about their strategy during training and then indicated which version of each feature they thought was most typical of each category (i.e., Did most Romeos have a head that was up or a head that was forward?). Last, participants were verbally debriefed about the study.

fMRI data acquisition

Scanning was completed on a 3 T Siemens MAGNETOM Skyra scanner at the University of Oregon Lewis Center for Neuroimaging using a 32-channel head coil. Head motion was minimized using foam padding. The scanning session started with a localizer scan followed by seven functional runs using a multiband gradient echo pulse sequence [TR = 2000 ms; TE = 26 ms; flip angle = 90°; matrix size = 100 × 100; 72 contiguous slices oriented 15° off the anterior commissure–posterior commissure line to reduced prefrontal signal dropout; interleaved acquisition; FOV = 200 mm; voxel size = 2.0 × 2.0 × 2.0 mm; generalized autocalibrating partially parallel acquisitions (GRAPPA) factor = 2]. One hundred fifty volumes were collected for the resting-state run, 100 volumes for each passive-viewing run, and 106 volumes for each categorization run. A standard high-resolution T1-weighted MPRAGE anatomical image (TR = 2500 ms; TE = 3.43 ms; TI = 1100 ms; flip angle = 7°; matrix size = 256 × 256; 176 contiguous slices; FOV = 256 mm; slice thickness = 1 mm; voxel size = 1.0 × 1.0 × 1.0 mm; GRAPPA factor = 2) was collected following all functional runs. Scanning concluded with a custom anatomical T2 coronal image (TR = 13,520 ms; TE = 88 ms; flip angle = 150°; matrix size = 512 × 512; 65 contiguous slices oriented perpendicularly to the main axis of the hippocampus; interleaved acquisition; FOV = 220 mm; voxel size = 0.4 × 0.4 × 2 mm; GRAPPA factor = 2).

Statistical analysis

Behavioral accuracies. Training data were examined with a one-way, repeated-measures ANOVA comparing accuracies across the five training blocks. A significant positive linear effect of block was used to evaluate whether significant learning occurred across the group during training. For the generalization phase, categorization accuracies were computed for new items (separately at each distance 0, 1, 2, and 3 to the prototypes) and for training items (all two features from their respective prototypes). A one-way repeated-measures ANOVA was used to compare accuracies across all distances from category prototypes for new items. A significant negative linear effect (i.e., better accuracy closer to prototypes) was used to test for a typicality effect (Rosch et al., 1976). Paired-samples *t* tests comparing categorization accuracies for old items and for new items at distance 2 from their respective prototypes was used to test for categorization advantage for training items compared with new items of the same typicality.

Additionally, we examined the relationship between categorization performance and explicit knowledge of typical feature values for each category (from the debriefing questionnaire). We computed Pearson's correlations between overall accuracy on the debriefing questionnaire and categorization accuracy, separately for old and new items.

Prototype and exemplar model fitting. Prototype and exemplar models were fit to trial-by-trial categorization data in individual subjects. The conceptual representations of the models are depicted in Figure 1*D*. Prototype models assume that categories are represented by their prototypes (i.e., the combination of typical category features from all training items in each category). Consistent with prior modeling literature (Minda and Smith, 2001; Maddox et al., 2011), the similarity of each categorization stimulus to each prototype was computed, assuming that perceptual similarity is an exponential decay function of physical similarity (Shepard, 1957) and taking into account potential differences in attention to individual features. Formally, the relationships were computed as follows:

$$S_A(x) = \exp[-c\sum_{i=1}^{m}(w_i|x_i - \text{proto}_{Ai}|^r)^{1/r}], \quad (1)$$

where $S_A(x)$ is the similarity of item x to category A, x_i represents the value of the item x on the ith dimension of its m binary dimensions (m = 8 in our study), proto_A is the prototype of category A, r is the distance metric (fixed at 1 for the city-block metric for the binary dimension stimuli). Parameters that were estimated from the pattern of behavioral responses, separately for each participant, were w (a vector with eight weights, one for each of the eight stimulus features) and c (sensitivity; the rate at which similarity declines with distance). More details of the parameter estimation procedure will follow after the description of the exemplar model.

Exemplar models assume that categories are represented by their exemplars, and test items are classified into the category with the highest summed similarity across category exemplars. Formally (Nosofsky, 1987; Zaki et al., 2003b), similarity of an item x to category A is computed as follows:

$$S_A(x) = \sum_{y\in A}\exp[-c\sum_{i=1}^{m}(w_i|x_i - y_i|^r)^{1/r}], \quad (2)$$

where y represents the individual training stimuli from category A, and the remaining notation and parameters are as in Equation 1.

For both models, the probability of assigning a stimulus x to category A is equal to the similarity to category A divided by the summed similarity to categories A and B, formally, as follows:

$$P(A|x) = \frac{S_A(x)}{S_A(x) + S_B(x)}. \quad (3)$$

Using these equations, the best fitting w_{1-8} (attention to each feature) and c (sensitivity) parameters were estimated from the behavioral data of each participant, separately for each model. For each trial, the probability of the participant's response under the assumptions of each model was computed. For a given set of model parameters (w_{1-8}, c), there will be a specific probability value for each trial. These trial-by-trial model predictions are then compared with the participant's actual series of responses. For example, if the participant chose category A on a trial where the model predicted a 70% chance of picking category A, then there is an error of 30%. Model parameters (w_{1-8}, c) are then tuned so that the model predictions are as close as possible to the actual observed pattern of responses. Specifically, an error metric (negative log likelihood of the whole sequence of responses) was computed for each model by summing the negative of log-transformed probabilities. This summed value was minimized by adjusting w attention weights and c sensitivity parameters using standard maximum likelihood methods, implemented using the "fminsearch" function in MATLAB (MATLAB 2015a, MathWorks). Parameters for each model and each participant were optimized separately as there are currently no procedures developed for trial-by-trial behavioral fits of both models simultaneously. After optimization, prototype and exemplar model fits were (1) compared between models across the group using a paired-samples *t* test to determine whether the group as a whole was better fit by the prototype or exemplar model; (2) compared within each participant to each other and to chance using Monte Carlo simulations (described in the following paragraph) to determine for each participant whether they used a prototype or exemplar strategy; and (3) used to generate neuroimaging regressors (described in the fMRI analysis section) to identify regions tracking predictions of each model.

To classify individual participants as "exemplarists" or "prototypists," we tested whether one model fit reliably better than the other on an individual participant basis using a Monte Carlo simulation. For each subject, a vector of random responses to the actual sequence of stimuli observed by a given participant was generated and used to fit both prototype and exemplar models as described above. This procedure was repeated 10,000 times to generate a subject-specific null distribution of model fits for each model. We then compared the observed prototype and exemplar model fits to this null distribution to determine whether one or both models fit the participant's data better than chance. This was determined by comparing the actually observed model fit to the null distribution of fits and testing whether the observed model fit appeared by chance with a frequency of <5% ($p < 0.05$, one-tailed). Both models showed above-chance performance in all subjects. To determine whether one model fit reliably better than the other, we compared the observed difference in model fits to the null distribution of differences in model fits generated by the Monte Carlo simulation. One model was deemed a winner for the given participant when that difference score appeared by chance with a frequency of <5% ($p < 0.05$, two-tailed).

fMRI preprocessing. Raw dicom images were converted to Nifti format using the dcm2nii function from MRIcron (https://www.nitrc.

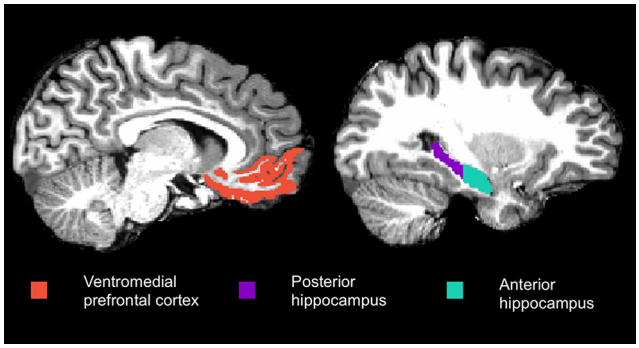

Figure 2. Regions of interest from a representative subject. Regions were defined in the native space of each subject using automated segmentation in Freesurfer.

org/projects/mricron). Functional images were skull stripped using BET (brain extraction tool), which is part of FSL version 5.0.9 (www.fmrib.ox.ac.uk/fsl). Motion correction was computed within each functional run using MCFLIRT in FSL to realign all volumes to the middle volume. Across-run realignment was computed using ANTs (Advanced Normalization Tools; http://stnava.github.io/ANTs/) with the first functional volume serving as the reference volume. The first volumes of all other runs were registered to the reference volume, and the transformation computed was applied to all other images in the run. Brain-extracted and motion-corrected images from each categorization run were entered into the FEAT (fMRI Expert Analysis Tool) in FSL for high-pass temporal filtering (100 s) and spatial smoothing using a 4 mm FWHM kernel. For whole-brain group analyses, functional data were registered to standard stereotaxic space following coregistration with the T1 image using the FLIRT (FMRIB's Linear Image Registration Tool) in FSL.

Regions of interest. Regions of interest (ROIs; Fig. 2) were defined anatomically in individual subjects' native space using the cortical (VMPFC) parcellation and subcortical (hippocampus) segmentation from Freesurfer version 6 (https://surfer.nmr.mgh.harvard.edu/) and collapsed across hemispheres. VMPFC (medial orbitofrontal label in Freesurfer) was expected to track prototype predictors based on its role in episodic generalization (Zeithamova et al., 2012a) and some types of categorization (Zeithamova et al., 2008). Past research has indicated that generalization effects in the hippocampus may be specific to the anterior hippocampus while the posterior hippocampus maintains event-specific representations (Collin et al., 2015; Schlichting et al., 2015). As such, we tested for anterior/posterior differences by dividing each subject-specific hippocampal ROI into an anterior half and a posterior half, splitting at the middle slice. When a participant had an odd number of hippocampal slices, the middle slice was assigned to the posterior hippocampus. The anterior hippocampus was expected to track prototype predictors, whereas the posterior hippocampus was expected to track exemplar predictors.

Model-based fMRI analysis. Data were modeled using the GLM. Three regressors were included in each model: one for all trial onsets, one that included modulation for each trial by prototype model predictions, and one that included modulation for each trial by exemplar model predictions. The regressor for all trial onsets was included with the model regressors to account for activation that is associated with performing a categorization task generally, but does not track either model specifically. The model-based fMRI predictors were computed as the summed similarity to both categories, formally the denominator in Equation 3 under the category representation assumptions of each model (as formalized in Eqs. 1 and 2). Because summed similarity indexes how similar the current item is to the existing category representations as a whole (regardless of which category it is closer to), it has been used in model-based studies for the localization of regions that contain such category representations. The summed similarity metric has also been previously called "recognition strength" (Davis et al., 2014) or "representational match" (Mack et al., 2013). Including exemplar and prototype predictors in the same GLM accounts for shared variance between the model predictions. Thus, the model adopted here ensures that only regions that show trial-by-trial response variability predicted by the prototype or exemplar model are identified when comparing each to baseline.

All test items were included in the GLM without an imposed distinction between old and new items because the models already differ in how they treat these two types of items. Exemplar models generally predict better classification of old items compared with new items at the same level of typicality, whereas prototype models do not differ in their treatment of old and new items when they are the same distance from the prototypes. Likewise, we did not include correct/incorrect explicitly in the GLM because the model-fitting process involves predicting an individual subject's pattern of responses, including experimenter-defined errors (e.g., a given classification response may be consistent with a particular participant's exemplar representation, even though it does match the experimenter-defined category). We also did not impose a distinction between individual subjects whose behavioral data were best fit by the prototype model compared with those who were not because of the following: (1) the neuroimaging regressors that we generated were unique to each subject and driven by their own responses; (2) it is possible that subjects maintain both prototype and exemplar representations even when one dominates behavioral responses in the majority of participants; and (3) the number of exemplarists in the current study would be too low to permit group comparisons based on the best fitting strategy. However, exploratory analyses split by strategy showed generally similar patterns of neural model fits, with numerically more pronounced prototype correlates in participants identified as prototypists. Events were modeled with a 5 s duration, which was the fixed length of the stimulus presentation, regardless of the timing of participant's response. Onsets were then convolved with the canonical hemodynamic response function as implemented in FSL (gamma function with a phase of 0 s, an SD of 3 s, and a mean lag time of 6 s).

ROI analyses were computed in native space. The mean parameter estimate from each region was extracted and divided by the SD of parameter estimates to compute the effect size of how much the BOLD signal tracked each model predictor. Normalizing the β-values by their error of estimate provides a mean to deweigh values associated with large uncertainty, similar to how lower-level variance estimates are used in group analyses as implemented in FSL (Smith et al., 2004). However, using raw parameter estimates did not affect the observed pattern of results. Normalized β-values were then averaged across runs and submitted to group analyses. Given our a priori predictions that the VMPFC and anterior hippocampus would track prototype predictors, we computed one-sample, one-tailed t tests to determine whether prototype effects were significantly greater than zero. A similar procedure was used to test for significant exemplar effects in the posterior hippocampus. Paired-sample t tests were used to determine whether a given region tracked one model predictor more than the other. For these ROI analyses, the α-level was set to $p < 0.016$ to correct for multiple comparisons across the three ROIs. To test for a representational dissociation along the anterior–posterior axis of the hippocampus (Collin et al., 2015; Schlichting et al., 2015), we computed a 2 (model: prototype, exemplar) $\times$ 2 (hippocampal ROI: anterior, posterior) repeated-measures ANOVA on the model-fit values and tested for an interaction.

Whole-brain maps in normalized space were generated for exploratory purposes. Parameter estimates were averaged across runs within individual subjects using a fixed-effects analysis. Group-level contrasts were computed using Mixed-Effects FLAME 1 in FSL, comparing the prototype and exemplar regressors to baseline (i.e., the unmodeled ITI during which the fixation cross was presented) to identify regions tracking the predictions of each model. Whole-brain maps were computed with a voxelwise threshold of $z > 3.1$ and cluster corrected at $p < 0.05$ using the false discovery rate (FDR) method in FSL.

Results

Behavioral

Categorization accuracy (Fig. 3*A*) increased linearly across training (repeated-measures ANOVA linear effect: $F_{(1,28)} = 137.67$, $p < 0.001$, partial $\eta^2 = 0.83$). At the generalization test, the mean overall accuracy was 86.2% (SD, 10.9%) on the training items

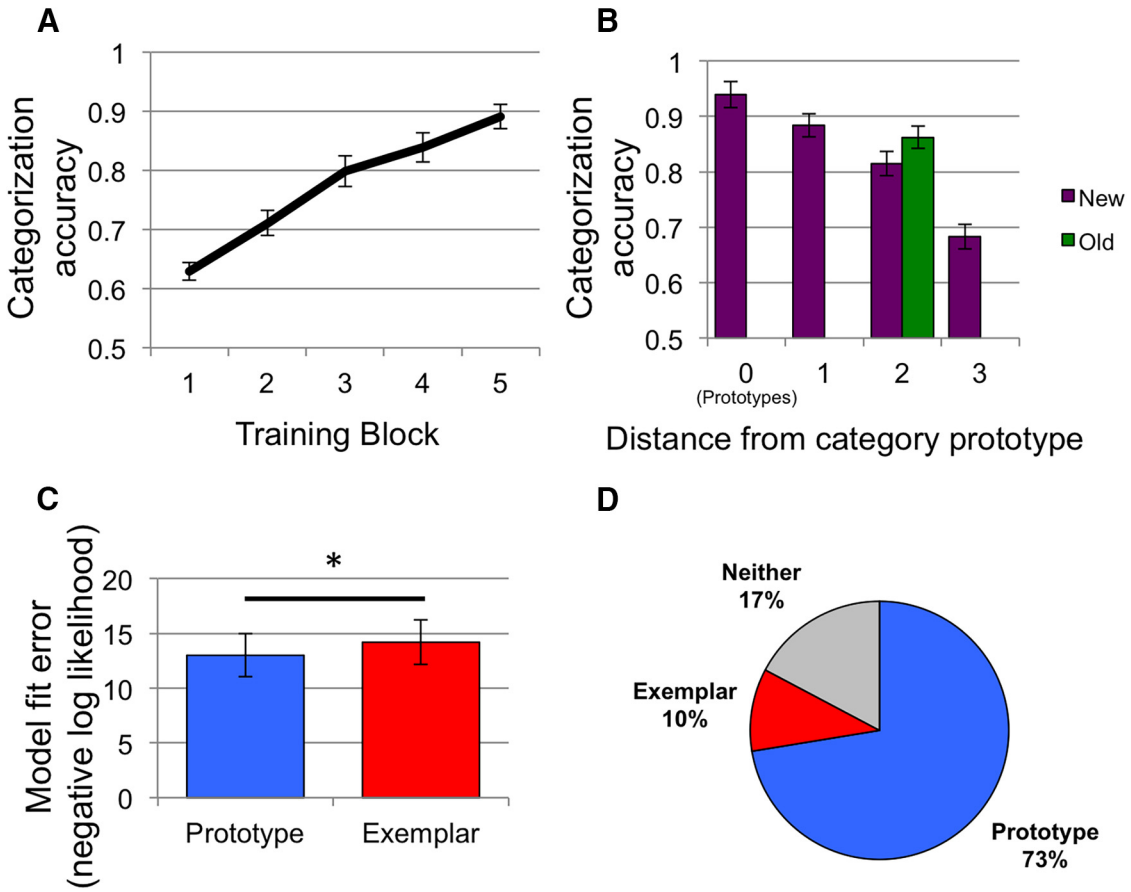

Figure 3. Behavioral results. ***A***, Proportion of correct categorization responses across five training blocks. ***B***, Proportion of correct responses during the generalization test for test items at each distance from their respective category prototype. Separate accuracies are presented for training items (old; all distance 2) and new items. ***C***, Model fit errors (negative log likelihood) for the prototype and exemplar models. Lower values indicate better fit. Asterisk represents a significant ($p < 0.05$) paired-samples difference in mean fit error between the models. In ***A–C***, error bars represent across-subject SEM. ***D***, The percentage of individual subjects best fit by the prototype model (blue), the exemplar model (red), and those for whom model fits did not differ significantly from one another (neither, gray).

and 80.5% (SD, 7.9%) on new transfer items. Categorization accuracy for all test items split by their typicality (distance to the prototypes) is presented in Figure 3*B*. A one-way, repeated-measures ANOVA on categorization accuracy for new stimuli across the four distances showed a significant linear effect ($F_{(1,28)} = 67.05, p < 0.001$, partial $\eta^2 = 0.71$) with better accuracy for items closer to its category prototype. A paired *t* test comparing accuracy on old items and new items at the same distance to the prototypes showed significantly better accuracy for old items ($t_{(28)} = 2.26, p = 0.03, d = 0.42$).

A paired-samples *t* test comparing prototype and exemplar model fits showed significantly better fit (i.e., lower negative log likelihood) across the group for the prototype model ($t_{(28)} = 3.61, p = 0.001, d = 0.67$; Fig. 3*C*). Results from the Monte Carlo simulation showed that the prototype model significantly outperformed the exemplar model in 21 subjects, the exemplar model outperformed the prototype model in 3 subjects, and the fit for the two models did not differ reliably in 5 subjects (Fig. 3*D*).

In the debriefing, the mean accuracy in identifying the most common version of the features for each category was 80.6% across the entire group (SD, 17.7%; range, 37.5–100%). Further, features that participants paid the most attention to (as estimated by the models) were also those that the participant had the best explicit knowledge of; attention weight estimates generated by each model were significantly higher for features labeled correctly during the debriefing compared with those labeled incorrectly (exemplar: $t_{(19)} = 3.11, p = 0.006$; prototype: $t_{(19)} = 4.76, p < 0.001$). Accuracy on the debriefing measure was not correlated with classification accuracy for training (old) items presented during the generalization test ($r = 0.03, p = 0.88$), but it was positively related to accuracy for new items ($r = 0.42, p = 0.02$), meaning that explicit knowledge of which feature values were associated with each category did benefit generalization.

Model-based fMRI

ROI analysis

Prototype and exemplar parameter estimates (normalized β values) for each ROI are depicted in Figure 4*A*. One-sample *t* tests showed above-chance prototype correlates in the VMPFC ($t_{(28)} = 3.55, p < 0.001, d = 0.66$) and anterior hippocampus ($t_{(28)} = 2.32, p = 0.014, d = 0.43$). A paired *t* test in the VMPFC showed greater prototype than exemplar correlates ($t_{(28)} = 2.33, p = 0.014, d = 0.43$), whereas the difference in model fits in the anterior hippocampus did not reach significance ($t_{(28)} = 1.50, p = 0.073, d = 0.28$). A one-sample *t* test of exemplar correlates in the posterior hippocampus was not significant ($t_{(28)} = 1.10, p = 0.14, d = 0.20$) nor was the paired comparison between exemplar and prototype effects ($t_{(28)} = 1.01, p = 0.16, d = 0.19$). To test whether model fits in the anterior hippocampus differed from those in the posterior hippocampus, we performed a 2 (model: prototype, exemplar) × 2 (hippocampal ROI: anterior, posterior) repeated-measures ANOVA, which showed no main effect of

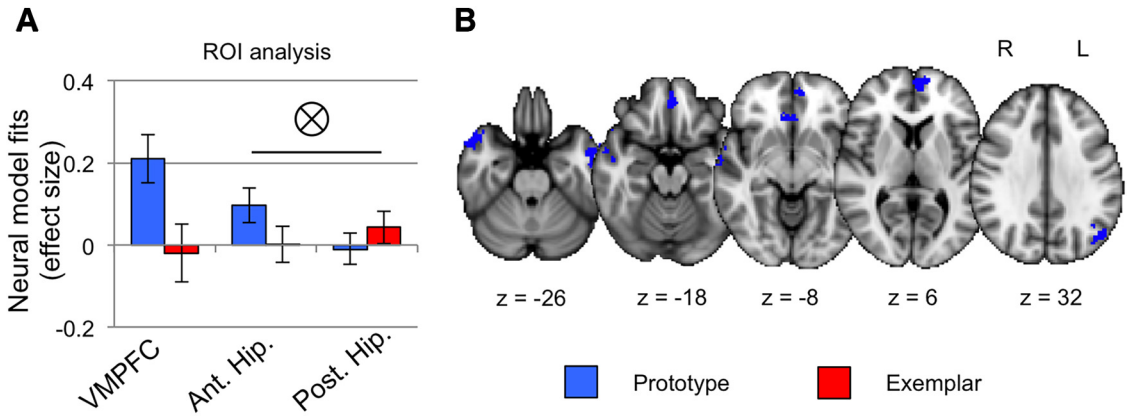

Figure 4. Model-based fMRI. ***A***, Prototype (blue) and exemplar (red) neural model fits in the three ROIs. Effect size is the mean/SD of β-values within each ROI, averaged across runs. Ant., Anterior; Post., posterior; Hip., hippocampus. Error bars represent the across-subject SEM. Tensor symbol represents a significant interaction between model fits and hippocampal ROIs (anterior vs posterior). ***B***, Representative slices from a whole-brain prototype > baseline contrast denoting regions in which activation reliably tracked predictors derived from the prototype model. No suprathreshold clusters of activation were found for the exemplar > baseline contrast. R, Right; L, left; z, *z*-coordinate in MNI standard space.

model ($F_{(1,28)} = 0.16$, $p = 0.69$, partial $\eta^2 = 0.006$), no main effect of hippocampal ROI ($F_{(1,28)} = 1.77$, $p = 0.20$, partial $\eta^2 = 0.06$), but a significant model × hippocampal ROI interaction ($F_{(1,28)} = 6.63$, $p = 0.02$, $\eta^2 = 0.19$).

Whole-brain analysis

Whole-brain activation maps showing regions tracking prototype and exemplar predictors are presented in Figure 4*B*, with cluster information in Table 2. There were seven clusters significantly tracking prototype predictors, including clusters in VMPFC, bilateral middle temporal gyrus, bilateral temporal pole, and left superior lateral occipital cortex. No regions tracked exemplar predictors at this threshold. Because a prior study identified exemplar correlates when using a category structure that resulted in stronger exemplar strategy use (Mack et al., 2013), we wanted to evaluate whether exemplar correlates could be identified if we relaxed the threshold. At a lenient threshold ($z = 2$; FDR cluster correction, $p = 0.1$), exemplar correlates were identified in left lateral occipital cortex (peak: MNI coordinates, −22, −88, −22; $z = 3.23$; 1376 voxels), right lateral occipital cortex (peak: MNI coordinates, 44, −86, −4; $z = 3.78$; 1346 voxels), precuneus (peak: MNI coordinates, 2, −64, 56; $z = 3.56$; 571 voxels), right postcentral gyrus/inferior parietal cortex (peak: MNI coordinates, 46, −32, 54; $z = 3.53$; 454 voxels), and left postcentral gyrus/inferior parietal cortex (peak: MNI coordinates, −58, −22, 48; $z = 3.04$; 423 voxels).

Table 2. Regions significantly tracking prototype model predictors

Region	Hemisphere	Cluster size	*z*-statistic	Peak coordinate		
				x	*y*	*z*
Frontal pole	L	315	4.78	−10	62	14
VMPFC	M	124	4.2	0	44	−16
Anterior cingulate cortex	R	65	4.14	4	30	−10
Temporal pole	L	266	4.13	−44	16	−32
Temporal pole	R	269	4.34	58	12	−30
Middle temporal gyrus	L	98	4.08	−62	−14	−12
Superior lateral occipital cortex	L	131	4.49	−46	−70	30

Cluster size is the number of voxels; peak coordinate is given in MNI space. L, Left; R, right; M, medial.

Discussion

We tested whether hippocampal and VMPFC processes that support generalization in episodic memory do so in concept learning. Furthermore, we aimed to determine whether such generalization relies on abstract category representations (i.e., prototypes) or item representations (i.e., exemplars). Participants learned to classify binary-feature stimuli into two categories, each organized around a prototype containing the version of each feature most typical of the category. While undergoing fMRI scanning, participants completed a generalization phase where they classified training stimuli and new items that had not been given an explicit category label. Fitting formal prototype and exemplar models to behavior revealed that most participants relied on prototype representations abstracted across the training set to make their generalization judgments. The dominance of the prototype model in behavior was accompanied by prototype correlates in the VMPFC and anterior hippocampus, suggesting that these regions contribute to concept generalization by representing abstract category information. These results indicate that memory integration mechanisms supported by the VMPFC and hippocampus contribute to generalization across multiple cognitive domains.

In episodic inference, the VMPFC has been shown to form generalized representations via memory integration processes that link information across episodes (for review, see Schlichting and Preston, 2017). The VMPFC also tracks generalization success in a task similar to that in the current study (Zeithamova et al., 2008). However, the computations or representations reflected in categorization-related VMPFC activity were unclear. By linking neural activation to predictions from formal categorization models, we show that the VMPFC tracks prototype-based model predictors, indicating that it contributes to classification by representing abstract category information. We propose that the VMPFC may form prototype representations by integrating information across exemplars that share a category label, which are then accessed to inform generalization judgments. Such a role for the VMPFC is consistent with episodic memory studies showing that the VMPFC supports integration of current experience with prior knowledge (Zeithamova et al., 2012a; Richter et al., 2016; Liu et al., 2017), facilitating memory for schema-consistent information (Tse et al., 2007, 2011; van Kesteren et al., 2010) and inference of new relationships across overlapping events (DeVito et al., 2010; Zeithamova et al., 2012a; Schlichting et al., 2015). Further, prior research has shown that the VMPFC contributes to

decision-making in novel situations by integrating across relevant experiences (Behrens et al., 2008; Barron et al., 2013) and may play a larger role in relating memory representations to current decision-making demands (Kaplan et al., 2017). Together, these results suggest that the VMPFC links information across episodes to represent abstract information not experienced directly, playing an important role in knowledge generalization across multiple cognitive domains.

The present results also revealed portions of the hippocampus that tracked prototype predictors during generalization. While many theories have posited that learning systems outside the medial temporal lobes are the primary drivers of category learning (Knowlton and Squire, 1993; Ashby et al., 1998; Shohamy et al., 2004), the current study and other evidence (Zeithamova et al., 2008; Kumaran et al., 2009) demonstrate a clear role for the hippocampus in concept generalization. These results challenge traditional multiple-systems views that posit a division of labor between medial temporal systems supporting memory for specific events and other learning systems for concept formation (Squire and Knowlton, 1995). Furthermore, evidence for generalized (prototype) concept representations in the hippocampus also challenges single-system views that posit hippocampal generalization based on representations of individual category members (Medin and Schaffer, 1978; Nosofsky, 1986; Koenig et al., 2008). Instead, the results suggest an update to multiple-systems views to incorporate growing evidence that the hippocampus contributes to several forms of memory generalization in addition to its well known role in memory for specific events.

Our results also demonstrated differences in concept representations along the long axis of the hippocampus, with prototype–model correlates specific to the anterior hippocampus. The existence of abstract concept representations within anterior but not posterior hippocampus is consistent with previous studies showing integrated representations in this region during associative inference (Schlichting et al., 2015) and when linking scenes to form large-scale narratives (Collin et al., 2015). The anterior hippocampus may be better suited than posterior hippocampus to forming such generalized representations because its cells have larger receptive fields that may better facilitate integration across time and space (Kjelstrup et al., 2008; Stensola et al., 2012; Poppenk et al., 2013). Anterior hippocampus also responds similarly to related events, suggesting a shared representation of similar information, whereas posterior hippocampus tends to represent similar information distinctly, forming unique representations of overlapping information (Komorowski et al., 2013; Brunec et al., 2017). These features of hippocampal coding identified in animal research have thus far been tested in humans primarily in studies of episodic memory, but are likely broadly applicable across memory domains. Prototype correlates unique to the anterior hippocampus that were observed in the current study are entirely consistent with this model of hippocampal function. While exemplar correlates in the posterior hippocampus did not reach significance, this may have been driven by the category structure that promoted prototype representations in the majority of participants. Together with prior evidence, our results contribute to the idea that memory representations vary along the long axis of the hippocampus, with the unique role of the anterior hippocampus in supporting novel decisions based on generalized representations abstracted across experiences.

An exploratory whole-brain analysis revealed additional prototype correlates in left middle temporal cortex, bilateral temporal pole, and left superior lateral occipital cortex. While not a part of the canonical category-learning network (Seger and Miller, 2010), prototype correlates in lateral temporal regions are noteworthy given their role in semantic processing (Martin and Chao, 2001) and false memories resulting from reliance on generalized information (Garoff-Eaton et al., 2006; Dennis et al., 2014; Turney and Dennis, 2017). A recent study of false memory (Turney and Dennis, 2017) showed increasing activation in bilateral middle temporal cortices, along with VMPFC, as the similarity of lures to targets increased. A recent categorization study (Davis et al., 2017) demonstrated that activation in lateral temporal cortices, also along with the VMFPC, tracks typicality (decision evidence) during category generalization. Together, these studies suggest that these regions may be sensitive to graded typicality in multiple memory domains. Future research may elucidate the possibility that the lateral temporal cortices play a previously underappreciated role in memory generalization that may be linked to generalization processes subserved by the VMPFC.

In contrast to robust prototype correlates, the whole-brain analysis revealed no significant exemplar correlates. This prototype–model dominance in the brain matched the prototype–model dominance in behavior, with the prototype model reliably outperforming the exemplar model in 73% of participants. However, while the current data inform the decades-long "prototype versus exemplar" debate on the nature of concept representation (Homa et al., 1981; Busemeyer et al., 1984; Nosofsky et al., 2012), the strong prototype fit identified here should not be overinterpreted as evidence that categories are always represented by their prototypes. For instance, a model-based fMRI study by Mack et al. (2013) found better exemplar fit to behavior matched by exemplar correlates in the brain when using a different category structure in which prototypes were not as readily extracted or as useful for categorization. Even in our prototype-dominant study, behavioral model fits indicated that several participants (10%) relied on exemplar representations, and the group as a whole showed better classification of training items than new items of the same typicality. Thus, specific exemplars had some influence on behavior, albeit weaker than that of prototypes. In line with these behavioral indicators, several regions tracked exemplar predictors at a more lenient threshold and were consistent with the exemplar-tracking regions identified in the study by Mack et al. (2013), including lateral occipital and parietal regions. Thus, weak exemplar representations may have formed along with prototype representations. In contrast, the lack of overlap between the prototype regions identified in the current study and the exemplar regions identified in the study by Mack et al. (2013) are consistent with the idea that specific and generalized memory representations rely on partially dissociable neural systems (Preston and Eichenbaum, 2013; Collin et al., 2015; Schlichting et al., 2015). Taking these results together, we propose that factors such as the category structure (Rosch, 1975; Medin and Schaffer, 1978) and the category-training format (Aizenstein et al., 2000; Reber et al., 2003; Zeithamova et al., 2008; Zeithamova and Maddox, 2009) may bias the nature of concept representations formed during learning and accessed during generalization. Furthermore, just as representations of large-scale narratives are proposed to form alongside memories for individual events (Collin et al., 2017), prototype and exemplar representations may form in parallel across many tasks, with their relative strengths in brain and behavior varying according to the task demands.

By using latent variables from well established categorization models in an fMRI analysis, we show that the VMPFC and anterior hippocampus support concept generalization by accessing abstract prototype information. These data inform the prototype versus exemplar debate by providing novel neural evidence for

the existence of generalized concept representations. Furthermore, together with prior studies on generalization in episodic memory, the data indicate that VMPFC–hippocampal memory integration mechanisms contribute to knowledge generalization across cognitive domains.

References

Aizenstein HJ, MacDonald AW, Stenger VA, Nebes RD, Larson JK, Ursu S, Carter CS (2000) Complementary category learning systems identified using event-related functional MRI. J Cogn Neurosci 12:977–987. CrossRef Medline

Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM (1998) A neuropsychological theory of multiple systems in category learning. Psychol Rev 105:442–481. CrossRef Medline

Barron HC, Dolan RJ, Behrens TE (2013) Online evaluation of novel choices by simultaneous representation of multiple memories. Nat Neurosci 16:1492–1498. CrossRef Medline

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. Nature 456:245–249. CrossRef Medline

Bozoki A, Grossman M, Smith EE (2006) Can patients with Alzheimer's disease learn a category implicitly? Neuropsychologia 44:816–827. CrossRef Medline

Brunec IK, Bellana B, Ozubko JD, Man V, Robin J, Liu Z-X, Grady C, Rosenbaum RS, Winocur G, Barense MD, Moscovitch M (2017) Differential spatiotemporal representations along the hippocampal long axis in humans. bioRxiv. Advance online publication. Retrieved February 7, 2018. doi:10.1101/179655.

Busemeyer JR, Dewey GI, Medin DL (1984) Evaluation of exemplar-based generalization and the abstraction of categorical information. J Exp Psychol Learn Mem Cogn 10:638–648. CrossRef Medline

Collin SH, Milivojevic B, Doeller CF (2015) Memory hierarchies map onto the hippocampal long axis in humans. Nat Neurosci 18:1562–1564. CrossRef Medline

Collin SH, Milivojevic B, Doeller CF (2017) Hippocampal hierarchical networks for space, time, and memory. Curr Opin Behav Sci 17:71–76. CrossRef

Davis T, Xue G, Love BC, Preston AR, Poldrack RA (2014) Global neural pattern similarity as a common basis for categorization and recognition memory. J Neurosci 34:7472–7484. CrossRef Medline

Davis T, Goldwater M, Giron J (2017) From concrete examples to abstract relations: the rostrolateral prefrontal cortex integrates novel examples into relational categories. Cereb Cortex 27:2652–2670. CrossRef Medline

Dennis NA, Bowman CR, Peterson KM (2014) Age-related differences in the neural correlates mediating false recollection. Neurobiol Aging 35:395–407. CrossRef Medline

DeVito LM, Lykken C, Kanter BR, Eichenbaum H (2010) Prefrontal cortex: role in acquisition of overlapping associations and transitive inference. Learn Mem 17:161–167. CrossRef Medline

Eichenbaum H (2000) A cortical-hippocampal system for declarative memory. Nat Rev Neurosci 1:41–50. CrossRef Medline

Filoteo JV, Maddox WT, Davis JD (2001) Quantitative modeling of category learning in amnesic patients. J Int Neuropsychol Soc 7:1–19. CrossRef Medline

Garoff-Eaton RJ, Slotnick SD, Schacter DL (2006) Not all false memories are created equal: the neural basis of false recognition. Cereb Cortex 16:1645–1652. CrossRef Medline

Homa D, et al (1973) Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. J Exp Psychol 101:116–122. CrossRef

Homa D, Sterling S, Trepel L (1981) Limitations of exemplar-based generalization and the abstraction of categorical information. J Exp Psychol Hum Learn Mem 7:418–439. CrossRef

Kaplan R, Schuck NW, Doeller CF (2017) The role of mental maps in decision-making. Trends Neurosci 40:256–259. CrossRef Medline

Kéri S, Kelemen O, Benedek G, Janka Z (2001) Intact prototype learning in schizophrenia. Schizophr Res 52:261–264. CrossRef Medline

Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser MB (2008) Finite scale of spatial representation in the hippocampus. Science 321:140–143. CrossRef Medline

Knowlton BJ, Squire LR (1993) The learning of categories: parallel brain systems for item memory and category knowledge. Science 262:1747–1749. CrossRef Medline

Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. Science 273:1399–1402. CrossRef Medline

Koenig P, Smith EE, Troiani V, Anderson C, Moore P, Grossman M (2008) Medial temporal lobe involvement in an implicit memory task: evidence of collaborating implicit and explicit memory systems from fMRI and Alzheimer's disease. Cereb Cortex 18:2831–2843. CrossRef Medline

Komorowski RW, Garcia CG, Wilson A, Hattori S, Howard MW, Eichenbaum H (2013) Ventral hippocampal neurons are shaped by experience to represent behaviorally relevant contexts. J Neurosci 33:8079–8087. CrossRef Medline

Kruschke JK (1992) ALCOVE: an exemplar-based connectionist model of category learning. Psychol Rev 99:22–44. CrossRef Medline

Kumaran D, Summerfield JJ, Hassabis D, Maguire EA (2009) Tracking the emergence of conceptual knowledge during human decision making. Neuron 63:889–901. CrossRef Medline

Lamberts K (1995) Categorization under time pressure. J Exp Psychol Gen 124:161–180. CrossRef

Liu ZX, Grady C, Moscovitch M (2017) Effects of prior-knowledge on brain activation and connectivity during associative memory encoding. Cereb Cortex 27:1991–2009. CrossRef Medline

Mack ML, Preston AR, Love BC (2013) Decoding the brain's algorithm for categorization from its neural implementation. Curr Biol 23:2023–2027. CrossRef Medline

Maddox WT, Glass BD, Zeithamova D, Savarie ZR, Bowen C, Matthews MD, Schnyer DM (2011) The effects of sleep deprivation on dissociable prototype learning systems. Sleep 34:253–260. CrossRef Medline

Martin A, Chao LL (2001) Semantic memory and the brain: structure and processes. Curr Opin Neurobiol 11:194–201. CrossRef Medline

Medin DL, Schaffer MM (1978) Context theory of classification learning. Psychol Rev 85:207–238. CrossRef

Minda JP, Smith JD (2001) Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. J Exp Psychol Mem Cogn 27:775–799. CrossRef

Nomura EM, Reber PJ (2012) Combining computational modeling and neuroimaging to examine multiple category learning systems in the brain. Brain Sci 2:176–202. CrossRef Medline

Nosofsky RM (1986) Attention, similarity, and the identification–categorization relationship. J Exp Psychol Gen 115:39–61. CrossRef Medline

Nosofsky RM (1987) Attention and learning processes in the identification and categorization of integral stimuli. J Exp Psychol Learn Mem Cogn 13:87–108. CrossRef Medline

Nosofsky RM, Denton SE, Zaki SR, Murphy-Knudsen AF, Unverzagt FW (2012) Studies of implicit prototype extraction in patients with mild cognitive impairment and early Alzheimer's disease. J Exp Psychol Learn Mem Cogn 38:860–880. CrossRef Medline

O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci 1104:35–53. CrossRef Medline

Pajkert A, Finke C, Shing YL, Hoffmann M, Sommer W, Heekeren HR, Ploner CJ (2017) Memory integration in humans with hippocampal lesions. Hippocampus 27:1230–1238. CrossRef Medline

Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013) Long-axis specialization of the human hippocampus. Trends Cogn Sci 17:230–240. CrossRef Medline

Posner MI, Keele SW (1968) On the genesis of abstract ideas. J Exp Psychol 77:353–363. CrossRef Medline

Preston AR, Eichenbaum H (2013) Interplay of hippocampus and prefrontal cortex in memory. Curr Biol 23:R764–R773. CrossRef Medline

Preston AR, Shrager Y, Dudukovic NM, Gabrieli JD (2004) Hippocampal contribution to the novel use of relational information in declarative memory. Hippocampus 14:148–152. CrossRef Medline

Reber PJ, Gitelman DR, Parrish TB, Mesulam MM (2003) Dissociating explicit and implicit category knowledge with fMRI. J Cogn Neurosci 15:574–583. CrossRef Medline

Reed SK (1972) Pattern recognition and categorization. Cogn Psychol 3:382–407. CrossRef

Richter FR, Chanales AJH, Kuhl BA (2016) Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. Neuroimage 124:323–335. CrossRef Medline

Rosch E (1975) Cognitive representations of semantic categories. J Exp Psychol Gen 104:192–233. CrossRef

Rosch E, Simpson C, Miller RS (1976) Structural bases of typicality effects. J Exp Psychol Hum Percept Perform 2:491–502. CrossRef

Schlichting ML, Preston AR (2015) Memory integration: neural mechanisms and implications for behavior. Curr Opin Behav Sci 1:1–8. CrossRef Medline

Schlichting ML, Preston AR (2017) The hippocampus and memory integration: building knowledge to navigate future decisions. In: The hippocampus from cells to system: structure, connectivity, and functional contributions to memory and flexible cognition (Duff MC, Hannula DE, eds), pp 405–437. New York, NY: Springer.

Schlichting ML, Mumford JA, Preston AR (2015) Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat Commun 6:8151. CrossRef Medline

Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. J Neuropsychiatry Clin Neurosci 20:11–21. CrossRef Medline

Seger CA, Miller EK (2010) Category learning in the brain. Annu Rev Neurosci 33:203–219. CrossRef Medline

Shepard RN (1957) Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. Psychometrika 22:325–345. CrossRef

Shohamy D, Wagner AD (2008) Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron 60:378–389. CrossRef Medline

Shohamy D, Myers CE, Onlaor S, Gluck MA (2004) Role of the basal ganglia in category learning: how do patients with Parkinson's disease learn? Behav Neurosci 118:676–686. CrossRef Medline

Smith JD, Minda JP (1998) Prototypes in the mist: the early epochs of category learning. J Exp Psychol Learn Mem Cogn 24:1411–1436. CrossRef

Smith JD, Redford JS, Haas SM (2008) Prototype abstraction by monkeys (Macaca mulatta). J Exp Psychol Gen 137:390–401. CrossRef Medline

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 [Suppl 1]:S208–S219. CrossRef Medline

Squire LR, Knowlton BJ (1995) Learning about categories in the absence of memory. Proc Natl Acad Sci U S A 92:12470–12474. CrossRef Medline

Stensola H, Stensola T, Solstad T, Frøland K, Moser MB, Moser EI (2012) The entorhinal grid map is discretized. Nature 492:72–78. CrossRef Medline

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RG (2007) Schemas and memory consolidation. Science 316:76–82. CrossRef Medline

Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Bito H, Morris RG (2011) Schema-dependent gene activation and memory encoding in neocortex. Science 333:891–895. CrossRef Medline

Tulving E, Markowitsch HJ (1998) Episodic and declarative memory: role of the hippocampus. Hippocampus 8:198–204. CrossRef Medline

Turney IC, Dennis NA (2017) Elucidating the neural correlates of related false memories using a systematic measure of perceptual relatedness. Neuroimage 146:940–950. CrossRef Medline

van Kesteren MT, Rijpkema M, Ruiter DJ, Fernandez G, Fernández G (2010) Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. J Neurosci 30:15888–15894. CrossRef Medline

Zaki SR (2004) Is categorization performance really intact in amnesia? A meta-analysis. Psychon Bull Rev 11:1048–1054. CrossRef Medline

Zaki SR, Nosofsky RM, Jessup NM, Unverzagt FW (2003a) Categorization and recognition performance of a memory-impaired group: evidence for single-system models. J Int Neuropsychol Soc 9:394–406. CrossRef Medline

Zaki SR, Nosofsky RM, Stanton RD, Cohen AL (2003b) Prototype and exemplar accounts of category learning and attentional allocation: a reassessment. J Exp Psychol Learn Mem Cogn 29:1160–1173. CrossRef Medline

Zeithamova D, Maddox WT (2009) Learning mode and exemplar sequencing in unsupervised category learning. J Exp Psychol Learn Mem Cogn 35:731–741. CrossRef Medline

Zeithamova D, Maddox WT, Schnyer DM (2008) Dissociable prototype learning systems: evidence from brain imaging and behavior. J Neurosci 28:13194–13201. CrossRef Medline

Zeithamova D, Dominick AL, Preston AR (2012a) Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron 75:168–179. CrossRef Medline

Zeithamova D, Schlichting ML, Preston AR (2012b) The hippocampus and inferential reasoning: building memories to navigate future decisions. Front Hum Neurosci 6:70. CrossRef Medline