Behavioral/Cognitive

Global Neural Pattern Similarity as a Common Basis for Categorization and Recognition Memory

Tyler Davis,[1] Gui Xue,[6] Bradley C. Love,[7] Alison R. Preston,[3,4,5] and Russell A. Poldrack[2,3,4,5]

[1]Department of Psychology, Texas Tech University, Lubbock, Texas 79409, [2]Imaging Research Center, [3]Department of Psychology, [4]Center for Learning and Memory, and [5]Department of Neuroscience, The University of Texas at Austin, Austin, Texas 78712, [6]National Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, People's Republic of China, and [7]University College London, London WC1E 6BT, United Kingdom

Familiarity, or memory strength, is a central construct in models of cognition. In previous categorization and long-term memory research, correlations have been found between psychological measures of memory strength and activation in the medial temporal lobes (MTLs), which suggests a common neural locus for memory strength. However, activation alone is insufficient for determining whether the same mechanisms underlie neural function across domains. Guided by mathematical models of categorization and long-term memory, we develop a theory and a method to test whether memory strength arises from the global similarity among neural representations. In human subjects, we find significant correlations between global similarity among activation patterns in the MTLs and both subsequent memory confidence in a recognition memory task and model-based measures of memory strength in a category learning task. Our work bridges formal cognitive theories and neuroscientific models by illustrating that the same global similarity computations underlie processing in multiple cognitive domains. Moreover, by establishing a link between neural similarity and psychological memory strength, our findings suggest that there may be an isomorphism between psychological and neural representational spaces that can be exploited to test cognitive theories at both the neural and behavioral levels.

Introduction

Familiarity is a ubiquitous part of everyday cognition. For example, when encountering a familiar face on a subway, one may pause to determine whether or not the person is a friend. Operationally, familiarity is thought to reflect the strength of memory traces associated with particular items or categories and is a key component of formal cognitive models in many research domains (Gillund and Shiffrin, 1984; Hintzman, 1988; Nosofsky, 1988, 1991; Norman and O'Reilly, 2003).

The convergence of research on categorization and recognition memory suggests that there may be a shared neural substrate for the processes underlying familiarity. For example, although there is strong divergence in terms of the neural mechanisms that support recognition memory and some types of category learning (for review, see Ashby and Maddox, 2005; Seger and Miller, 2010), measures of memory strength in both domains have been found to track mean activation in the medial temporal lobe (MTL) (Ranganath et al., 2004; Daselaar et al., 2006; Seger et al., 2011; Davis et al., 2012a,b). The underlying mechanisms by which neural activity gives rise to memory strength, however, have not been elucidated, making it unclear whether this common activation reflects the same computational processes in both domains.

Computational theories of long-term memory and categorization can aid in relating activation within different brain regions with cognitive mechanisms (Daw, 2011; Forstmann et al., 2011). In formal models of long-term memory and categorization, memory strength reflects how similar an item is to all other representations stored in memory, which is often referred to as "global similarity" (Gillund and Shiffrin, 1984; Hintzman, 1988; Nosofsky, 1988, 1991). In long-term memory experiments, global similarity between representations of items stored in memory is thought to contribute to a number of behavioral measures, including recognition memory and recall (Raaijmakers and Shiffrin, 1992; Clark and Gronlund, 1996). In categorization, global similarity, along with the similarity an item to its own category, is one of the pieces of information that is used to decide how to classify an item (Nosofsky, 1988; Love et al., 2004). Although classification is the primary role of global similarity in categorization models, category members that have high global similarity are also often associated with stronger recognition memory (Sakamoto and Love, 2004, 2006), even when they have not been previously encountered (Nosofsky, 1988).

Previous studies suggest that overall activation of the MTL and mathematical global similarity measures are both associated with behavioral measures of familiarity, which suggests that MTL may be engaged in a global similarity computation that underlies

Received Aug. 7, 2013; revised March 25, 2014; accepted April 21, 2014.

Author contributions: T.D., G.X., B.C.L., A.R.P., and R.A.P. designed research; T.D. and G.X. performed research; T.D. analyzed data; T.D., G.X., B.C.L., A.R.P., and R.A.P. wrote the paper.

This project is partially supported by the National Natural Science Foundation of China (31130025) and the 973 Program (2014CB846102) to G.X., the James S. McDonnell Foundation to R.P., and the National Institute of Mental Health (MH091523) to B.C.L. and A.R.P. We thank Frances Fawcett for use of beetle stimuli, and Ken Norman for comments on a previous version of this manuscript.

Correspondence should be addressed to Tyler Davis, Department of Psychology, MS 2051 Psychology Building, Texas Tech University, Lubbock, TX 79409. E-mail: tyler.h.davis@ttu.edu.

DOI:10.1523/JNEUROSCI.3376-13.2014

 0270-6474/14/347472-13$15.00/0

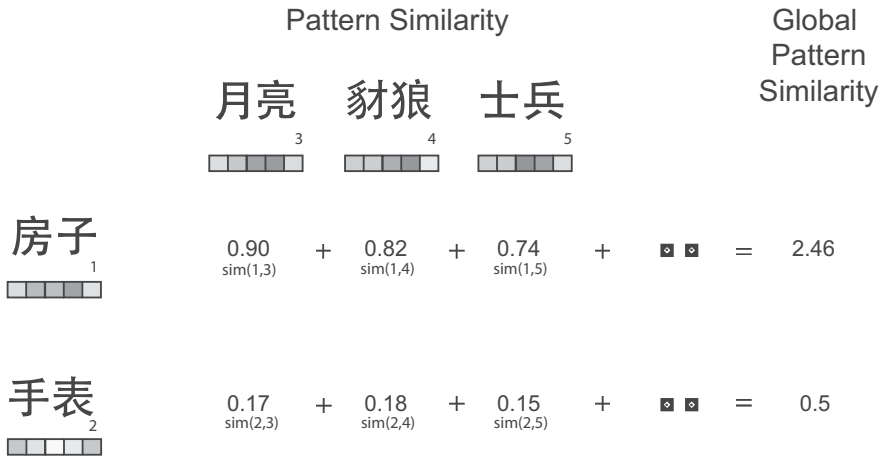

Figure 1. A depiction of the neural global pattern similarity measure. Patterns of activation elicited for a given target word (depicted by 1 and 2) are compared with patterns elicited for all other words (depicted by 3, 4, and 5). Words whose patterns of activation are highly similar to those of other items have higher aggregate matches, which is the basis for global pattern similarity. The neural global pattern similarity measure relates to global similarity processes in categorization and long-term memory models. Models predict that the higher the global similarity of an item with respect to representations of other items, the more strongly it will be remembered.

familiarity. In the present study, we develop a mathematical measure of global neural similarity (Fig. 1) that is inspired by formal memory and categorization models. In both categorization and long-term memory tasks, we find that items associated with stronger psychological memory strength elicit patterns of activation that are most similar to those of other studied items, consistent with the hypothesis that the MTL engages a global similarity process.

Materials and Methods

Here we detail our computational framework and current analysis, as well as give a brief overview of fMRI acquisition and behavioral methods. Detailed methods describing participants, experimental design, computational modeling, and behavioral and imaging data acquisition and processing can be found in Davis et al. (2012a) for the categorization task and Xue et al. (2010) for the long-term memory task.

Neural global similarity measure

Our multivoxel neural global pattern similarity measure (Fig. 1) is inspired by computational models of global similarity that measure how similar an item is to all of the items in a task from psychological or physical representations of stimuli (Nosofsky, 1988, 1991). Instead of measuring similarities between psychological or physical representation of stimuli, neural global pattern similarity measures similarities between multivoxel activation patterns elicited for stimuli in the task. Stimuli that are associated with high neural global pattern similarity (GPS) are central in a neural activation space.

Formally, the global pattern similarity of an object i arises from the similarity between the multivoxel neural activation pattern elicited for item i (A_i) and the activation patterns for each j item in the set (K) of all items encountered in the task, as follows:

$$\mathrm{GPS}(i) = \sum_{j=1}^{K} \mathrm{sim}(A_i, A_j), \quad (1)$$

where the similarity between the patterns of items i and j is a function of the distance between their elicited activation patterns, as follows:

$$\mathrm{sim}(A_i, A_j) = \exp^{-d(A_i, A_j)}. \quad (2)$$

The exponential function in Equation 2 institutes a similarity gradient that scales the distances between stimuli such that, as the distance between the patterns for items i and j increases, the similarity between the items decreases exponentially. Although there are a variety of functions that can be used as a similarity gradient (Lesot et al., 2009), the exponential has been among the most successful in explaining how distances in psychological space relate to generalization between two stimuli (Shepard, 1987), and thus serves as a starting point for testing our neural global pattern similarity measure.

The distance between i and j is computed from the Pearson correlation distance (Kriegeskorte et al., 2008) between the multivoxel patterns elicited for stimuli i and j, as follows:

$$d(A_i, A_j) = (1 - \mathrm{corr}(A_i, A_j)). \quad (3)$$

We use a correlation distance metric in the present context for consistency with the majority of previous similarity-based neuroimaging analysis where correlations are used due to their automatic removal of the mean activation/engagement across voxels (Kriegeskorte et al., 2008). However, as with the similarity gradient described above, a number of choices are possible for the precise distance metric (Lesot et al., 2009), and there has yet to be a systematic investigation into which metrics are best for similarity-based neuroimaging analysis (Davis and Poldrack, 2013a).

After neural global pattern similarity is computed for each stimulus, the correlation can be estimated between the resulting item-wise global similarities and item-wise behavioral and computational measures of memory strength. In the long-term memory task (Xue et al., 2010), we estimate the correlation between the neural global pattern similarity measure and subjects' subsequent recognition memory confidence. Recognition memory confidence is hypothesized to reflect the outcome of global similarity processes operating on the psychological representations (P_i) of stimuli in a task (Raaijmakers and Shiffrin, 1992; Clark and Gronlund, 1996). Significant correlations between recognition memory confidence and our neural global pattern similarity measure are thus evidence for informational overlap between P_i and A_i in terms of which stimuli are the most central within the two spaces. Given that we are measuring global similarity during encoding in the long-term memory task, the key assumption is that the activation patterns present at encoding contain information about how the representations of stimuli will overlap and influence later retrieval. The more stimuli elicit similar neural activation patterns, the more they are predicted to overlap in their representations, and hence, according to global similarity models, the more likely they are to be subsequently remembered.

In the categorization task (Davis et al., 2012a), we examined the relationship between the neural global pattern similarity measure and a psychological measure of global similarity, termed "recognition strength," computed from a category learning model SUSTAIN (see Fig. 3*B*; for details, see Davis et al., 2012a). For the present purposes, the key difference between the recognition strength measure and the neural global pattern similarity measure of SUSTAIN is that, whereas neural global pattern similarity is computed from similarity between multivoxel activation patterns A_i, the recognition strength measure of SUSTAIN is computed from psychological representations P_i, which it learns from the stimulus space in the task. Thus, significant correlations between neural global pattern similarity and the recognition strength measure of SUSTAIN are evidence that there is information overlap between P_i and A_i in terms of which stimuli are most central within the two spaces.

fMRI acquisition

Long-term memory task. Imaging data were acquired on a 3.0 T Siemens MRI scanner in the MRI Center at Beijing Normal University. Structural images were acquired using a T1-weighted, three-dimensional, gradient echo pulse sequence (TR = 2530 ms; TE = 3.39 ms; $\theta = 7°$; FOV = 256 × 256 mm; matrix = 192 × 256; slice thickness = 1.33 mm). One hundred twenty-eight sagittal slices were acquired to provide high-resolution

structural images of the whole brain. Functional images were acquired using a single-shot T2*-weighted gradient echo EPI sequence (TR = 2000 ms; TE = 30 ms; $\theta = 90°$; FOV = 200 × 200 mm; matrix = 64 × 64; slice thickness = 4 mm). Thirty contiguous axial slices parallel to the anterior commissure–posterior commissure (AC–PC) line were obtained to cover the whole cerebrum and partial cerebellum.

Categorization task. Imaging data were acquired on a 3.0 T GE Healthcare Signa MRI scanner in the MRI Center at University of Texas at Austin. Structural images were acquired using a T2-weighted, flow-compensated spin-echo pulse sequence (TR = 3 s; TE = 68 ms; 256 × 256 matrix; 1 × 1 mm in-plane resolution) using 31 3-mm-thick oblique axial slices (0.6 mm gap), ~20° off the AC–PC line, oriented for the best whole-brain coverage. Functional images were acquired using a multiecho GRAPPA parallel imaging EPI sequence using the same slice prescription as the structural images (TR = 2 s; TE = 30 ms; 2 shot; flip angle = 90°; 64 × 64 matrix; 3.75 × 3.75 mm in-plane resolution).

Behavioral methods

Long-term memory task. Male ($n = 11$) and female ($n = 11$) human subjects were scanned during a semantic memory task in which 60 common Chinese words were presented three times over the course of three scanning runs. On each trial, subjects were presented with a 1 s fixation followed by the presentation of a word. Subjects were asked to decide whether the word corresponded to a living or nonliving object. Subjects had 3 s to key in their response. After responding, subjects completed 8 s of a self-paced visual orientation judgment task in which they chose whether a 45° Gabor patch was tilting to the left or right. Thirty minutes after the scan, subjects were asked to perform two surprise memory tasks. First, subjects were asked to freely recall words they remembered from the scanning session by writing them down in any order. Next, subjects completed a recognition memory task containing the original 60 words plus an additional 60 foils. Subjects were asked to give their recognition confidence for each word on a scale of 1–6, ranging from definitely new to definitely old.

Similarity rating task. To supplement the primary results, we conducted an additional similarity rating task. The goal of this task was to provide a measure of the semantic relationships between words that would allow us to test hypotheses about the representational content of activation patterns that are used as inputs into the global neural similarity measure. Semantic similarity between an item and representations stored in memory is one of the features thought to feed into memory strength contributions, according to global similarity models (Steyvers et al., 2004).

For the similarity rating task, male ($n = 4$) and female ($n = 4$) native Chinese speakers provided pairwise similarity ratings for each of the words used in the long-term memory task. None of the subjects in the similarity rating task were scanned in the long-term memory task. On each trial, subjects were asked to rate the similarity between each pair of words on a 1–7 scale. Subjects were instructed to base their similarity ratings on word meaning (i.e., semantic similarity). Because of the high number of similarity ratings (1770 unique pairs), the ratings were randomly spread over nine sessions, which subjects completed over the course of 2 h.

Categorization task. Male ($n = 9$) and female ($n = 13$) human subjects completed a rule-plus-exception category learning task during fMRI scanning. In this task, subjects learn by trial and error to classify schematic beetles into categories (Hole A or Hole B beetles) based on their features (Fig. 2*A*,*B*; Table 1). Most beetles in the task are rule-following items and can be accurately classified by using a simple rule based on a single stimulus dimension (e.g., if the beetle has thick legs, it belongs in Hole A). However, there are also exceptions to this rule that must be represented distinctly from the rule-following items (e.g., a specific Hole A beetle with thin legs). Before learning, subjects were informed which stimulus dimension would be the rule dimension (Table 1, first digit of the abstract representations) and given a hint to check this dimension on each trial.

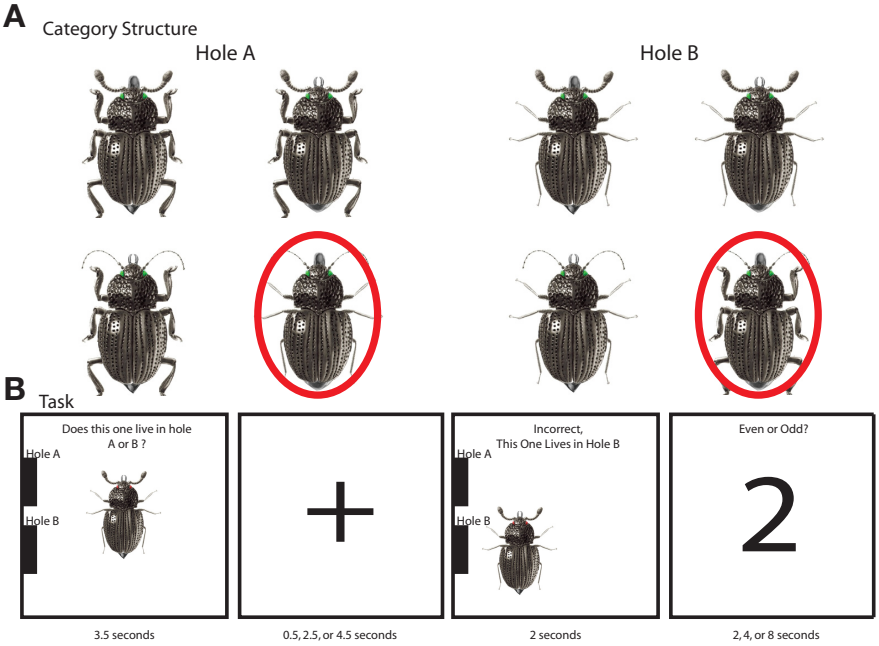

Figure 2. ***A***, An example category structure. There are a total of eight beetles in the task. Most of the beetles can be easily classified based on the following rule: for example, if it has thick legs, it belongs in Hole A. However, each category also contains an exception item (circled) that looks as if it should belong in the opposite category. ***B***, An example trial sequence. On each trial, subjects are presented with a beetle and are asked to supply the correct category label. They then receive feedback about the correct category. An even/odd digit task separates the trials.

Table 1. Abstract category structure

Category	Structure
Hole A beetles	2 2 2 2[a]
	1 1 1 2
	1 1 2 1
	1 2 1 1
Hole B beetles	1 2 2 2[a]
	2 1 1 2
	2 1 2 1
	2 2 1 1
Recognition test foils	1 1 1 1
	1 1 2 2
	1 2 1 2
	1 2 2 1
	2 2 2 2
	2 2 1 1
	2 1 2 1
	2 1 1 2

Each row represents a unique stimulus (i.e., beetle). The four values assigned to a stimulus denote the four stimulus dimensions (e.g., legs, antennae) assigned to a beetle. Each numeric value (1 or 2) represents a specific feature instantiation (e.g., red or green eyes). The first dimension represents the rule-relevant dimension. Most Hole A beetles have a 1 on the first dimension (e.g., thick legs), whereas most Hole B beetles have a 2 (e.g., thin legs). The first stimulus in each of the columns is therefore an exception. The recognition test foils were never presented in the categorization task and were saved for the surprise 2AFC recognition memory task. In each block of the 2AFC recognition task, each Hole A and Hole B beetle was paired with each of the recognition test foils with the same value on the rule dimension one time. For each subject, which physical dimension (eyes, tail, antennae, fangs, and legs) corresponded to the abstract feature dimensions (1s and 2s) was randomized, and the fifth physical dimension was held fixed.

[a]Exception item.

Subjects were trained on the rule-plus-exception task for 30 blocks in which each of the eight beetle stimuli were presented once in a random order. The 30 blocks were divided equally among six scanning runs, each of which lasted 8 min and 27 s. On each trial of the task (Fig. 2*B*), subjects were presented with a stimulus for 3.5 s during which time they were asked which one of the two categories (Hole A or Hole B) the stimulus belonged to. The stimulus presentation time was followed by variable fixation (0.5, 2.5, or 4.5 s) after which subjects were presented with feedback detailing the correct category assignment. Feedback was followed by an active even/odd digit task baseline of 2, 4, or 8 s (Stark and Squire, 2001). Following the category learning task, subjects completed a self-paced, two-alternative forced choice recognition memory task outside of the scanner. On each trial of the recognition task, subjects were presented with two beetles: one that was presented during the category learning phase and a foil (Table 1) that was not presented during the category learning phase. Subjects were asked to identify the old item presented during the scanned rule-plus-exception task.

fMRI image preprocessing and analysis

FSL was used for image processing. Functional time series were skull stripped, motion corrected, prewhitened, and high-pass filtered (cutoff: category learning task = 100 Hz; long-term memory task = 60 Hz). No spatial smoothing was performed on the functional data. First-level statistical maps were registered to the Montreal Neurological Institute (MNI)-152 template using 7 df to align the functional image to the structural image, and 12 df to align the structural image to the MNI-152 template.

Long-term memory task. Multivoxel activation patterns for each stimulus were computed in two ways. For the primary analysis, the voxelwise GLM was used to compute a *t*-map giving the difference between baseline activation and the activation elicited for each of the 60 unique stimuli in the task. Stimulus presentation times for each unique stimulus were modeled with a double-gamma hemodynamic response and its temporal derivatives. Unconvolved motion parameters were included as nuisance regressors. In a second model, which was designed to allow the computation of self-similarity across repetitions a stimulus, a β-map was computed for each repetition of each stimulus by using an LS-S procedure (Mumford et al., 2012). The multivoxel response patterns for each stimulus were registered to standard space to facilitate cross-run and cross-subject comparisons.

Multivoxel activation patterns (*t*-maps) for each of the unique stimuli were used as the inputs into the neural global similarity measure. For each stimulus, the similarity between its multivoxel pattern and all of the other 60 items was computed and summed according to the neural global pattern similarity measure. A correlation was then computed between this 60-item global neural similarity vector and the behavioral memory strength measures (recognition confidence and recall).

The spatial localization of the voxels within the *t*-map used to compute the neural global pattern similarity measure was selected using a searchlight algorithm (Kriegeskorte et al., 2006) with a 3 voxel radius. For each searchlight, a correlation between the observed global neural similarity and the behavioral memory measures was computed and stored at the voxel corresponding to the center of the searchlight. The resulting subject-level correlation maps were transformed using Fisher's *z* test and combined for second-level between-subjects analysis using a one-sample *t* test.

The results presented in the body of the manuscript use a small-volume anatomical cluster correction based on an MTL mask (Harvard-Oxford), a cluster-forming threshold of $p < 0.05$ and a corrected extent threshold of $p < 0.05$ using a permutation test. Although the results presented in the article are significant when using the MTL-based anatomical correction, there were no significant clusters observed at conventional whole-brain corrected thresholds for the long-term memory task.

In addition to the primary analyses, we conducted a series of analyses examining the relationship of the global pattern similarity to self-similarity, recall, and recognition confidence within only the nonrecalled items. In the recall analysis, we computed the correlation between a dummy coded recall variable (1 = recalled; 0 = nonrecalled) and global pattern similarity for each searchlight. The resulting statistical maps give the regions in which global pattern similarity was significantly higher for recalled items. In a second, nonrecalled analysis, we restricted all of the analyses to nonrecalled items such that only nonrecalled items were included in the global pattern similarity computation, and the correlation between global pattern similarity and recognition confidence included only nonrecalled items. Finally, in the partial similarity analysis, we partialled out the effect of mean self-similarity across repetitions on recognition confidence before estimating the correlation between recognition confidence and global pattern similarity, and vice versa. The resulting statistical maps give the regions in which there is significant variance in recognition confidence accounted for by mean self-similarity or global pattern similarity after controlling for the other.

Finally, pairwise similarity analysis was conducted to test whether semantic relationships between words were coded in the item-wise activation patterns. The semantic similarity between an item and representations stored in memory is one of the features thought to feed into memory strength contributions, according to global similarity models (Steyvers et al., 2004). In the pairwise similarity analysis, we examined how the information contained in the rated pairwise similarities between the words collected in the similarity rating task related to the pairwise similarities between neural activation patterns elicited in the long-term memory task that feed into the global pattern similarity measure. Instead of using the average raw pairwise similarity ratings collected in the similarity rating task, however, we computed a second-order similarity measure in which each pairwise similarity rating was computed as the correlation between the raw rating vectors of the two words. The logic behind this measure is that words that are similar and dissimilar to the same words will be more similar to each other than words that are similar and dissimilar to other words. Conceptually related second-order similarity measures are often used in computational linguistics (Landauer and Dumais, 1997; Islam and Inkpen, 2006; second-order pointwise mutual information; e.g., latent semantic analysis) and gene network analysis (Ravasz et al., 2002). This second-order similarity measure reduced within-subject noise in the ratings, thus resulting in enhanced reliability of the mean pairwise similarities across subjects (intraclass correlation for raw ratings = 0.65; intraclass correlation for second-order similarities = 0.77). This analysis used the same searchlight procedure used in the other analyses except that the correlation stored for each searchlight was the correlation between the lower triangle of the pairwise word similarity matrix and the lower triangle of the pairwise neural similarity matrix (Kriegeskorte et al., 2008; see Fig. 6*A*).

Categorization task. Voxelwise GLM analysis was conducted in FEAT to obtain a *t*-map for each of the eight stimuli in each of the six scanning runs (Fig. 2*A*; Table 1). Stimulus presentation and feedback times were modeled using a double-gamma hemodynamic response and its temporal derivatives. To account for between-item differences in time on task, reaction time was controlled for by including a regressor in which the durations of the hemodynamic response varied according to reaction time. This regressor was orthogonalized with respect to stimulus presentation regressors. Unconvolved motion parameters were included as nuisance variables.

A *t*-map giving voxelwise activation differences during stimulus presentation from baseline was computed for each stimulus-by-scanning run combination, yielding 48 unique multivariate stimulus representations. These *t*-maps were registered to standard space to ensure alignment of anatomical regions between runs and used as neural stimulus representations to compute the neural global pattern similarity measure (see Eqs. 1–3 above).

To evaluate how neural global pattern similarity changed over time for each stimulus, the neural global pattern similarity measure was calculated for each stimulus by comparing it only to the other seven stimuli within the same run. A correlation was then computed between the neural global pattern similarity measure and a psychological global similarity measure, recognition strength, generated from a computational model, SUSTAIN (Love et al., 2004; but see Davis et al., 2012a). Model derivation and parameter settings used to generate the recognition strength predictions of SUSTAIN are the same as those detailed in Davis et al. (2012a). Subject-level correlation maps obtained from the searchlight algorithm were transformed using Fisher's *z* test, and combined for

second-level between-subjects analysis. Subject-level maps were submitted to a group t test. The primary results use a small-volume anatomical cluster correction based on an MTL mask (Harvard-Oxford), a cluster-forming threshold of $p < 0.05$, and a corrected extent threshold of $p < 0.05$ using a permutation test. Whole-brain results use a $p < 0.001$ cluster-forming threshold and a $p < 0.05$ corrected extent threshold using a permutation test.

In addition to the primary results, we included pairwise similarity analyses to test which aspects of the psychological category structure were coded in the activation patterns used as inputs into our neural global pattern similarity measure. For this analysis, we used the anatomically based MTL mask and calculated the pairwise similarities (i.e., correlations) between the activation patterns elicited for each of the eight stimuli in the final category learning run. Subjects' final-run correlation matrices were used to compute group-level t tests to test whether items were more similar to members of their own or other categories. Additionally, the average (across-subjects) pairwise similarity matrix was submitted to a multidimensional scaling analysis using a Sammon mapping procedure to visualize the information coded in the activation patterns.

Results

Application 1: long-term memory

Global similarity plays a prominent role in models of long-term memory and is theorized to contribute to judgments of familiarity and recognition confidence (Raaijmakers and Shiffrin, 1992; Clark and Gronlund, 1996). According to formal long-term memory models, recognition confidence and familiarity during recognition memory tests derive from the extent to which stimuli match the representations stored in memory during encoding. Here we test whether the global pattern similarity between activation patterns elicited for words during incidental encoding is associated with higher subsequent memory strength in a word list-based long-term memory task (Xue et al., 2010, Experiment 3). We hypothesize that if representational overlap is driving memory strength, activation patterns in the MTL should be more globally similar for subsequently remembered items.

The previous results from the study by Xue et al. (2010) focused primarily on how encoding variability between repeated presentations of stimuli related to long-term memory. The authors found that pattern similarity across repetitions of words during encoding (i.e., self-similarity) was significantly higher for subsequently recalled words compared with forgotten words. In the present analyses, we test whether words that are subsequently remembered also exhibit higher global neural pattern similarity to the activation patterns of other items elicited during encoding. This prediction follows from formal long-term memory models, which predict that memory strength is driven by the similarity relationships between the representations of all stimuli encoded in memory and not just a stimulus self-similarity. Additionally, we examine whether this global similarity accounts for additional variance in subsequent memory strength not accounted for by the previous self-similarity findings. These analyses focus on the relationship between neural similarity measures and subjects' recognition confidence ratings, which provide a more graded measure of memory strength than the recall measure used in previous analyses by Xue et al. (2010).

Recognition confidence analysis

Our first hypothesis was that there would be a significant correlation between global neural pattern similarity in the MTL and subjects' recognition confidence. To test this hypothesis, we computed the correlation between the item-wise neural global pattern similarity measure and subjects' recognition confidence using a searchlight procedure. In this procedure, voxels were iteratively selected from groupings of adjacent voxels across the brain and used to compute the neural global pattern similarity measure. The correlation between the item-wise neural global pattern similarity measure and subjects' recognition confidence was stored at the central voxel in each searchlight.

We found that there was a significant correlation between subjects' recognition confidence and global pattern similarity in

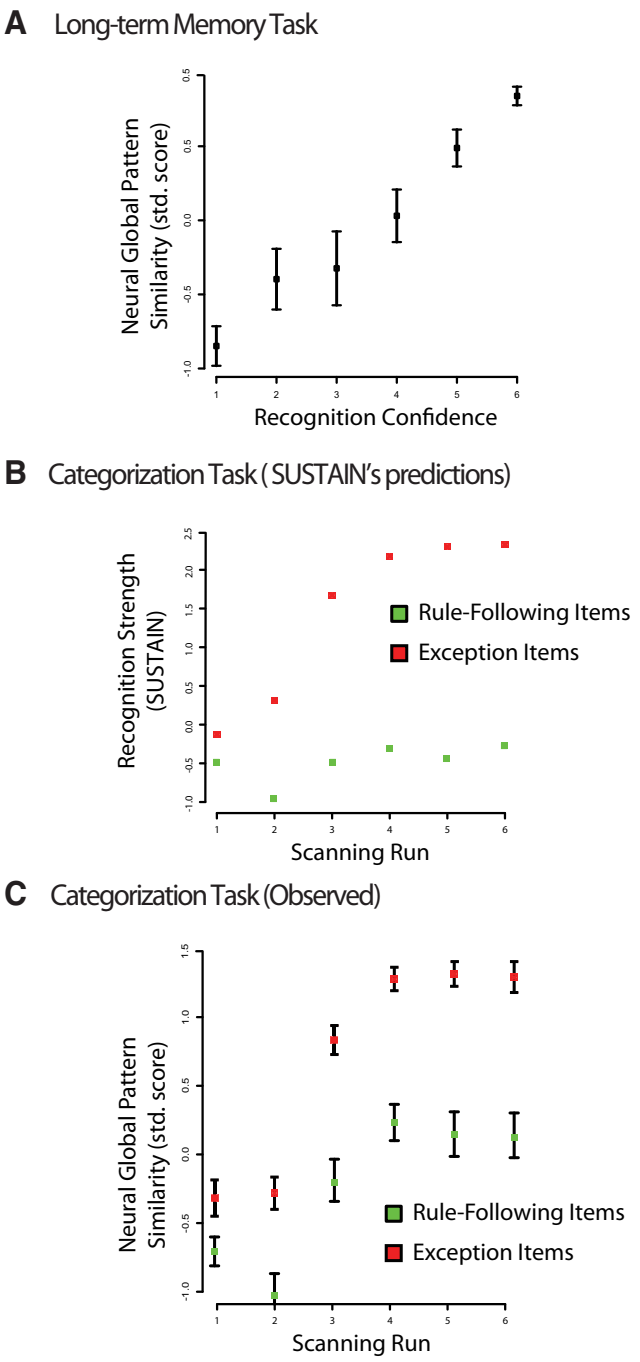

Figure 3. ***A***, Global pattern similarity values extracted from 6 mm spheres around subjects' peak correlation values in the MTL in the recognition memory experiment. ***B***, Predictions from SUSTAIN for how recognition strength changes for the item types over the course of learning. Exceptions are depicted in red, and rule-following items are depicted in green. ***C***, Global pattern similarity values extracted from 6 mm spheres around subjects' peak correlation values in the MTL in the categorization experiment. All global pattern similarity values are standardized within subjects to neutralize individual differences in mean similarity and scale. Although not independent from the statistical maps presented in Figure 4, these plots help to assess the effect sizes and viability of the linearity assumptions (Friston, 2012).

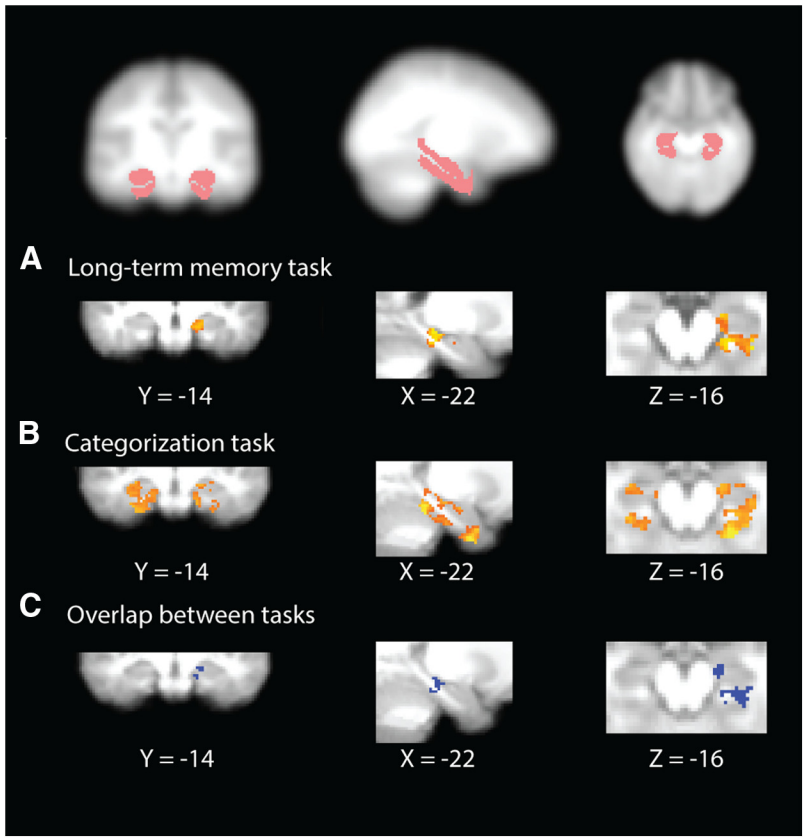

Figure 4. Statistical results from our tests of the global pattern similarity measure for the long-term memory and categorization tasks. The top row indicates the anatomical basis of the MTL mask used in our primary analyses. ***A–C***, Statistical maps depict regions in the MTLs in which the neural global pattern similarity measure tracks. ***A***, Subsequent memory confidence in the long-term memory task. ***B***, The predicted recognition strength of SUSTAIN in the category learning task. The statistical maps depict group-level *t* statistics masked with clusters obtained from an anatomically based small-volume cluster correction. ***C***, Depicts the overlap in significant voxels between ***B*** and ***C***.

Table 2. MTL small-volume corrected clusters in which a significant correlation was observed between the neural global pattern similarity measure and the recognition strength measure of SUSTAIN in the categorization task, and recognition confidence in the long-term memory task

Task	Region	Cluster	Cluster size (voxels)	Cluster *p* value	MNI space *x*	*y*	*z*	Peak (*z* score)
Categorization task	Anterior parahippocampal gyrus	1	952	0.002	−28	−2	−36	5.28
	Hippocampus	1			−26	−12	−20	2.47
	Anterior parahippocampal gyrus	2	794	0.003	24	−22	−30	4.22
	Posterior parahippocampal gyrus	2			14	−36	−4	3.09
	Hippocampus	2			24	−22	−10	2.46
Long-term memory task	Hippocampus	1	322	0.015	−34	−30	−14	5.48
	Hippocampus	1			−26	−40	−6	2.7

Peaks represent local maxima separated by at least 12 mm. Coordinates are in MNI space.

the MTL cortex and hippocampus (Figs. 3*A*, 4*A*; Table 2). When searchlights were constrained to subregions within the MTL, correlations between recognition confidence and global similarity were marginal in left hippocampus ($p = 0.06$) and left parahippocampal cortex ($p = 0.078$).

These results support the hypothesis that a global similarity process underlies MTL computations in long-term memory tasks. As theorized by computational models of long-term memory, items that elicited patterns of activation at encoding that were the most globally similar were patterns that were the most confidently recognized. Importantly, however, these confidently remembered items were more globally similar in a neural activation space—not just in terms of the psychological representational spaces posited by long-term memory models. These results suggest that there is informational overlap between the psychological representations of items and the neural activation patterns elicited for these items in the MTL, at least in terms of which items are the most central or similar to other items.

Relating global and self-similarity

The present results are independent of those presented by Xue et al. (2010) because, in the present study, global pattern similarity is computed solely from between-item similarities that are based on a single activation pattern averaged over the three repetitions. The original analysis by Xue et al. (2010) focused only on similarity between repetitions of the same stimulus. Although these results were not anticipated by any of the original analyses by Xue et al. (2010), it is worthwhile to consider the relationships between the self-similarity findings and the present global pattern similarity findings to assess how global and self-similarity relate. In a second series of analyses, we examined whether, statistically, the global pattern similarity measure is tapping into information not revealed in the original analyses by Xue et al. (2010), which examined the relationship between self-similarity and recall.

One question that we examined is how global similarity relates to recall, the primary memory measure used in the original analysis by Xue et al. (2010). In terms of recall measures, recalled words are likely to be those that are very strongly remembered and may be associated with distinct pattern-separated representations that do not overlap with representations of other items (O'Reilly and Norman, 2002). On the other hand, global similarity is also thought to contribute to recall performance (Gillund and Shiffrin, 1984). Consistent with the latter explanation, we found that neural global pattern similarity was significantly higher for recalled items (Fig. 5*A*), suggesting that overlap in MTL activation patterns between items may also lead to successful recall.

This finding of higher neural global pattern similarity for recalled items raises an important question with respect to the global similarity analyses presented in the previous section (Recognition confidence analysis): because recalled items tend to be associated with high recognition memory confidence, is the correlation between recognition confidence and neural global pattern similarity driven by the difference between recalled and nonrecalled items? To test this question, we examined whether

the correlation between recognition confidence and neural global pattern similarity was significant within only the nonrecalled items. We found significant correlations between global neural pattern similarity and recognition confidence in the MTL (Fig. 5*B*). Collectively, these results suggest that neural global pattern similarity may underlie memory strength computations for both recalled items and memories with intermediate strength (recognized but not recalled items).

In addition to focusing on recall, Xue et al. (2010) also restricted their neural similarity analyses to self-similarity between repetitions of a stimulus. This raises the question of whether neural global pattern similarity is able to explain unique variability in memory strength that is not accounted for by self-similarity, and vice versa. After partialling out the effect of self-similarity between repetitions of words, we found that the correlation between neural global pattern similarity within the MTL and subsequent recognition memory confidence remained significant (Fig. 5*C*). Likewise, the correlation between self-similarity in the MTL and recognition confidence remained significant after partialling out the effect of global pattern similarity (Fig. 5*D*).

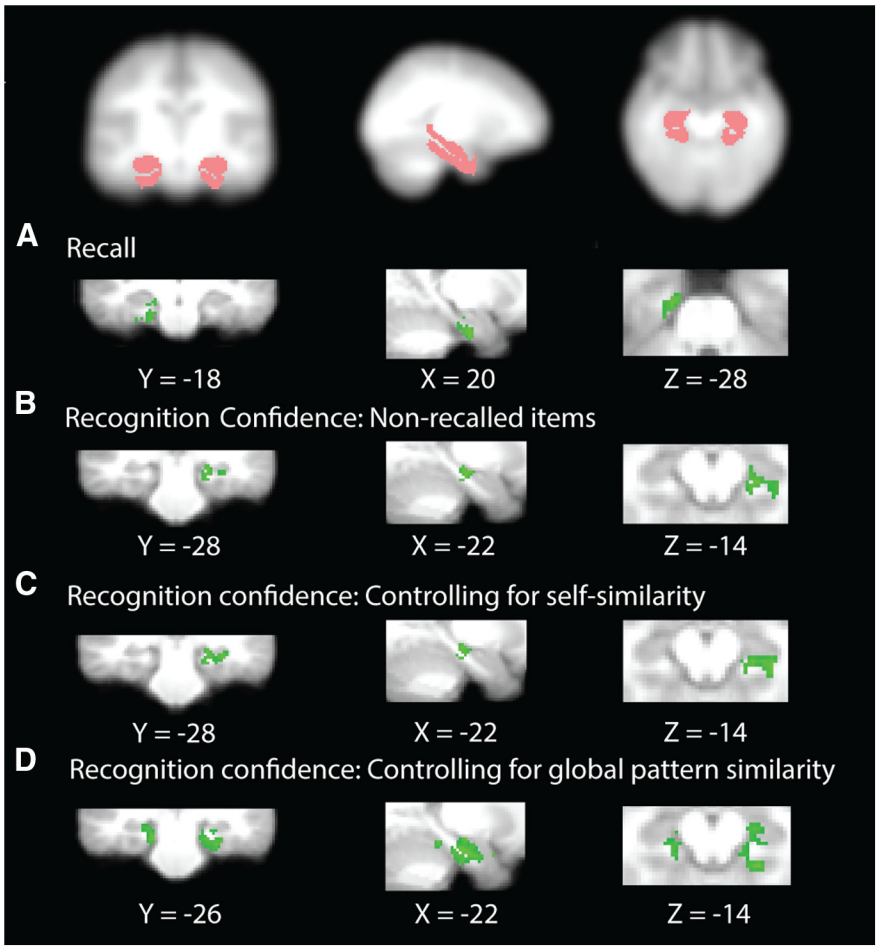

Figure 5. Statistical maps from supporting analyses in the long-term memory task. Statistical maps depict group-level *t* statistics masked with clusters obtained from an MTL-based anatomical correction. ***A***, MTL clusters in which the global pattern similarity measure is significantly correlated with recall. ***B***, MTL clusters in which global pattern similarity is significantly correlated with recognition memory confidence after excluding recalled items. ***C***, MTL clusters in which recognition confidence and global pattern similarity are significantly correlated after partialling out self-similarity. ***D***, MTL clusters in which recognition confidence and self-similarity are significantly correlated after partialling out global pattern similarity. The statistical maps depict group-level *t* statistics masked with clusters obtained from an anatomically based small-volume cluster correction.

Altogether, these four results augment our primary findings by showing that both the pattern similarity of an item to itself across repetitions and its pattern similarity to other items are important for recognition memory judgments. Further, global similarity is associated with subsequent memory, even for the confidently recognized recalled items. These findings are consistent with predictions from mathematical models of human memory and suggest that the neural activation space underlying MTL function may obey the same principles as the psychological spaces underlying these mathematical models.

It is notable that global pattern similarity in both the hippocampus and MTL cortex relates to both recognition memory confidence and recall. In many neurobiological theories of long-term memory, the MTL cortex is thought to be critical for global item-based familiarity processes, whereas the hippocampus is thought to encode more contextual, relational, and associative aspects of memory, akin to the construct of recollection, which is critical for accurate recall (Brown and Aggleton, 2001; Diana et al., 2007; Eichenbaum et al., 2007; but see Squire et al., 2007). Previous neuroimaging results have supported this distinction by showing that activation in the MTL cortex is often graded with respect to subsequent memory performance, whereas activation in the hippocampus tends to be all or none (Ranganath et al., 2004). However, when we test more mechanistic conceptions of memory strength via our global similarity measure, it appears that both the hippocampus and MTL cortex activation patterns contain enough overlap to support familiarity-based processing, and that global similarity is also predictive of recall performance and not just intermediate levels of recognition confidence. Thus, our results may support theories that suggest the MTL as a whole contributes to recollection and familiarity (Squire et al., 2007). Importantly, however, specific anatomical claims regarding MTL subregions must be tempered by the relatively low resolution of the fMRI data in this study (Carr et al., 2010).

Pairwise similarity analysis

The above analyses suggest that words that are remembered more strongly are more central in terms of the neural activation space of the MTL. Because remembered items are also thought to be more globally similar in a psychological memory space, according to models of long-term memory, these results suggest that activation patterns in the MTL and psychological memory representations overlap at least in terms of which items are the most central. However, an important question is whether these two similarity spaces overlap in terms of other representational aspects of the words in the task. Indeed, in many global similarity models, featural or semantic overlap between words is thought to be a key contributor to global similarity and heightened memory strength (Steyvers et al., 2004). Our analyses so far do not directly answer whether semantic similarity is reflected in the similarity relationships between neural activation patterns within the MTL

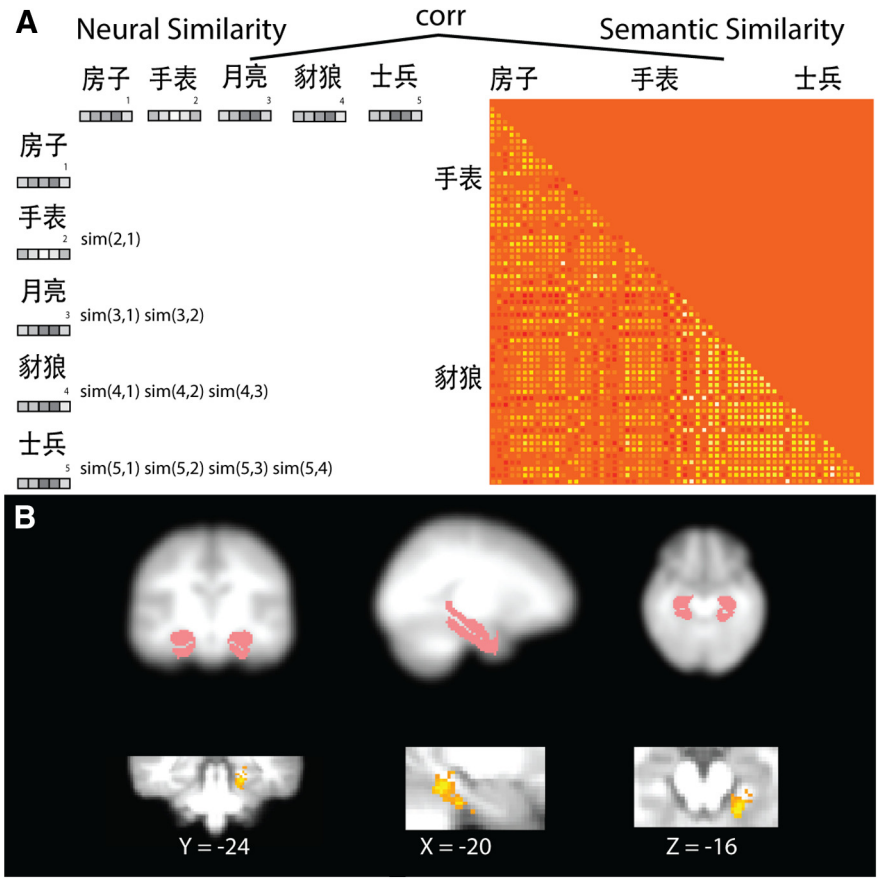

Figure 6. ***A***, A depiction of how the pairwise similarity analysis was conducted in the long-term memory task. A correlation was computed between the lower triangle of the pairwise neural similarity matrix and the lower triangle of the pairwise semantic similarity matrix gathered from the similarity rating task. ***B***, Results of the pairwise similarity analysis. Statistical maps depict clusters in which there was a significant correlation between the neural and semantic similarities between words.

Table 3. MTL small-volume corrected clusters in which a significant correlation was observed between the lower triangle of the pairwise neural similarity matrix and the pairwise semantic similarity matrix gathered in the similarity rating task

Region	Cluster	Cluster size (voxels)	Cluster *p* value	MNI space x	y	z	Peak (*z* score)
Parahippocampal cortex	1	151	0.007	−20	−34	−16	3.85

and thus do not answer whether semantic similarity contributes to the correlation between neural global pattern similarity and recognition confidence.

To test whether the semantic relationships between words are reflected in the activation patterns of the MTL, we collected independent ratings of the semantic similarity between the words used in our long-term memory task and tested how these rated relationships relate to the pairwise neural similarities observed in the task (Fig. 6*A*). Using the same searchlight procedure as used for the primary analysis, we observed a cluster in parahippocampal cortex in which there was a significant correlation between the rated semantic similarity relationships and the pairwise neural similarities between words (Fig. 6*B*; Table 3). These findings suggest that semantic relationships between words may be coded in the MTL and thus may contribute to the heightened neural global pattern similarity for words, as anticipated by formal global similarity models in long-term memory research.

Application 2: categorization task

The above analyses suggest that neural global pattern similarity may drive recognition confidence in long-term memory tasks. Global similarity is also thought to be a key computation underlying behavior in categorization tasks, and thus testing our neural global pattern similarity measure in a categorization task is a critical test of its generality across domains. In the second application, we test how the global similarity between activation patterns elicited for items in a categorization task relates to a model-based measure of psychological memory strength that is estimated from subjects' behavior. This model-based measure, termed recognition strength, is conceptually related to our neural global similarity measure, but instead of measuring how globally similar an activation pattern is compared with those elicited by other stimuli in a task, it estimates how globally similar psychological representations of items are in relation to category representations stored in memory. In a previous analysis of the current dataset (Davis et al., 2012a), significant correlations were observed between this recognition strength measure and trial-by-trial activation in the MTL. In the current analysis, instead of examining trial-by-trial changes in the activation of the MTL, we examine how the psychological recognition strength measure relates to block-by-block measures of global neural similarity for items in the task. This allows us to test whether the MTL may be engaging a global similarity process during categorization, as predicted by similarity-based category learning models and our theory that the MTL engages a global similarity process that supports both long-term memory and categorization behavior.

The categorization task used to test our global neural similarity measure is a rule-plus-exception task in which subjects learn to assign schematic beetles to one of two categories using trial and error. Most of the beetles can be classified using a rule based upon a single dimension (e.g., if the beetle has thick legs, it belongs in Hole A). However, each category also contains an exception item that looks like it ought to belong in the opposing category based on its feature value along the rule dimension. Previous findings suggest that, over the course of learning, the exception items become more prominent in memory. Although there remains some debate about the processes and representations that lead to this exception advantage (Palmeri and Nosofsky, 1995; Sakamoto and Love, 2004; Sakamoto et al., 2004), models that are able to explain the heightened memory strength for exception items predict that it occurs due to a recoding of the stimulus space such that the unique features of exception items are emphasized in memory so that they can be individuated, whereas the unique features of rule-following items are de-emphasized in memory because these features are not important for accurate categorization (Palmeri and Nosofsky, 1995; Sakamoto and Love, 2004; Sakamoto et al., 2004). This recoding causes the global similarity for exception items to increase over the course of learning. Con-

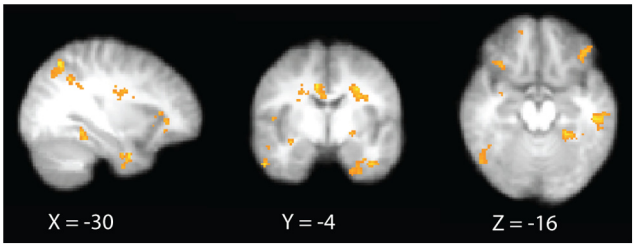

Figure 7. Whole-brain clusters in which there was a significant correlation between the neural global pattern similarity measure and the predicted recognition strength of SUSTAIN in the category learning task.

sistent with these predictions, in a previous analysis of the current dataset, Davis et al. (2012a) found that subjects were more accurate in recognizing exception items in a postlearning recognition memory test. Moreover, Davis et al. (2012a) observed a significant correlation between activation in the MTL during the category learning task and a model-based psychological measure of recognition strength, suggesting that a global similarity process may underlie MTL activation during category learning.

The previously published activation-based analyses raise the question of whether the same basic similarity processes are underlying the activation of the MTL and the model-based recognition strength measure. The model-based mechanisms that predict the greater recognition strength for exception items are engaging a global similarity process; activation, however, only tells how much a region is engaged for different item types. In the present analysis, we attempt to draw a stronger parallel between the model-based global similarity mechanisms that produce the recognition strength measure and the mechanisms underlying MTL function. To do this, we test whether neural global pattern similarity between the MTL activation patterns elicited for different items in the task tracks the model-based recognition strength measure.

We found that there were significant correlations between neural global pattern similarity (Eq. 1) and the model-based global similarity measure in widespread regions of the MTL cortex and hippocampus (Figs. 3*C*, 4*B*; Table 2; for whole-brain results see Fig. 7; Table 4). When searchlights were additionally constrained to anatomical subregions within the MTL, significant clusters were observed bilaterally in hippocampus (right: $p = 0.033$; left: $p = 0.026$), parahippocampal cortex (right: $p = 0.04$; left: $p = 0.004$), and perirhinal cortex (right: $p = 0.004$; left: $p = 0.003$). These results suggest the MTL may perform a global similarity computation during categorization behavior, as predicted by the model-based recognition strength measure.

The present neural global pattern similarity results could not have been predicted to correlate in any specific way with the model-based recognition strength measure based on previous activation results alone. For example, intuitively, the activation-based results may have just as easily predicted the opposite pattern of results: because there are three times as many rule-following items as exceptions, if items differed only with respect to their overall activation level in the MTL, it may have been the case that activation patterns for rule-following items would be more globally similar than exceptions due to their higher frequency.

Pairwise similarity analysis

Even though the present similarity results are not predictable from previous activation-based results, as with the long-term memory results, it is important to further assess what information about the stimuli is present in the MTL activation patterns.

Table 4. Whole-brain corrected clusters in which a significant correlation was observed between the neural global pattern similarity and the recognition strength measure of SUSTAIN in the categorization task

Region	Cluster size (voxels)	Cluster *p* value	MNI space *x*	*y*	*z*	Peak (*z* score)
Anterior cingulate	780	<0.001	12	−12	30	8.59
Posterior cingulate	685	<0.001	4	−46	36	9.57
Putamen	470	<0.001	26	14	0	6.95
Lateral occipital cortex, superior division	441	0.001	−28	−62	52	8.13
Inferior temporal gyrus	375	0.002	50	−52	−22	9.13
Insula	330	0.003	−22	26	2	7.22
Frontal pole	282	0.004	6	56	−6	6.88
Inferior temporal gyrus	275	0.004	−56	−28	−18	7.6
Middle temporal gyrus	251	0.005	56	2	−28	8.91
Cuneal cortex	247	0.006	−2	−88	14	7.48
Posterior cingulate	234	0.006	16	−44	4	8.57
Supracalcarine cortex	229	0.006	−22	−60	14	8.22
Anterior cingulate	225	0.006	−26	−2	34	8.09
Frontal pole	188	0.009	8	48	−26	7.85
Lateral occipital cortex, inferior division	185	0.009	40	−60	6	13.5
Anterior parahippocampal gyrus	177	0.01	−28	0	−32	6.69
Temporal fusiform	176	0.01	−24	−40	−20	11.6
Precentral gyrus	144	0.014	8	−26	56	7.74
Frontal pole	136	0.016	−46	38	−8	8.39
Inferior frontal gyrus	134	0.016	52	20	28	7.38
Temporal pole	115	0.022	−40	6	−40	8.53
Putamen	110	0.024	32	−10	−12	7.33
Cerebellum	110	0.024	18	−62	−28	5.86
Thalamus	104	0.027	−16	−24	6	7.74
Middle temporal gyrus	78	0.041	−48	−54	0	6.64
Frontal pole	77	0.043	−8	58	30	5.71
Superior temporal gyrus	76	0.043	58	−18	2	6.12
Putamen	74	0.043	−18	6	−2	7.52
Angular gyrus	73	0.044	30	−48	22	6.19
Frontal pole	69	0.048	12	66	2	7.05
Temporal pole	69	0.048	46	16	−32	7.01
Cerebellum	67	0.05	−26	−54	−40	7.38

Peaks represent the maximum correlation within each cluster.

Here we use pairwise similarity analysis to examine how the beetle space is represented in the MTL such that exception items are more globally similar than rule-following items, even though both item types are equally globally similar in terms of the physical/perceptual beetle space (i.e., feature values are evenly distributed across rule-following and exception items; Table 1). If the MTL is recoding the space like the global-similarity models that are able to learn rule-plus-exception tasks, it should de-emphasize the unique features of rule-following items and differentiate rule-following items primarily based upon which category they belong in. Likewise, unique features of the exception items should be emphasized in memory so that these items can be individuated.

One critical aspect of the category structure is that each item in the task mirrors the non-rule-following dimension features of an item in the opposing category (Table 1). This property of the category structure allows us to test whether exceptions are more differentiated in memory than rule-following items by testing whether exceptions are more similar to their mirrored items (i.e., other exceptions) than rule-following items are to their mirrored items (i.e., other rule-following items).

Consistent with the hypothesis that the exception item features are more differentiated in memory, we found that exception items from opposing categories were more similar than rule-following items from opposing categories ($t_{(14)} = 3.37$, $p = 0.005$).

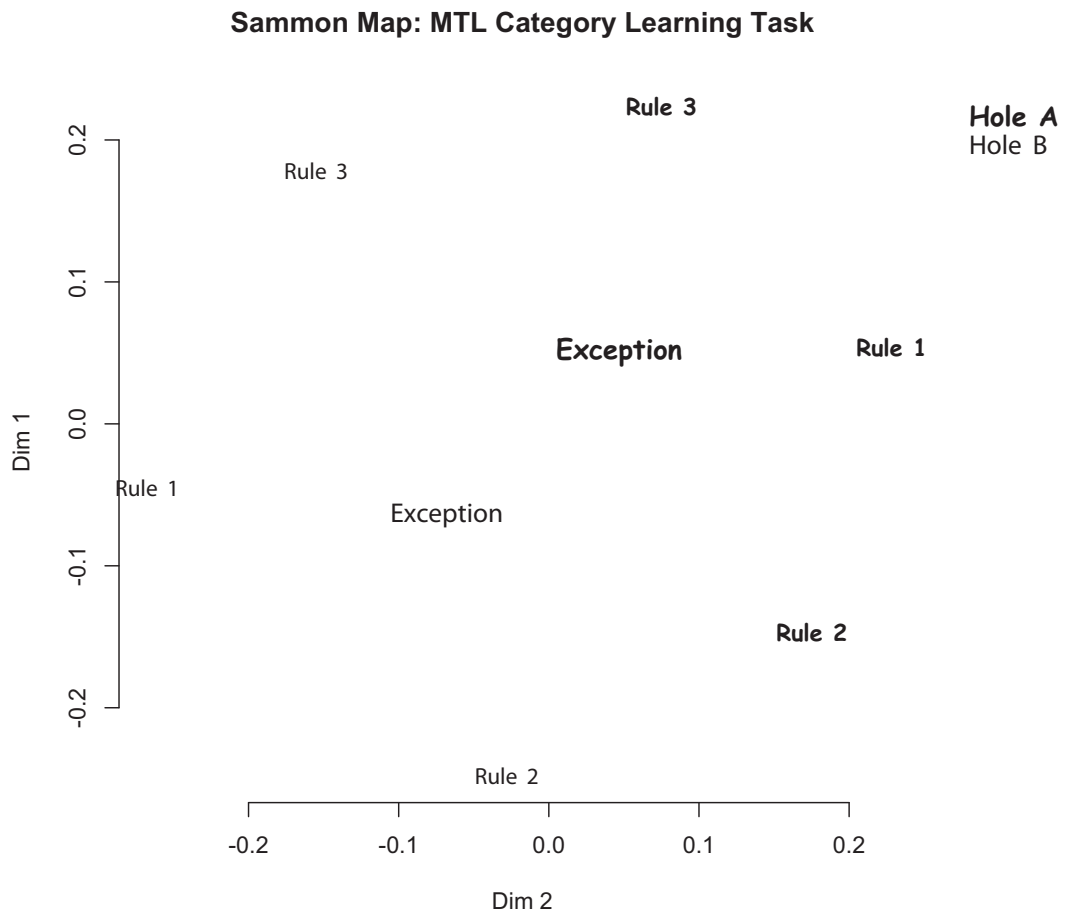

Figure 8. A Sammon map of pairwise similarities between MTL activation patterns elicited for the different item types at the end of the category learning task (sixth scanning run). Rules 1, 2, and 3 refer to different rule-following items used in the task (Table 1). Hole A beetles are depicted in Comic Sans, and Hole B beetles are depicted in Papyrus. The Sammon map projects the pairwise similarities observed in the final run of the category learning task onto a two-dimensional space. The two dimensions on the *x*- and *y*-axes depict the two dimensions of the Sammon map. Distances between items in the plot depict dissimilarities between their respective activation patterns in the reduced space.

As a second test of whether the activation patterns of the MTL contained information about the category structure, we tested whether rule-following items from the same category were more similar than rule-following items from opposing categories. Consistent with the hypothesis that the MTL contains information about the category structure, we found that rule-following items were more similar to the other rule-following members in their category than to rule-following items in the opposing category ($t_{(14)} = 2.24$, $p = 0.04$). These relationships between items were captured in a two-dimensional multidimensional scaling (i.e., Sammon mapping) of the pairwise similarities observed in the MTL at the end of training (Fig. 8). Stress, a measure of how well the two-dimensional solution accounted for the pairwise similarities between the items, was 0.07, indicating a good fit. In the Sammon map, exceptions are most central and close to each other, whereas rule-following items from opposing categories are clustered on opposite sides of the space.

Together, these results suggest that the MTL activation patterns present a recoding of the physical stimulus space into a space in which exceptions are represented as the most central—similar in some respects to both categories—whereas rule-following items from opposing categories are represented as dissimilar even when they otherwise share the same number of common physical features with the opposing category as exceptions do. Importantly, these results suggest that the activation patterns of the MTL contain important information about category membership, and not just differences in activation between exceptions and rule-following items, as demonstrated in the study by Davis et al. (2012a).

Conjunction analysis across studies

We conducted a conjunction analysis to examine the degree to which common regions were involved in global similarity processing across the long-term memory and categorization studies. The regions in which global pattern similarity in the MTL correlated with memory strength showed considerable overlap between the two tasks (Fig. 4*C*). This overlap suggests that computational processes supporting familiarity or memory strength appear to share anatomical substrates in the MTL across both task types.

Discussion

Similarity-based models from several cognitive domains posit a central role for global similarity in computations of memory strength. Despite this central role for global similarity in cognitive theory, neurobiological studies on the neural basis of memory strength have largely relied upon univariate activation measures as surrogates for global similarity processes. Here we developed a novel method for quantifying the global similarity of an item with respect to patterns of activation elicited for all stimuli within a task. Analysis of independent categorization and long-term memory datasets revealed significant correlations between this

neural global pattern similarity measure and psychological measures of memory strength thought to derive from global similarity. These results suggest that similarity between activation patterns in the brain can contain important information about cognitive states and may allow the testing of cognitive theories at a finer-grained level than previously thought.

Collectively, our results significantly extend the findings of related categorization and long-term memory studies and previously published analyses of the current datasets (Xue et al., 2010; Davis et al., 2012a). Foremost, the present results highlight a mechanistic overlap between MTL-based categorization and long-term memory processes that could not be determined based on previous univariate analyses of MTL activation (Davis et al., 2012a). According to mathematical categorization and long-term memory models, both categorization and long-term memory engage a global similarity process (Gillund and Shiffrin, 1984; Hintzman, 1988; Nosofsky, 1988, 1991; Norman and O'Reilly, 2003), but the activation-based measures previously used do not provide any information about the similarity relationships between items in terms of MTL processing. The present results strengthen these theories by showing that the similarity relationships between activation patterns elicited for items are related to measures of memory strength in both task domains. Likewise, our results strengthen global similarity theory in both domains by revealing that both the similarity of an item to its own activation patterns and its similarity to activation patterns of other items contribute to memory strength.

Despite the strong formal relationships between categorization and long-term memory models (Estes, 1996; Love and Gureckis, 2007; Nosofsky et al., 2012), the idea that the same mechanisms support similarity processing in both domains has been controversial. Based on evidence that amnesic individuals could learn some categorization tasks but could not recognize the same stimuli in follow-up memory tasks, early neuropsychological research suggested that the neural mechanisms for categorization and memory diverged substantially (Knowlton and Squire, 1993). More recent neuroimaging research has demonstrated that the MTL plays a critical role in categorization and recognition at various points (Poldrack et al., 2001; Seger et al., 2011; Davis et al., 2012a,b), and for particular types of categorization problems (Reber et al., 2003; for review, see Ashby and Maddox, 2005; Zeithamova et al., 2008; Seger and Miller, 2010). Furthermore, global similarity models that use the same representations for categorization and recognition memory have been successful at accounting for patterns of patient and imaging data in both types of tasks (Nosofsky and Zaki, 1998; Love and Gureckis, 2007; Nosofsky et al., 2012).

Our study is the first to use similarity-based analysis to directly test whether overlap between activation patterns in the MTL drives memory strength in both categorization and long-term memory tasks. However, it is important to note several potential points for consideration for future studies. First, rule-plus-exception tasks may be unique in encouraging subjects to use explicit MTL-based memory processes to memorize the exceptions, particularly when subjects are cued with the rule and told beforehand that there will be exceptions, as we did in our experiment. Thus, it will be important to test how the neural global similarity measure relates to memory strength in a number of different categorization tasks, particularly ones that may not depend on the MTL (Reber et al., 2003; Nomura et al., 2007; Zeithamova et al., 2008). Second, our results and theory do not suggest that categorization and long-term memory will overlap in all of their underlying psychological and neural mechanisms. Indeed, both categorization and long-term memory likely contain a number of subprocesses beyond similarity and memory strength computations that do not overlap (e.g., hypothesis testing, response selection), and are not hypothesized to depend upon the MTL (Poldrack and Foerde, 2008).

Our global pattern similarity measure shares relationships with recent work examining how neural pattern similarities relate to long-term memory. Recent studies have examined how pattern similarities between repetitions of an item (Xue et al., 2010), between an item and other members of its category (Kuhl et al., 2012), or similarity to other categories (LaRocque et al., 2013) relate to long-term memory. Contrastingly, our neural global pattern similarity measure follows models of categorization and memory in suggesting that the similarity relationships of an item to all items combine additively to influence memory. However, this does not mean that each of the different components that go into global similarity (at the item level, within a category, and between categories) will always be positively correlated with memory. For example, in our rule-plus-exception task, the similarity of the exception to members of the opposing category is predicted to be a central part of their high memory strength. In other contexts, the relationship between similarity to opposing categories and memory strength may be negative (e.g., when categories are defined by well separated prototypes). Future studies will need to be developed to fully test the consequences of this additive model in a variety of designs with well defined representational structures.

One important question with respect to the present results is how neural pattern similarity relates to similarity computations in long-term memory and categorization models (Davis and Poldrack, 2013b). In formal long-term memory and categorization models, familiarity or memory strength is modeled as arising from global representational similarity (Gillund and Shiffrin, 1984; Hintzman, 1988; Nosofsky, 1988, 1991; Norman and O'Reilly, 2003). Depending upon the specific model, this representational similarity can include any information about the overlap in featural, category-level, or other contextual or associative information between stimuli in a task. Multivoxel activation patterns like those that make up our global similarity measure are often interpreted in the broader literature as measures of neural representation (Haxby et al., 2001; Eger et al., 2008; Kriegeskorte et al., 2008; Weber et al., 2009; for review, see Davis and Poldrack, 2013b). If this representational interpretation is correct, the neural global pattern similarity measure is related directly to computations in formal models.

It is critical to note, however, that multivoxel activation patterns likely contain a number of nonrepresentational signals that can overlap between stimuli, such as information about engagement of cognitive processes or anything else that differs between the stimuli (Todd et al., 2013). For example, it is possible that activation patterns contain voxels that are more reliably engaged due to higher signal-to-noise ratio (SNR) for strongly remembered items, and that this more reliable activation is driving high neural global pattern similarity for strongly remembered items as opposed to representational overlap per se. For example, analogous to their action on visual processing (Boynton, 2005), attentional processes may increase the reliability of a common memory strength activation pattern for strongly remembered items via decreases in SNR. Indeed, although our correlation-based measure is insensitive to mean activation or average differences in variability between activation patterns for stimuli (Kriegeskorte et al., 2008), recent findings suggest that such modulation of shared activation patterns will result in increased

shared variability between items, which is not easily removed (LaRocque et al., 2013; Davis et al., 2014).

Several pieces of evidence argue against the hypothesis that engagement of memory strength processes is the only information that is contained in the activation patterns of the MTL in the present experiments. First, in the category learning task, rule-following members were found to elicit activation patterns that are similar to members of their own category and dissimilar to members of the opposing category, even though rule-following items from opposing categories would be associated with equivalent memory strength or attentional modulation of SNR. Likewise, in the long-term memory task, information related to the semantic relationships between words was found to be coded in the activation patterns of the MTL, suggesting that the activation patterns of the MTL contain information about the psychological representations of stimuli and not just how strongly they are remembered. Critically, however, our results do not definitively indicate that cognitive models and the neural global pattern similarity measure are taking into account identical representational information. To draw stronger parallels between our global pattern similarity measure and cognitive models, it will be important for future research to develop techniques that are increasingly able to directly measure the informational contents of activation patterns.

In conclusion, formal cognitive models posit that global similarity processes play a key role in cognitive processing in a variety of domains. Using a novel neural global pattern similarity measure, we found that MTLs function in both categorization and long-term memory tasks may operate via the same principles as predicted by global similarity models. These results extend our knowledge of the neural processes that give rise to memory and suggest a remarkable consistency in terms of the neural mechanisms that support cognition across categorization and long-term memory domains.

Notes

Supplemental material for this article is available at https://drive.google.com/file/d/0Bz-7C2DKgeKoaTlkVkVfa2dpaDQ/edit?usp=sharing, consisting of model equations and fit details for SUSTAIN. This material has not been peer reviewed.

References

Ashby FG, Maddox WT (2005) Human category learning. Annu Rev Psychol 56:149–178. CrossRef Medline

Boynton GM (2005) Attention and visual perception. Curr Opin Neurobiol 15:465–469. CrossRef Medline

Brown MW, Aggleton JP (2001) Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat Rev Neurosci 2:61–62. CrossRef Medline

Carr VA, Rissman J, Wagner AD (2010) Imaging the human medial temporal lobe with high-resolution fMRI. Neuron 65:298–308. CrossRef Medline

Clark SE, Gronlund SD (1996) Global matching models of recognition memory: how the models match the data. Psychon Bull Rev 3:37–60. CrossRef Medline

Daselaar SM, Fleck MS, Cabeza R (2006) Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. J Neurophysiol 96:1902–1911. CrossRef Medline

Davis T, Poldrack RA (2013a) Quantifying the internal structure of categories using a neural typicality measure. Cereb Cortex. Advance online publication. Retrieved April 27, 2014. doi:10.1093/cercor/bht014. CrossRef Medline

Davis T, Poldrack RA (2013b) Measuring neural representations with fMRI: practices and pitfalls. Ann N Y Acad Sci 1296:108–134. CrossRef Medline

Davis T, Love BC, Preston AR (2012a) Learning the exception to the rule: model-based fMRI reveals specialized representations for surprising category members. Cereb Cortex 22:260–273. CrossRef Medline

Davis T, Love BC, Preston AR (2012b) Striatal and hippocampal entropy and recognition signals in category learning: simultaneous processes revealed by model-based fMRI. J Exp Psychol Learn Mem Cogn 38:821–839. CrossRef Medline

Davis T, LaRocque KL, Mumford JA, Norman KA, Wagner AD, Poldrack RA (2014) What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. Neuroimage. Advance online publication. Retrieved April 21, 2014. doi:10.1016/j.neuroimage.2014.04.037. CrossRef Medline

Daw ND (2011) Trial-by-trial data analysis using computational models. In: Decision making, affect, and learning: attention and performance XXIII (Delgado MR, Phelps EA, Robbins TW, eds), pp 3–38. Oxford, UK: Oxford UP.

Dennis S, Humphreys MS (2001) A context noise model of episodic word recognition. Psychol Rev 108:452–478. CrossRef Medline

Diana RA, Yonelinas AP, Ranganath C (2007) Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn Sci 11:379–386. CrossRef Medline

Eger E, Ashburner J, Haynes JD, Dolan RJ, Rees G (2008) fMRI activity patterns in human loc carry information about object exemplars within category. J Cogn Neurosci 20:356–370. CrossRef Medline

Eichenbaum H, Yonelinas AP, Ranganath C (2007) The medial temporal lobe and recognition memory. Annu Rev Neurosci 30:123–152. CrossRef Medline

Estes WK (1996) Classification and cognition. New York: Oxford UP.

Forstmann BU, Wagenmakers EJ, Eichele T, Brown S, Serences JT (2011) Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract? Trends Cogn Sci 15:272–279. CrossRef Medline

Friston K (2012) Ten ironic rules for non-statistical reviewers. Neuroimage 61:1300–1310. CrossRef Medline

Gillund G, Shiffrin RM (1984) A retrieval model for both recognition and recall. Psychol Rev 91:1–67. CrossRef Medline

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430. CrossRef Medline

Hintzman DL (1988) Judgments of frequency and recognition memory in a multiple-trace memory model. Psychol Rev 95:528. CrossRef

Islam A, Inkpen D (2006) Second order co-occurrence PMI for determining the semantic similarity of words. Paper presented at the International Conference on Language Resources and Evaluation 2006, Genoa, Italy, May.

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

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868. CrossRef Medline

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—connecting the branches of systems neuroscience. Front Syst Neurosci 2:4. CrossRef Medline

Kuhl BA, Rissman J, Wagner AD (2012) Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. Neuropsychologia 50:458–469. CrossRef Medline

Landauer TK, Dumais ST (1997) A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol Rev 104:211–240. CrossRef

LaRocque KF, Smith ME, Carr VA, Witthoft N, Grill-Spector K, Wagner AD (2013) Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. J Neurosci 33:5466–5474. CrossRef Medline

Lesot MJ, Rifqi M, Benhadda H (2009) Similarity measures for binary and numerical data: a survey. International Journal of Knowledge Engineering and Soft Data Paradigms 1:63–84. CrossRef

Love BC, Gureckis TM (2007) Models in search of a brain. Cogn Affect Behav Neurosci 7:90–108. CrossRef Medline

Love BC, Medin DL, Gureckis TM (2004) SUSTAIN: a network model of category learning. Psychol Rev 111:309–332. CrossRef Medline

Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643. CrossRef Medline

Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitelman DR, Parrish TB,

Mesulam MM, Reber PJ (2007) Neural correlates of rule-based and information-integration visual category learning. Cereb Cortex 17:37–43. CrossRef Medline

Norman KA, O'Reilly RC (2003) Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol Rev 110:611–646. CrossRef Medline

Nosofsky RM (1988) Exemplar-based accounts of relations between classification, recognition, and typicality. J Exp Psychol Learn Mem Cogn 14: 700. CrossRef

Nosofsky RM (1991) Tests of an exemplar model for relating perceptual classification and recognition memory. J Exp Psychol Hum Percept Perform 17:3–27. CrossRef Medline

Nosofsky RM, Zaki SR (1998) Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. Psychol Sci 9:247. CrossRef

Nosofsky RM, Little DR, James TW (2012) Activation in the neural network responsible for categorization and recognition reflects parameter changes. Proc Natl Acad Sci U S A 109:333–338. CrossRef Medline

O'Reilly RC, Norman KA (2002) Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. Trends Cogn Sci 6:505–510. CrossRef Medline

Palmeri TJ, Nosofsky RM (1995) Recognition memory for exceptions to the category rule. J Exp Psychol Learn Mem Cogn 12:548–568. CrossRef Medline

Poldrack RA, Foerde KF (2008) Category learning and the memory systems debate. Neurosci Biobehav Rev 32:197–205. CrossRef 17869339

Poldrack RA, Clark J, Paré-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA (2001) Interactive memory systems in the human brain. Nature 414:546–550. CrossRef Medline

Raaijmakers JG, Shiffrin RM (1992) Models for recall and recognition. Annu Rev Psychol 43:205–234. CrossRef Medline

Ranganath C, Yonelinas AP, Cohen MX, Dy CJ, Tom SM, D'Esposito M (2004) Dissociable correlates of recollection and familiarity within the medial temporal lobes. Neuropsychologia 42:2–13. CrossRef Medline

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297: 1551–1555. 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

Sakamoto Y, Love BC (2004) Schematic influences on category learning and recognition memory. J Exp Psychol Gen 133:534–553. CrossRef Medline

Sakamoto Y, Love BC (2006) Vancouver, Toronto, Montreal, Austin: enhanced oddball memory through differentiation, not isolation. Psychon Bull Rev 13:474–479. CrossRef Medline

Sakamoto Y, Matsuka T, Love BC (2004) Dimension-wide vs. exemplar-specific attention in category learning and recognition. In: Proceedings of the Sixth International Conference on Cognitive Modeling (Lovett MC, Schunn CD, Lebiere C, Munro P, eds), pp 261–266. Abingdon, UK: Taylor & Francis.

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

Seger CA, Dennison CS, Lopez-Paniagua D, Peterson EJ, Roark AA (2011) Dissociating hippocampal and basal ganglia contributions to category learning using stimulus novelty and subjective judgments. Neuroimage 55:1739–1753. CrossRef Medline

Shepard RN (1987) Toward a universal law of generalization for psychological science. Science 237:1317–1323. CrossRef Medline

Shiffrin RM, Steyvers M (1997) A model for recognition memory: REM—retrieving effectively from memory. Psychon Bull Rev 4:145–166. CrossRef Medline

Squire LR, Wixted JT, Clark RE (2007) Recognition memory and the medial temporal lobe: a new perspective. Nat Rev Neurosci 8:872–883. CrossRef Medline

Stark CE, Squire LR (2001) When zero is not zero: the problem of ambiguous baseline conditions in fMRI. Proc Natl Acad Sci U S A 98:12760–12766. CrossRef Medline

Steyvers M, Shiffrin RM, Nelson DL (2004) Word association spaces for predicting semantic similarity effects in episodic memory. In: Experimental cognitive psychology and its applications (Healy AF, ed), pp 237–249. Washington, DC: American Psychological Association.

Todd MT, Nystrom LE, Cohen JD (2013) Confounds in multivariate pattern analysis: theory and rule representation case study. Neuroimage 77: 157–165. CrossRef Medline

Weber M, Thompson-Schill SL, Osherson D, Haxby J, Parsons L (2009) Predicting judged similarity of natural categories from their neural representations. Neuropsychologia 47:859–868. CrossRef Medline

Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA (2010) Greater neural pattern similarity across repetitions is associated with better memory. Science 330:97–101. 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