



Neural pattern similarity across concept exemplars predicts memory after a long delay

Heather Bruett^{a,b,*}, Regina C. Calloway^{a,b}, Natasha Tokowicz^{a,b}, Marc N. Coutanche^{a,b,c}

^a Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

^b Learning Research & Development Center, University of Pittsburgh, Pittsburgh, PA, USA

^c Brain Institute, University of Pittsburgh, Pittsburgh, PA, USA

ARTICLE INFO

Keywords:

Episodic
Semantic
Memory
Pattern similarity
Encoding
Retrieval

ABSTRACT

The irregularities of the world ensure that each interaction we have with a concept is unique. In order to generalize across these unique encounters to form a high-level representation of a concept, we must draw on similarities between exemplars to form new conceptual knowledge that is maintained over a long time. Two neural similarity measures — pattern robustness and encoding-retrieval similarity — are particularly important for predicting memory outcomes. In this study, we used fMRI to measure activity patterns while people encoded and retrieved novel pairings between unfamiliar (Dutch) words and visually presented animal species. We address two underexplored questions: 1) whether neural similarity measures can predict memory outcomes, despite perceptual variability between presentations of a concept and 2) if pattern similarity measures can predict subsequent memory over a long delay (i.e., one month). Our findings indicate that pattern robustness during encoding in brain regions that include parietal and medial temporal areas is an important predictor of subsequent memory. In addition, we found significant encoding-retrieval similarity in the left ventrolateral prefrontal cortex after a month's delay. These findings demonstrate that pattern similarity is an important predictor of memory for novel word-animal pairings even when the concept includes multiple exemplars. Importantly, we show that established predictive relationships between pattern similarity and subsequent memory do not require visually identical stimuli (i.e., are not simply due to low-level visual overlap between stimulus presentations) and are maintained over a month.

1. Introduction

We often interact with the same concepts throughout our lives, but we rarely encounter a concept in the same way each time. The definition of “concept” has taken many forms over the centuries (Smith and Medin, 1981), but one of its key characteristics is helping to efficiently connect the varied featural information associated with a particular item, distributed largely across occipital and temporal regions (Tyler et al., 2004; Tyler and Moss, 2001). The concept of “giraffe,” for instance, is connected to features such as tall, spotted, and animate, as well as its semantic label. To effectively understand and recognize a “giraffe”, a person must generalize low-level perceptual differences that are irrelevant to the core concept (such as small variations in coloring or size) and retain this knowledge over a long period of time.

Successful memory retrieval can be predicted based on similarity in patterns of neural activity. This is true across presentations during

encoding (pattern robustness; LaRocque et al., 2013), as well as between encoding and retrieval (encoding-retrieval similarity, ERS; Ritchey et al., 2013). When information is first encoded into memory, the robustness (reliability over multiple trials) of the neural pattern during encoding predicts how well the association is later remembered (LaRocque et al., 2013; Xiao et al., 2016). The directionality of this robustness measure (i.e., higher or lower similarity within, compared to than between, conditions) can reflect two alternative forms of representation. The most commonly examined direction (within > between) indicates that an item is robustly represented relative to other items. The alternative direction (between > within) has also been observed in the perirhinal cortex (PrC) and parahippocampal cortex (PHC), though the reason for this is, as yet, unknown (LaRocque et al., 2013). The second form of pattern similarity, the degree of pattern reinstatement between encoding and retrieval (within > between ERS), also positively predicts subsequent memory performance (Kuhl and Chun, 2014; Lee et al., 2018; Mack and Preston,

* Corresponding author. Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA.

E-mail address: heb52@pitt.edu (H. Bruett).

<https://doi.org/10.1016/j.neuroimage.2020.117030>

Received 18 March 2020; Received in revised form 14 May 2020; Accepted 5 June 2020

Available online 8 June 2020

1053-8119/© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2016; Ritchey et al., 2013; Xue et al., 2010).

These signatures of neural similarity have been linked to subsequent memory for *particular* word-image associations (e.g., Ritchey et al., 2013; Staresina et al., 2012; Wing et al., 2015), but truly having a concept in memory requires a neural activity pattern that transcends specific images to generalize across exemplars and view-points — a requirement that has not been tested for pattern similarity. It is thus not known whether the predictive ability of pattern similarity is restrained to simple and very specific word-image episodic associations, or also applies to forming conceptual knowledge, where words are associated with a variety of visual presentations (e.g., giraffes from multiple angles, exemplars, and shading). A second feature of conceptual knowledge is its long-term nature (Squire and Alvarez, 1995). Studies of pattern similarity and memory have typically tested encoding/retrieval delays of 5 min (Kuhl and Chun, 2014; Lee et al., 2018; Mack and Preston, 2016; Ward et al., 2013) or closer to an hour (Xiao et al., 2016; Xue et al., 2010), with only occasional delays of up to a week (Tompary and Davachi, 2017; Wirebring et al., 2015), leaving uncertainty over whether any conceptual pattern similarity would maintain a relationship with subsequent memory across a longer time period, such as a month. Here, we investigate how two forms of pattern similarity – pattern robustness and ERS – predict subsequent memory for novel word-item associations, where an association cannot be accounted for by a single image (and associated low-level features), and where memory is tested one month after encoding.

The question of whether neural signatures of successful memory encoding are linked to low-level visual features (or allow cross-image generalization) is particularly critical for analysis approaches that draw on multi-voxel patterns, which can pick-up on visually-specific predictors of subsequent memory (Kuhl and Chun, 2014; Lee et al., 2018; Mack and Preston, 2016; Ward et al., 2013). Unlike univariate approaches (Coutanche, 2013), multivariate analyses are sensitive to differences in visual stimuli with great specificity, including at the superordinate level (e.g., birds vs. mammals; Connolly et al., 2012). Indeed, at a given moment, the brain represents visual stimuli at multiple levels of granularity. For example, when one retrieves memory for a known concept such as a carrot, its perceptual features, such as its shape and color information are represented in the lateral occipital complex (LOC) and V4, respectively, while its object-level representation is elsewhere (Coutanche and Thompson-Schill, 2015). This is in line with the hub-and-spoke model of semantic memory (Lambon Ralph, 2014), which suggests the memory for a known concept is encoded within regions that represent their features (“spokes”), as well as in a higher-level hub, such as the anterior temporal lobe (ATL) and angular gyrus (AG) (Lambon Ralph, 2014; Lambon Ralph et al., 2017).

With the increased sensitivity in multi-voxel patterns, new methodological considerations arise. With the ability to detect information at the single-image level (Kay et al., 2008), it becomes necessary to actively examine whether a memory-relevant metric of pattern similarity is drawing on such single-image representations, or more generalizable cross-image concepts, not least because visual features at the image-level can predict the memorability of a stimulus (Koch et al., 2019). To examine neural representations for concepts, rather than images, it is necessary to take efforts to ensure that robustness and ERS effects reflect more than overlapping low-level visual patterns, which cannot be ruled-out from past studies where associations are learned with just one image across all encoding and retrieval trials (Lee et al., 2018; Mack and Preston, 2016; Ward et al., 2013; Xue et al., 2010). In this study, we aim to mitigate perceptually-driven effects by examining ERS of the same concept (e.g., a species of bird) from different visual angles (e.g., standing face-forward; sitting looking to the side) and exemplars (different members of the species). By incorporating diverse images into the concept being learned, we decrease the likelihood that effects are driven by perceptual similarity between encoding trials or between encoding and retrieval, and instead test for image-invariant conceptual representations, leading to the hypothesis that we will observe ERS in regions

relevant to conceptual processing (i.e., semantic hubs, ATL and AG).

The shift from images to concepts brings an additional consideration in the form of semantic granularity, such as a concept’s item- (e.g., “sparrow”) and superordinate- (e.g., “bird”) levels, to consider just two. A person’s ability to access a learned concept through these dimensions is important for memory, with some levels being remembered, while others are not (Lee et al., 2018). To foreshadow our findings, our results indicate that regions of the brain show robustness and ERS that is predictive of subsequent memory at a relatively fine level of analysis (e.g., bird vs. mammal). This itself is a unique finding given that it is more common to examine pattern similarity at much coarser levels of granularity (e.g., faces vs. objects vs. scenes; Lee et al., 2018). Given our success at this fine level of granularity, we followed up our superordinate-level analyses with additional analyses at an even finer level (e.g., bird A vs. bird B) to ask whether similarity at the two levels converge or diverge.

Most studies investigating pattern robustness and ERS have also relied on demonstrating a maintenance of patterns over the course of a few minutes (e.g., Kuhl and Chun, 2014; Lee et al., 2018; Mack and Preston, 2016; Wirebring et al., 2015) or several days (e.g., Tompary and Davachi, 2017). One study investigated neural similarity between encoding and retrieval over a longer time span (i.e. 2-6 weeks), but did so using a fear conditioning paradigm, a type of associative learning that differs from declarative memory in a number of respects (Coutanche and Thompson-Schill, 2012; Visser et al., 2013). Importantly, the typical time span examined in these studies is relatively restricted compared to the timespan of a typical memory, particularly given that the neural basis for memories can continue to change for weeks, months and even years after initial encoding (Squire and Alvarez, 1995). Our study, on the other hand, tests patterns after a longer delay period (i.e., one month), which is also more aligned with the focus on longer-term conceptual processing, rather than word-image episodic associations. The timeline of pattern similarity effects is particularly important to consider when examining the encoding of novel (i.e., unfamiliar) stimuli. Although previously familiar stimuli will have existing representations in memory that can help support new associations being formed in an episodic memory task, novel stimuli (e.g., an unfamiliar word and object) will not (van Kesteren et al., 2012). Here, we present participants with pairings of novel words and unfamiliar animals, to examine how representations are first formed, unlike prior studies that examined associations between familiar concepts. This further motivates the importance of examining longer time-courses than prior ERS studies, to allow for both item- and association-based memories to form and be consolidated over time.

In this study, we scanned individuals using functional magnetic resonance imaging (fMRI) to probe the information present in multi-voxel patterns as people encoded 60 pairs of novel Dutch words and unfamiliar animals, and later retrieved the pairings one month later. Based on our design of incorporating visual images and novel words, we hypothesized that regions associated with high-level visual and semantic level representations (e.g., VT) would contain patterns predictive of subsequent memory, as would the visual word form area (VWFA), which contains patterns that distinguish known words from non-words (e.g., Carlos et al., 2019). Regions in the medial temporal lobe (HC, PHC, PrC), the left ventrolateral prefrontal cortex (vlPFC), and medial parietal cortex (MPC) were also predicted to show ERS due to their involvement in memory encoding and retrieval (Badre and Wagner, 2007; Davachi et al., 2003; Hutchinson et al., 2009; Rugg and King, 2018; Staresina et al., 2012; Wagner et al., 2016). Finally, we did not expect that early visual cortex (EVC) would show significant ERS, based on the higher-level conceptual ERS we examined, which is encoded in multi-voxel patterns later in the ventral stream (Coutanche et al., 2016).

2. Materials and methods

2.1. Participants

Data from 22 right-handed participants (female = 14; male = 8; Age:

Mean (M) = 25.3, standard deviation (SD) = 5.0) were collected over the course of three sessions. The first two sessions occurred on consecutive days, and the final session took place approximately one month later (range = 24–44 days after the first session). All participants provided written informed consent at the start of each day of data collection and were compensated for their time. At the start of the first session, participants completed a demographics questionnaire and color blindness test. All participants received perfect scores on the color blindness test and were free of psychiatric, neurological, learning, and attention disorders. All participants were native English speakers who did not learn any languages other than English in their home growing up and had no prior experience with Dutch or German languages. Two participants' data were removed due to attrition on the final day of the study (giving $n = 20$). All procedures were approved by the University of Pittsburgh Institutional Review Board. Any data associated with this article will be made available by reasonable request in a way that complies with policies of the University of Pittsburgh Institutional Review Board.

2.2. Design

Session I (approximately 1 h) consisted primarily of fMRI data collection (see Fig. 1), including an anatomical scan, localizer run, and two functional runs of encoding associations in Set 1. After scanning, participants completed a cued recall task. Session II (approximately 2 h) incorporated retrieval of the prior day's set of associations, followed by encoding associations from Set 2, while EEG data was collected (not analyzed here). Participants then completed a cued recall task for Set 2. Session III (approximately 1.5 h) consisted of another fMRI scan, with an anatomical scan, three runs of a retrieval task for Set 1, followed by three runs of a retrieval task for Set 2. After the scan, participants completed a categorization task (not analyzed here) and final survey of pre-experimental familiarity with the animals on a scale of 1 (not at all familiar) to 7 (very familiar), as well as typing the name of animals if they knew them. For the purposes of this paper, we will present data from fMRI sessions (I and III) only.

2.3. Stimuli

2.3.1. Images

Twenty-six unique images (25 for the encoding and retrieval tasks, one for cued recall) of 40 animals were administered. Each image was cropped to the center of a white background of 252×252 pixels with a resolution of 72 pixels/inch. Images of each animal species (e.g., shoe-bill) presented different individual animals (i.e., distinct individuals) from different viewpoints (see Fig. 1). Animals were selected based on the low likelihood that participants were familiar with their appearance (established through norming from prior studies; Coutanche and Koch, 2018; Koch et al., 2020). Participants' prior (un)familiarity with the species was verified through the final survey, in which participants were asked to provide the English name for each species (if known). The animals consisted of ten birds, ten fish, six insects, twelve mammals, and two reptiles.

2.3.2. Words

The word materials were 40 Dutch words, three to eight letters in length ($M = 5.78$, $SD = 1.24$), based on the normalized dataset from Tokowicz et al. (2002). All words were concrete nouns and had one English translation. Form similarity between Dutch words and their English translation was low ($M = 1.40$, $SD = 0.32$; range = 1–2.38) on a scale of 1 ("low similarity") to 7 ("high similarity"), as rated by 24 Dutch-English bilinguals.

2.4. Tasks

The trial order within all tasks was pseudorandomized in advance and made to follow the order determined by the jitter optimization program, Optseq 2 (Greve, 2002). The set of words included in Set 1 and Set 2 were counterbalanced from Sets 1 and 2 across participants. Each fMRI run began with 2 TRs of fixation and ended with 6 TRs (localizer) or 5 TRs (encoding, retrieval) of fixation.

2.4.1. Localizer

We functionally localized the VWFA with a words > non-words

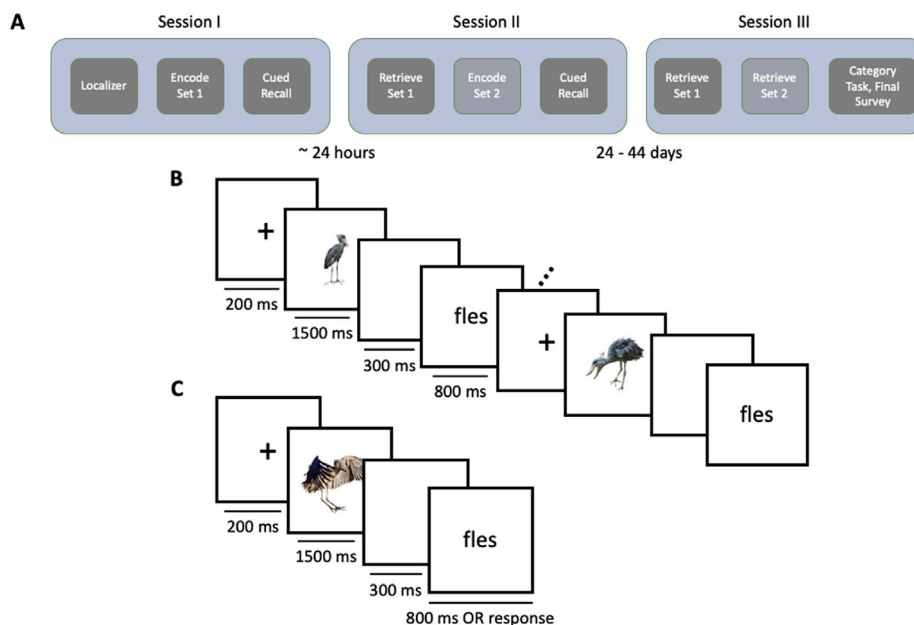


Fig. 1. Task layout and example trials for a given word-animal pair. A) Overview of experimental sessions. Set 2 stimuli (faded) were encoded during EEG and not analyzed here. B) Encoding showing two trials for a pair. Three dots indicate other intervening trials for other pairs. C) Retrieval trial where the correct answer is “match.”

contrast. During the localizer task, participants were presented with a letter string and prompted to indicate, by pressing one of three fingers, whether the letter string was a non-English (e.g., vlees) or English word, whether it was larger (e.g., elephant) or smaller (e.g., key) than a shoebox. This task was used to drive semantic processing of the presented words. Real word stimuli referred to either animals (animate) or inanimate objects. The localizer consisted of one run of 80 trials and was preceded by six practice trials.

2.4.2. Encoding

Participants were told they would be learning Dutch labels of various animals. Prior to the session, two sets of associations were created, each consisting of 20 unique Dutch word-animal pairings, the concepts learned in our study (counterbalanced across participants). Each participant was presented with associations from one of the counterbalanced sets. Participants passively viewed images of animals (described above) and Dutch labels over the course of two functional runs in the scanner. Each trial consisted of 200 ms of fixation, 1500 ms of the animal image, an interstimulus interval (ISI) of 300 ms, and 800 ms of the Dutch label, centrally located on the screen. All animals were presented once per iteration before there was a repeat, with at least four trials separating any repeats (at the end and start of the iterations).

2.4.3. Retrieval

Participants viewed trials in which images of animals were followed by a Dutch label. Participants responded as to whether or not the animals matched the label (based on the prior encoding). All mismatch pairings were consistent for each presentation (e.g., if the shoebill was labeled as “citreon” during encoding, its mismatch word might be “kicker”, but no other label) to ensure that participants could not use a continually changing foil as a cue (i.e., select the word that does not change). Trials began with 200 ms of fixation and an image presentation for 1500 ms. After a 300 ms ISI, the Dutch label appeared, and participants had 800 ms to respond. Trials ended when participants responded or once 800 ms had passed. To ensure that participants were comfortable with the task, the first retrieval task in the session was preceded by 10 practice trials. Forty match and 40 mismatch trials appeared during each run. All animals were presented once per iteration before there was a repeat, with at least four trials separating any repeats (at the end and start of the iterations).

2.4.4. Cued recall

After leaving the scanner, participants first viewed animal images that they had viewed previously during encoding. After each image disappeared, participants typed the name of the encoded Dutch label and pressed enter to start the next trial.

2.4.5. Categorization

As an additional test of participants' encoding of the Dutch labels and animals, a categorization task on Session III presented triads of animal names. Each triad consisted of two already familiar animals (e.g., grasshopper and snake), along with a word encoded during either Session I or Session II (e.g., fles). Two of the animals came from the same taxonomic group (fish, bird, reptile, mammal, and insect) with the third animal from a different group. Participants were told to indicate which animal did not match the taxonomic category of the other two. Each Dutch label was shown three times in different triads. The new Dutch label was the odd one out in one third of the trials. To establish the words in the triads, two sets were generated, randomly pairing existing animals within their taxonomic category to create 120 triads each. Existing animals were taken from the following number of animal categories: birds = 51, reptiles = 18, mammals = 63, fish = 30, and insects = 26. These existing animals were not shown more than twice per participant. Results from this test were not analyzed and will not be discussed further.

2.4.6. Final survey

At the end of the final session, participants provided the English name of each of the animals if they knew it, wrote if they had a strategy for remembering the animal names, and whether they rehearsed the animals during the delay between the three sessions.

2.5. Data acquisition

Participants were scanned using a Siemens 3-T head only Allegra magnet and standard radio-frequency coil equipped with mirror device to allow for fMRI stimuli presentation. Whole-brain imaging was conducted. T1-weighted images were acquired at the start of both sessions (TR = 1.540 s, TE = 3.04 s, voxel size = 1.0 × 1.0 × 1.0 mm). Localizer data collection took place over one functional run (TR = 2.25 s, TE = 25), encoding over 2 functional runs (TR = 2.8 s, TE = 25), and retrieval over 3 functional runs (TR = 2.8 s, TE = 25). All functional runs employed voxel sizes of 3.125 × 3.125 × 3.125 mm. A predetermined jitter was used in all functional runs. The optimal sequence was determined using Optseq 2 (<http://surfer.nmr.mgh.harvard.edu/optseq/>) and an average jitter length equivalent to the run's TR length.

2.6. Data preprocessing

Anatomical images from Session III were registered to those collected in Session I. Imaging data were preprocessed using the Analysis of Functional NeuroImages (AFNI) software package (Cox, 1996). Slice-time and motion corrections were applied to all functional images so as to register them to a mean functional volume. A high-pass filter was used to remove low-frequency trends below 0.01 Hz from all runs. For multivariate analyses, betas were calculated with this data using the Least Squares-Separate (LSS) method in a superordinate-wise and item-wise fashion based on the onset time of the image appearing on the screen (Mumford et al., 2012). Only match trials were included when calculating betas during retrieval to ensure trials were as similar as possible to those with which they were compared during encoding. All encoding trials were modeled with epochs of 2.6 s. Retrieval encoding trials were modeled with a variable epoch model to account for the different trial lengths (based on when the subjects responded). The option `stimtimes_am1` was used with AFNI's `3ddeconvolve`. If the subject did not respond on a trial, the maximum possible trial length (2.6 s) was included in the model. This produced a vector of beta coefficients for each ROI reflecting the BOLD response to the superordinate category or item, which was then subjected to multivariate analyses (described below). The functional data was not smoothed prior to the multivariate analyses. Classification analyses (see Supplementary Materials) used the same preprocessing steps as the other multivariate analyses. For univariate analyses, data were smoothed using a kernel with full width at half maximum (FWHM) of 6 mm and regions were standardized to Talairach space for group analyses (see Supplementary Materials).

2.7. Regions of interest

We identified 16 regions of interest (ROIs) for our analyses. To account for the number of ROIs included in our study, we divided our regions into primary a priori primary ROIs, and exploratory secondary ROIs in which neural analyses were corrected for multiple comparisons by accounting for false discovery rate (FDR) using the Benjamini-Hochberg (BH) Procedure (Benjamini and Hochberg, 1995). All neural data reported for secondary ROIs are BH-corrected *p*-values. We used a systematic approach to decide which ROIs would be primary or secondary. We first identified papers in the literature that tested encoded associations and found either pattern robustness or ERS that predicted subsequent memory performance. Our primary ROIs of interest were those with significant results in these past studies: bilateral VT, vPFC, PHC, and HC (Kuhl et al., 2012; Staresina et al., 2012; Wing et al., 2015; see Fig. 2). In addition, because many of the studies found significant effects

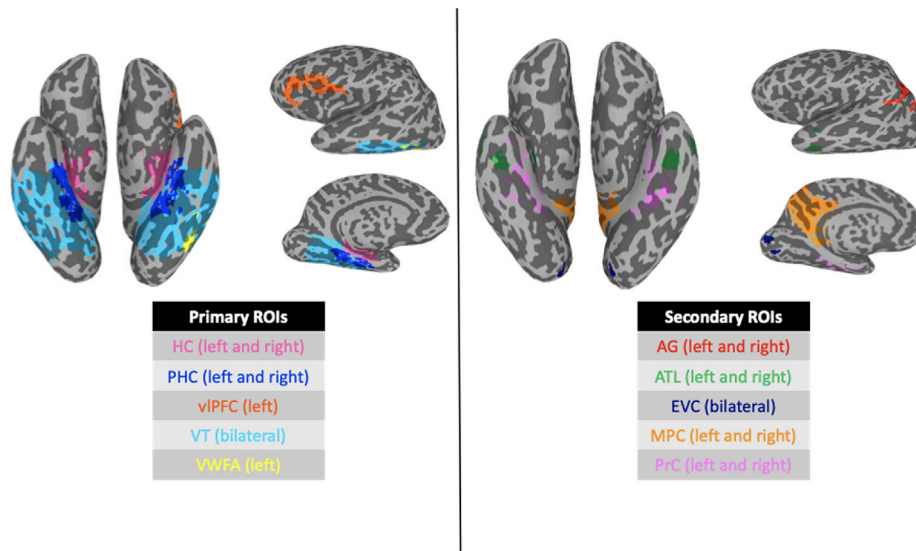


Fig. 2. Visualization of primary and secondary ROIs from a representative subject. The displayed VT ROI shows voxels from which the feature selection is conducted (123 most animacy-responsive voxels; see Methods and Materials).

in stimulus-sensitive cortex (Jonker et al., 2018; Koen et al., 2019; Ward et al., 2013), we also included the VWFA, which is sensitive to word stimuli. Our secondary ROIs were selected for exploratory analyses. The PrC and AG were included to test for memory effects (Davachi et al., 2003; Uncapher and Wagner, 2009), the ATL and EVC to test for superordinate and item level effects (Devlin et al., 2002; Eger et al., 2008; Pobric et al., 2010), and the MPC because it is related to long-term retrieval (Bird et al., 2015). All ROIs were created in the right and left hemisphere separately, except for the VWFA, which is left-lateralized (McCandliss et al., 2003), the vPFC, which is left-lateralized for semantic retrieval (Badre and Wagner, 2007), and the VT and EVC, which were examined bilaterally.

The HC, PHC, and MPC were created using FreeSurfer (Fischl et al., 2002). The MPC consists of the precuneus, subparietal sulcus, and posterior cingulate gyrus. The vPFC ROI was created using BA44, BA45, and BA47 from AFNI's TT_Daemon atlas. The VT ROI was created by selecting animate-sensitive voxels by choosing the 123 most-active voxels VT (in either right or left hemisphere) for animate relative to inanimate words in the functional localizer. EVC was defined by placing a 10 mm-radius sphere on the calcarine sulcus. The ATL was defined using coordinates reported in a previous study (10 mm-radius sphere at 41, 8, -17; -41, 8, -17; Coutanche and Thompson-Schill, 2015). Although the localizer was created to define the VWFA, it was not able to produce reliable word-form sensitivity in most participants. As such, we defined the region according to coordinates reported in a past study (10 mm-radius sphere at -42, -56, -12; Price and Devlin, 2003).

2.8. Behavioral

To evaluate behavioral performance on the cued recall task in a way that accounts for the accuracy of typed responses (e.g., “citeon” is very similar to the correct answer of “citeon”), we calculated the lexical similarity between the word typed by the participant and the correct word for the image shown, using an online resource of orthographic similarity (van Orden, 1987; <https://psico.fcep.urv.cat/utilitats/nim/eng/graphsim.php>). Performance on the recognition test was measured by calculating d' for each concept to account for both hits and false alarms.

2.9. Pattern robustness

Pattern robustness was calculated between conditions during

encoding. At the superordinate level, item-level beta coefficients were first averaged at this level for each run (e.g., for birds, all bird exemplar patterns were averaged for run 1 and again for run 2). Pattern robustness was measured as the difference between “within” and “between” similarities. Within similarity was calculated for each superordinate level as the correlation between matching superordinate level patterns from run 1 and run 2 (e.g., the bird pattern would be correlated between runs). Values were Fisher- z corrected and averaged across all superordinate levels. Between similarity was calculated in the same manner but across groups (e.g., for birds, all bird patterns from one run with patterns for other taxonomic groups in the other run). Trials were correlated across (rather than within) runs to prevent any spurious correlations that might occur by comparing trials in the same run (Mumford et al., 2014). Within and between similarities were the Fisher- z corrected Pearson's correlation coefficients from these analyses (Kuhl et al., 2012; LaRocque et al., 2013). Pattern robustness was calculated for each trial, as the difference of within and between similarities. For group-level analyses, the superordinate level analyses were averaged and tested against zero in a one-sample t -test.

Item-level analyses were similar to the superordinate level analyses, but with item-level beta coefficients (e.g., all shoebills, a type of bird) averaged for each run and correlated between runs to calculate within (e.g., shoebill - shoebill) and between (e.g., shoebill - other types of bird) similarity. The between similarity was always calculated within the same superordinate level (e.g., shoebill - other types of bird but not with mammals) to remove any influence of taxonomic categories on this measure.

2.10. Encoding-retrieval similarity

ERS was calculated at a superordinate level and item-level using superordinate-level and item-level beta coefficients, respectively. For superordinate analyses, the superordinate beta vector from encoding was correlated with the superordinate vector from retrieval, Fisher- z corrected, and then averaged. The correlated superordinate levels were either identical (“within”) or different (“between”). Item-level analyses correlated the item-level beta vectors during encoding and retrieval for the same concepts (within; e.g., shoebill with shoebill) or different concepts (between; e.g., shoebill with other birds). The correlations were Fisher- z corrected and averaged for within or between. The “between” similarity was subtracted from the within similarity to calculate the final ERS difference scores for each superordinate level, which were compared

against zero in a group-wise one-sample *t*-test.

2.11. Behavioral regressions

To test whether the neural metrics related to subject's behavioral performance, all regions with significant neural findings for a given measure (i.e., pattern robustness, ERS) were entered into two linear mixed effects models, each predicting a measure of behavioral performance: cued recall performance on Session I and recognition performance on Session III. The neural measure was the predictor of interest in the main model (pattern robustness or ERS), with the behavioral measure serving as the dependent variable. Subject was included as a random effect. Because pre-experimental familiarity has been shown to be an important indicator of memory performance for word-image pairs (Bruett et al., 2018), pre-experimental familiarity was included as a fixed effect in all models. Change in AIC values were calculated to determine statistical significance, where the null model contained all variables except for the neural measure. This method produced results consistent with those calculated when all variables were placed in one model and significance was determined by examining the *p*-value associated with the variable of interest. Plots of significant and marginally significant regression models can be found in Supplementary Materials.

3. Results

3.1. Behavioral

Two participants were removed from analyses due to below-chance behavioral performance on the Session II retrieval task (<50% recognition accuracy for Set 1 words), leaving 18 subjects for the final analyses. Participants' performance showed that they learned the word-animal pairings, as reflected in the cued recall task in Session I ($M = 0.56$, $SD = 0.43$), the recognition task in Session II ($d' M = 1.9$, $SD = 2.13$), and the recognition task in Session III ($d' M = 1.5$, $SD = 2.43$). Participants were generally unfamiliar with our animal stimuli prior to the study (on a 1 (not at all familiar) to 7 (very familiar) scale, $M = 2.5$, $SD = 1.95$). The responses across all subjects covered the full range (1 through 7), justifying our use of this measure in our regressions.

3.2. Pattern robustness

Pattern robustness is reflected in the overlap (similarity) of encoding presentations at the item or superordinate level. Superordinate-level pattern robustness for the primary ROIs are shown in Fig. 3. Significant superordinate-level pattern robustness was present in the left HC ($M = 0.082$, $t(17) = 2.91$, $p = .010$), left PHC ($M = 0.062$, $t(17) = 2.13$, $p = .048$), right PHC ($M = 0.086$, $t(17) = 2.92$, $p = .009$), vIPFC ($M = 0.135$, $t(17) = 5.18$, $p < .001$), with strongest robustness in stimulus-specific ROIs, the VWFA ($M = 0.346$, $t(17) = 9.62$, $p < .001$) and VT cortex ($M = 0.223$, $t(17) = 8.91$, $p < .001$). The right HC did not reach significance ($p = .086$). Of the secondary ROIs, the left AG ($M = 0.208$, $t(17) = 8.61$, $p < .001$), right AG ($M = 0.176$, $t(17) = 7.84$, $p < .001$), EVC ($M = 0.360$, $t(17) = 9.30$, $p < .001$), left MPC ($M = 0.152$, $t(17) = 4.74$, $p < .001$), right MPC ($M = 0.168$, $t(17) = 6.76$, $p < .001$), left PrC ($M = 0.111$, $t(17) = 4.14$, $p = .002$), and right PrC ($M = 0.097$, $t(17) = 2.72$, $p = .037$) showed significant superordinate-level robustness after correcting for FDR. No other secondary ROIs were significant after correction (see Supplementary Materials for full table).

In testing the relationship between pattern robustness and behavioral performance (cued recall and recognition memory accuracy), only regions with significant pattern robustness were examined. Of these regions, Session III recognition memory performance was predicted by superordinate-level pattern robustness in left vIPFC ($\beta = 2.13$, $SE = 0.99$, $\Delta AIC = 2.68$, $\chi^2 = 4.68$, $p = 0.031$), left MPC ($\beta = 2.47$, $SE = 0.92$, $\Delta AIC = 5.15$, $\chi^2 = 7.16$, $p = 0.007$), right PrC ($\beta = 1.95$, $SE = 0.83$, $\Delta AIC = 3.15$, $\chi^2 = 5.16$, $p = 0.023$), and EVC ($\beta = 1.36$, $SE = 0.66$, $\Delta AIC =$

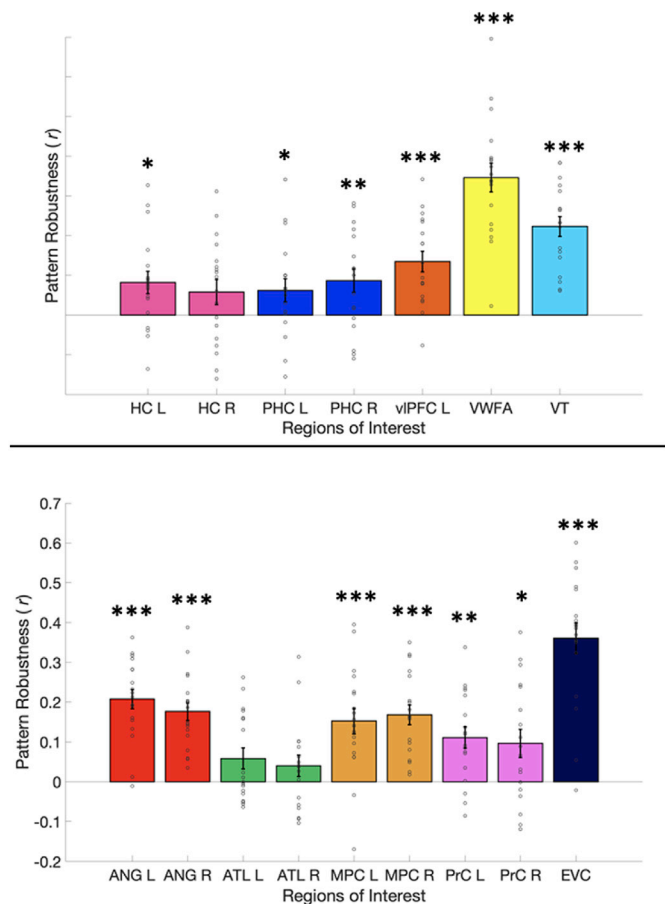


Fig. 3. Pattern robustness for superordinate analyses for primary (above) and secondary (below) ROIs. Pearson's *r* standardized with Fisher's *z*. Error bars represent standard errors of the mean. Circles represent means of the 18 participants. Right and left hemispheres indicated by "R" and "L", respectively. * $p < .05$, ** $p < .01$, *** $p < .001$.

2.32 , $\chi^2 = 4.32$, $p = 0.038$). Recognition memory was marginally predicted by pattern robustness in the left PHC ($\beta = 1.13$, $SE = 0.65$, $\Delta AIC = 1.09$, $\chi^2 = 3.09$, $p = 0.079$). Session I cued recall performance was marginally predicted by pattern robustness in the right AG ($\beta = 0.23$, $SE = 0.12$, $\Delta AIC = 1.67$, $\chi^2 = 3.67$, $p = 0.055$).

Because we saw superordinate-level pattern robustness effects, we tested whether we could detect effects at an even finer level of granularity by analyzing item-level robustness, with the caveat that significantly fewer trials contributed to this analysis. Item-level robustness was significant in the left PHC ($M = -0.041$, $t(17) = -2.40$, $p = .028$), but not in other primary regions ($ps > .353$) or secondary regions ($ps > .405$ after FDR correction). Left PHC item-level pattern robustness was not significantly related to Session I cued recall or Session III recognition performance ($ps > .257$).

3.3. Encoding-retrieval similarity

ERS is reflected in pattern overlap (similarity) between encoding and retrieval. Superordinate-level ERS was significant in the left PHC ($M = -0.41$, $t(17) = -2.42$, $p = .027$) and vIPFC ($M = 0.018$, $t(17) = 2.61$, $p = .018$), but no other primary ROIs ($ps > .154$; see Fig. 4). No secondary ROIs had significant superordinate-level ERS after FDR-correction.

ROIs with significant superordinate-level ERS were tested as potential predictors of Session I cued recall performance and Session III recognition memory. Superordinate-level ERS in the left PHC was marginally predictive of Session III recognition memory ($\beta = 1.88$, $SE = 0.50$, $\Delta AIC = 1.42$, $\chi^2 = 3.42$, $p = 0.064$). ERS in other regions did not significantly

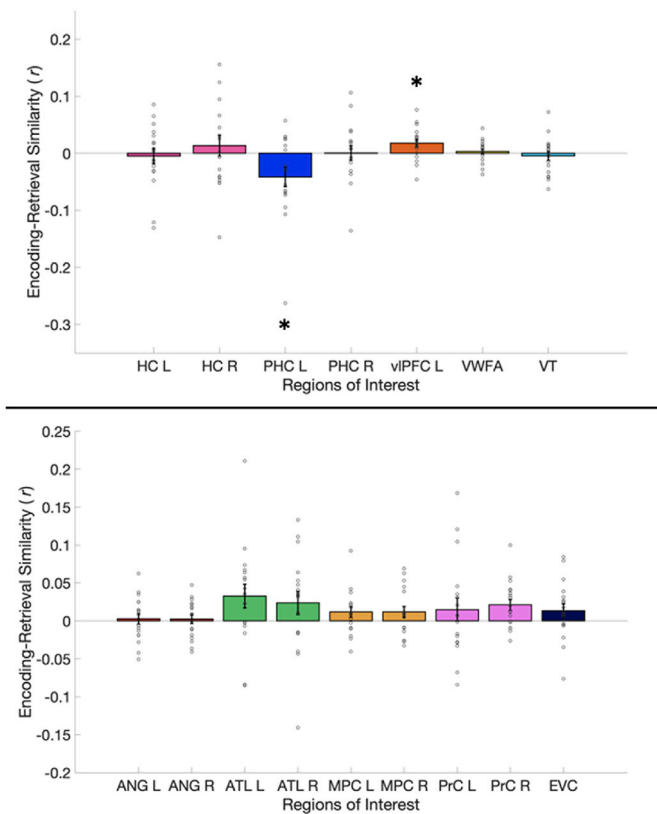


Fig. 4. ERS for superordinate analyses for primary (above) and secondary (below) ROIs. Pearson's r standardized with Fisher's z . Error bars represent standard errors of the mean. Right and left hemispheres are indicated by "R" and "L", respectively. * $p < .05$, ** $p < .01$, *** $p < .001$.

predict recognition memory ($ps > .978$) or cued recall performance ($ps > .117$). Item-level ERS did not reach significance in any primary regions ($ps > .104$) or secondary regions.

4. Discussion

In this study, we examined how two neural measures – pattern robustness and ERS – predict memory for concepts across visually-varied exemplars over an extended delay. By pairing each Dutch word with multiple images of a novel animal species, we ensured that pattern similarity effects were not driven by associations with low-level visual similarities between instances of a single image. Superordinate-level representations were examined through pattern robustness during encoding and ERS.

To first understand how patterns at encoding predicted memory performance, we measured pattern robustness across exemplars. A number of regions showed significant pattern robustness that was consistent across multiple exemplars (for each species), and across species of the same taxonomic category. Importantly, pattern robustness in many of these regions predicted memory performance after a month-long delay, including the vIPFC, and two regions of medial temporal cortex - right PrC and left PHC. These findings are consistent with past studies showing that consistent encoding representations in these regions facilitates associative memory formation (LaRocque et al., 2013; Wagner et al., 2016; Xue et al., 2010). Notably, pattern robustness in both the right and left MPC also predicted subsequent memory, in line with the MPC's involvement in long-term retrieval (Bird et al., 2015; Lee et al., 2018). Pattern robustness in EVC was also predictive of subsequent memory. Past research has suggested that high-level visual regions responsible for object categorization have long-range feedback loops to early visual cortex. One study, for instance, found higher activity in EVC

associated with category-level target detection despite the need for upstream semantic systems to define these targets, providing evidence for these connections (Hon et al., 2009). Given that our effect was present at the superordinate level, it is likely not representative of bottom-up early visual processing of concepts (as otherwise it would also be observed at the item-level), but instead likely reflects feedback mechanisms from higher level regions (Hon et al., 2009; Luck et al., 1997) which is known to affect multi-voxel patterns, as shown by the ability to decode visual properties from associated words in EVC (Borghesani et al., 2016). Finally, the right AG was the only region to marginally predict memory on the day of the scan. This aligns with past findings of univariate negative subsequent memory effects in the AG, where forgotten words tend to show more activity than subsequently remembered words (Clark and Wagner, 2003; Park et al., 2013), and pattern robustness measures predict subsequent memory after 3 days (Wagner et al., 2016).

In addition to encoding, we were also interested in how pattern similarity between encoding and retrieval predicted subsequent memory for concepts, measured through ERS across varying exemplars. Two regions, the left PHC and vIPFC, showed evidence of superordinate-level ERS, and none showed evidence at the item level. Interestingly, the left PHC, associated with both memory (Brewer et al., 1998; LaRocque et al., 2013; Xiao et al., 2016) and high-level visual processing (Bonner et al., 2016), showed superordinate-level ERS that was marginally related to recognition memory one month after encoding. This is consistent with other studies that have found the region to be particularly sensitive to processing superordinate-level information, relative to other regions in medial temporal cortex (Diana et al., 2008). Unexpectedly, the directionality of the PHC findings indicate that more negative ERS scores (i.e., between $>$ within similarity) predicted improved memory performance. Although this was not an anticipated direction, past studies have shown that the directionality of pattern similarity (albeit for robustness) in this region seems to depend on the stimulus viewed (LaRocque et al., 2013). In particular, a past study showed negative directionality for living things (i.e., faces, bodies) and positive directionality for non-living things (i.e., objects, scenes; LaRocque et al., 2013). Our negative direction for novel animals is consistent with this; however, given the lack of similar findings in other related studies, we note that caution should be taken in interpreting them at present. Future studies will want to investigate the role of the left PHC in ERS more directly to understand how these patterns benefit memory, and whether directionality interacts with the learned concept or stimulus.

These findings are particularly interesting due to their contribution to understanding the processing of concepts at multiple levels of abstraction or semantic granularity. Concepts can be considered at multiple levels of semantic granularity. One can think of a mammal, a dog, or their friend's pet, Gracie. Some levels might be forgotten over time, however. For example, one might remember that their friend has a dog, but not remember its name or its breed. Although it is clear that these levels exist, their relationship in the brain remains to be fully understood.

Our findings begin to speak to the issue of semantic granularity by demonstrating that superordinate level neural information is associated with item-level memory. One possible reason for this is that different levels of granularity may not be independent. This is consistent with a recent study in which participants learned to associate pseudowords with one of six categories of words (monkeys, donkeys, elephants, hammers, wrenches, screwdrivers; Malone et al., 2016). After learning these relationships, the representations of the pseudowords reflected the semantic relationships that are expected of the six categories, both at the basic and superordinate level. Importantly, during recognition, overlapping representations formed during encoding can be used to judge the familiarity of test cues, where a higher match leads to better memory (Hintzman, 1984). Given past findings that representations of concepts at different levels of granularity are not independent, one explanation for our findings may be that robust item- and category-level neural representations were created during encoding. As such, at retrieval, those items that had the most robust stored representations showed a higher

match with the cue's representation and were best remembered. Examining multiple levels of granularity using different neuroimaging analysis methods (i.e., univariate and multivariate) will likely be fruitful in understanding these levels better in the future.

It is surprising and worth noting that our results did not show evidence of ERS in the HC, given that this effect has been previously observed (e.g., Liang and Preston, 2017; Mack and Preston, 2016). A possible reason for this unexpected finding is the significant delay between encoding and retrieval. Although the HC is important for memory encoding and retrieval, new information from encoding is consolidated over time with sleep (Takashima et al., 2009). As a result, although HC ERS may be important on a shorter time scale, before much consolidation has occurred, our results suggest these patterns may change over longer time scales. In addition, it has also been suggested that, relative to the other regions of medial temporal cortex, there may be less superordinate sensitivity to stimuli in the HC (LaRocque et al., 2013). Finally, alternative imaging sequences may provide additional insights into HC, though often at the cost of reduced coverage. As we planned to collect whole-brain analyses, we did not collect high resolution fMRI data, and thus did not optimize in accordance with these constraints, which could also explain our lack of HC findings.

Our study examined item-level pattern robustness and ERS to further our investigation regarding semantic granularity. Our findings showed evidence of pattern robustness at the item-level in the left PHC, consistent with our finding at the superordinate level. This finding, though preliminary, suggests that the encoding representations in PHC may be important for multiple levels of semantic granularity, though future studies will be needed before conclusions can be drawn. A limitation of analysis at the item-level in this study was that each word-animal pairing was only presented eight times during encoding (four presentations in each of the two runs). Because the item-level analyses involved creating average patterns in each run, only four trials contributed to the average for each concept, so we were likely underpowered to detect item-level effects in regions with a smaller effect size or large variance. For this reason, it is not possible to directly compare item and superordinate levels. We anticipate that a study featuring more presentations of concepts might find stronger effects for pattern robustness and ERS, with relationships to subsequent memory. Finally, we note that our study includes a sample of 18 subjects, which may limit our findings. Although our sample size is consistent with other fMRI studies examining similar processing (e.g., LaRocque et al., 2013; Ritchey et al., 2013), it is possible that greater power may allow greater sensitivity in detecting item-level differences.

The timing of our experimental trials was such that words and images were presented close in time to one another. Because of this, one concern might be whether hemodynamic overlap affects the interpretation of our findings. For instance, the question arises whether the neural findings reported reflect an abstraction of regularities within a category or repeated word forms? We acknowledge that this question is valid; however, we note that overlap at the superordinate level is relatively small because each level contains multiple species, each with a different word string. Additionally, neural patterns in certain regions were able to predict subsequent recognition performance, but patterns in the VWFA (which are most likely to represent word form) could not. We argue that this is consistent with the interpretation that the reported neural patterns do not simply represent word-form information, but instead reflect higher-order semantic information, which in turn is associated with memory outcomes. Finally, past findings demonstrate that some of these regions are not associated with word processing. For example, in one study, authors did not find activation differences in the left MPC when retrieving words versus faces (Guerin and Miller, 2009). In addition, the PrC has been shown to represent semantic distances between words in a way that is not perceptually driven (Bruffaerts et al., 2013). These findings support the claim that the effects we show here are at least not driven entirely by word form.

This study provides evidence for how new concepts (with visually-

varying exemplars) are encoded and later retrieved on long time scales. Our findings highlight the predictive relationship between superordinate pattern robustness and conceptual memories after a month, and suggest that parietal and medial temporal cortical regions may be particularly important in this relationship.

CRediT authorship contribution statement

Heather Bruett: Conceptualization, Formal analysis, Data curation, Writing - original draft. **Regina C. Calloway:** Conceptualization, Data curation, Writing - review & editing. **Natasha Tokowicz:** Conceptualization, Funding acquisition, Writing - review & editing, Supervision. **Marc N. Coutanche:** Conceptualization, Funding acquisition, Writing - review & editing, Supervision.

Acknowledgements

This work was supported by the Learning Research and Development Center Internal Awards Program. The authors thank Griffin Koch for feedback on previous drafts. Angeleah Carreras, Allyson Distler, Shansa Gaddam, Jasmine Issa, Ankitha Iyer, Claire Kollhoff, John Paulus and Aarya Wadke assisted with stimuli and data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2020.117030>.

References

- Badre, D., Wagner, A.D., 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45 (13), 2883–2901.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* 57 (1), 289–300. JSTOR, 14426–14434. <https://doi.org/10.1523/JNEUROSCI.1774-15.2015>.
- Bird, C.M., Keidel, J.L., Ing, L.P., Horner, A.J., Burgess, N., 2015. Consolidation of complex events via reinstatement in posterior cingulate cortex. *J. Neurosci.* 35 (43), 14426–14434. <https://doi.org/10.1523/JNEUROSCI.1774-15.2015>.
- Bonner, M.F., Price, A.R., Peelle, J.E., Grossman, M., 2016. Semantics of the visual environment encoded in parahippocampal cortex. *J. Cognit. Neurosci.* 28 (3), 361–378. https://doi.org/10.1162/jocn_a_00908.
- Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E., Piazza, M., 2016. Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *Neuroimage* 143, 128–140. <https://doi.org/10.1016/j.neuroimage.2016.08.068>.
- Brewer, J., Zhao, Z.H., Gabrieli, J.D.E., 1998. Parahippocampal and frontal responses to single events predict whether those events are remembered or forgotten. *Science* 281, 1185–1187.
- Bruett, H., Fang, X., Kamaraj, D.C., Haley, E., Coutanche, M.N., 2018. Expertise moderates incidentally learned associations between words and images. *Front. Psychol.* 9, 2085.
- Bruffaerts, R., Dupont, P., Peeters, R., De Deyne, S., Storms, G., Vandenberghe, R., 2013. Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. *J. Neurosci.* 33 (47), 18597–18607.
- Carlos, B.J., Hirshorn, E.A., Durisko, C., Fiez, J.A., Coutanche, M.N., 2019. Word inversion sensitivity as a marker of visual word form area lateralization: an application of a novel multivariate measure of laterality. *Neuroimage* 191, 493–502.
- Clark, D., Wagner, A.D., 2003. Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia* 41 (3), 304–317. [https://doi.org/10.1016/S0028-3932\(02\)00163-X](https://doi.org/10.1016/S0028-3932(02)00163-X).
- Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. *J. Neurosci.* 32 (8), 2608–2618. <https://doi.org/10.1523/JNEUROSCI.5547-11.2012>.
- Coutanche, M.N., 2013. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cognit. Affect. Behav. Neurosci.* 13 (3), 667–673. <https://doi.org/10.3758/s13415-013-0186-2>.
- Coutanche, M.N., Koch, G.E., 2018. Creatures great and small: real-world size of animals predicts visual cortex representations beyond taxonomic category. *Neuroimage* 183, 627–634. <https://doi.org/10.1016/j.neuroimage.2018.08.066>.
- Coutanche, M.N., Solomon, S.H., Thompson-Schill, S.L., 2016. A meta-analysis of fMRI decoding: quantifying influences on human visual population codes. *Neuropsychologia* 82, 134–141. <https://doi.org/10.1016/j.neuropsychologia.2016.01.018>.
- Coutanche, M.N., Thompson-Schill, S.L., 2012. Reversal without remapping. *Perspect. Psychol. Sci.* 7 (2), 118–134. <https://doi.org/10.1177/1745691611434211>.

- Coutanche, M.N., Thompson-Schill, S.L., 2015. Creating concepts from converging features in human cortex. *Cerebr. Cortex* 25 (9), 2584–2593. <https://doi.org/10.1093/cercor/bhu057>.
- Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res. Int. J.* 29 (3), 162–173.
- Davachi, L., Mitchell, J.P., Wagner, A.D., 2003. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc. Natl. Acad. Sci. Unit. States Am.* 100 (4), 2157–2162.
- Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Moss, H.E., Fadili, M.J., Tyler, L.K., 2002. Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia* 40 (1), 54–75. [https://doi.org/10.1016/S0028-3932\(01\)00066-5](https://doi.org/10.1016/S0028-3932(01)00066-5).
- Diana, R.A., Yonelinas, A.P., Ranganath, C., 2008. High-resolution multi-voxel pattern analysis of category selectivity in the medial temporal lobes. *Hippocampus* 18 (6), 536–541. <https://doi.org/10.1002/hipo.20433>.
- Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R.J., Rees, G., 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. *J. Cognit. Neurosci.* 20 (2), 356–370. <https://doi.org/10.1162/jocn.2008.20019>.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33 (3), 341–355.
- Greve, D.N., 2002. Optseq2. Computer software. <http://Surfer.Nmr.Mgh.Harvard.Edu/Optseq>.
- Guerin, S.A., Miller, M.B., 2009. Lateralization of the parietal old/new effect: an event-related fMRI study comparing recognition memory for words and faces. *Neuroimage* 44 (1), 232–242.
- Hintzman, D.L., 1984. MINERVA 2: a simulation model of human memory. *Behav. Res. Methods Instrum. Comput.* 16 (2), 96–101.
- Hon, N., Thompson, R., Sigala, N., Duncan, J., 2009. Evidence for long-range feedback in target detection: the reaction of semantic targets modulates activity in early visual areas. *Neuropsychologia* 47 (7), 1721–1727. <https://doi.org/10.1016/j.neuropsychologia.2009.02.011>.
- Hutchinson, J.B., Uncapher, M.R., Wagner, A.D., 2009. Posterior parietal cortex and episodic retrieval: convergent and divergent effects of attention and memory. *Learn. Mem.* 16 (6), 343–356.
- Jonker, T.R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., Ranganath, C., 2018. Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proc. Natl. Acad. Sci. Unit. States Am.* 115 (43), 11084–11089. <https://doi.org/10.1073/pnas.1800061115>.
- Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. *Nature* 452 (7185), 352–355. <https://doi.org/10.1038/nature06713>.
- Koch, G.E., Akpan, E., Coutanche, M.N., 2019. Image Memorability is Predicted by Activity across Different Stages of Convolutional Neural Networks and the Human Ventral Stream. *BioRxiv*, p. 834796.
- Koch, G.E., Paulus, J.P., Coutanche, M.N., 2020. Neural patterns are more similar across individuals during successful memory encoding than during failed memory encoding. *Cerebr. Cortex*. <https://doi.org/10.1093/cercor/bhaa003> (New York, N.Y.: 1991).
- Koen, J.D., Hauck, N., Rugg, M.D., 2019. The relationship between age, neural differentiation, and memory performance. *J. Neurosci.* 39 (1), 149–162. <https://doi.org/10.1523/JNEUROSCI.1498-18.2018>.
- Kuhl, B.A., Chun, M.M., 2014. Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *J. Neurosci.* 34 (23), 8051–8060. <https://doi.org/10.1523/JNEUROSCI.4328-13.2014>.
- Kuhl, B.A., Brice, A., Rissman, J., Wagner, A.D., 2012. Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia* 50 (4), 458–469. <https://doi.org/10.1016/j.neuropsychologia.2011.09.002>.
- Lambon Ralph, M.A., 2014. Neurocognitive insights on conceptual knowledge and its breakdown. *Phil. Trans. Biol. Sci.* 369 (1634), 20120392. <https://doi.org/10.1098/rstb.2012.0392>.
- Lambon Ralph, M.A., Jefferies, E., Patterson, K., Rogers, T.T., 2017. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18 (1), 42–55. <https://doi.org/10.1038/nrn.2016.150>.
- LaRocque, K.F., Smith, M.E., Carr, V.A., Wittthoff, N., Grill-Spector, K., Wagner, A.D., 2013. Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J. Neurosci.* 33 (13), 5466–5474.
- Lee, H., Samide, R., Richter, F.R., Kuhl, B.A., 2018. Decomposing parietal memory reactivation to predict consequences of remembering. *Cerebr. Cortex* 29 (8), 3305–3318.
- Liang, J.C., Preston, A.R., 2017. Medial temporal lobe reinstatement of content-specific details predicts source memory. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 91, 67–78. <https://doi.org/10.1016/j.cortex.2016.09.011>.
- Luck, S.J., Chelazzi, L., Hillyard, S.A., Desimone, R., 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77 (1), 24–42. <https://doi.org/10.1152/jn.1997.77.1.24>.
- Mack, M.L., Preston, A.R., 2016. Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. *Neuroimage* 127, 144–157.
- Malone, P.S., Glezer, L.S., Kim, J., Jiang, X., Riesenhuber, M., 2016. Multivariate pattern analysis reveals category-related organization of semantic representations in anterior temporal cortex. *J. Neurosci.* 36 (39), 10089–10096.
- McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cognit. Sci.* 7 (7), 293–299.
- Mumford, J.A., Davis, T., Poldrack, R.A., 2014. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *Neuroimage* 103, 130–138. <https://doi.org/10.1016/j.neuroimage.2014.09.026>.
- Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* 59 (3), 2636–2643.
- Park, H., Kennedy, K.M., Rodrigue, K.M., Hebrank, A., Park, D.C., 2013. An fMRI study of episodic encoding across the lifespan: changes in subsequent memory effects are evident by middle-age. *Neuropsychologia* 51 (3), 448–456. <https://doi.org/10.1016/j.neuropsychologia.2012.11.025>.
- Pobric, G., Jefferies, E., Lambon Ralph, M.A., 2010. Amodal semantic representations depend on both anterior temporal lobes: evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia* 48 (5), 1336–1342. <https://doi.org/10.1016/j.neuropsychologia.2009.12.036>.
- Price, C.J., Devlin, J.T., 2003. The myth of the visual word form area. *Neuroimage* 19 (3), 473–481.
- Ritchey, M., Wing, E.A., LaBar, K.S., Cabeza, R., 2013. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebr. Cortex* 23 (12), 2818–2828. <https://doi.org/10.1093/cercor/bhs258> (New York, N.Y.: 1991).
- Rugg, M.D., King, D.R., 2018. Ventral lateral parietal cortex and episodic memory retrieval. *Cortex* 107, 238–250. <https://doi.org/10.1016/j.cortex.2017.07.012>.
- Smith, E.E., Medin, D.L., 1981. *Categories and Concepts*, vol. 9. Harvard University Press Cambridge, MA.
- Squire, L.R., Alvarez, P., 1995. Retrograde amnesia and memory consolidation: a neurobiological perspective. *Curr. Opin. Neurobiol.* 2 (5), 169–177.
- Staresina, B.P., Henson, R.N.A., Kriegeskorte, N., Alink, A., 2012. Episodic reinstatement in the medial temporal lobe. *J. Neurosci.* 32 (50), 18150–18156. <https://doi.org/10.1523/JNEUROSCI.4156-12.2012>.
- Takashima, A., Nieuwenhuis, I.L.C., Jensen, O., Talamini, L.M., Rijpkema, M., Fernández, G., 2009. Shift from hippocampal to neocortical centered retrieval network with consolidation. *J. Neurosci.: Off. J. Soc. Neurosci.* 29 (32), 10087–10093. <https://doi.org/10.1523/JNEUROSCI.0799-09.2009>.
- Tokowicz, N., Kroll, J.F., de Groot, A.M.B., van Hell, J.G., 2002. Number-of-translation norms for Dutch–English translation pairs: a new tool for examining language production. *Behav. Res. Methods Instrum. Comput.* 34 (3), 435–451. <https://doi.org/10.3758/BF03195472>.
- Tompary, A., Davachi, L., 2017. Consolidation promotes the emergence of representational overlap in the Hippocampus and medial prefrontal cortex. *Neuron* 96 (1), 228–241. <https://doi.org/10.1016/j.neuron.2017.09.005> e5.
- Tyler, Moss, H., 2001. Towards a distributed account of conceptual knowledge. *Trends Cognit. Sci.* 5 (6), 244–252. [https://doi.org/10.1016/S1364-6613\(00\)01651-X](https://doi.org/10.1016/S1364-6613(00)01651-X).
- Tyler, Stamatakis, E.A., Bright, P., Acres, K., Abdallah, S., Rodd, J.M., Moss, H.E., 2004. Processing objects at different levels of specificity. *J. Cognit. Neurosci.* 16 (3), 351–362. <https://doi.org/10.1162/089992904322926692>.
- Uncapher, M., Wagner, A.D., 2009. Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual attention theory. *Neurobiol. Learn. Mem.* 91 (2), 139. <https://doi.org/10.1016/j.nlm.2008.10.011>.
- van Kesteren, M.T.R., Ruiters, D.J., Fernández, G., Henson, R.N., 2012. How schema and novelty augment memory formation. *Trends Neurosci.* 35 (4), 211–219. <https://doi.org/10.1016/j.tins.2012.02.001>.
- van Orden, G.C., 1987. A ROWS is a ROSE: spelling, sound, and reading. *Mem. Cognit.* 15 (3), 181–198. <https://doi.org/10.3758/BF03197716>.
- Visser, R.M., Scholte, H.S., Beemsterboer, T., Kindt, M., 2013. Neural pattern similarity predicts long-term fear memory. *Nat. Neurosci.* 16 (4), 388–390. <https://doi.org/10.1038/nn.3345>.
- Wagner, I.C., van Buuren, M., Bovy, L., Fernández, G., 2016. Parallel engagement of regions associated with encoding and later retrieval forms durable memories. *J. Neurosci.* 36 (30), 7985–7995. <https://doi.org/10.1523/JNEUROSCI.0830-16.2016>.
- Ward, E.J., Chun, M.M., Kuhl, B.A., 2013. Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. *J. Neurosci.* 33 (37), 14749–14757. <https://doi.org/10.1523/JNEUROSCI.4889-12.2013>.
- Wing, E.A., Ritchey, M., Cabeza, R., 2015. Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *J. Cognit. Neurosci.* 27 (4), 679–691. https://doi.org/10.1162/jocn_a.00740.
- Wirebrink, L.K., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., Nyberg, L., 2015. Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *J. Neurosci.* 35 (26), 9595–9602. <https://doi.org/10.1523/JNEUROSCI.3550-14.2015>.
- Xiao, X., Dong, Q., Chen, C., Xue, G., 2016. Neural pattern similarity underlies the mnemonic advantages for living words. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 79, 99–111. <https://doi.org/10.1016/j.cortex.2016.03.016>.
- Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J.A., Poldrack, R.A., 2010. Greater neural pattern similarity across repetitions is associated with better memory. *Science (New York, N.Y.)* 330, 97–101. <https://doi.org/10.1126/science.1193125>, 6000.