ORIGINAL ARTICLE

Neural Patterns are More Similar across Individuals during Successful Memory Encoding than during Failed Memory Encoding

Griffin E. Koch[1,2,3], John P. Paulus[1,2] and Marc N. Coutanche[1,2,3,4]

[1]Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260, USA, [2]Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260, USA, [3]Center for the Neural Basis of Cognition, Pittsburgh, PA 15260, USA and [4]Brain Institute, University of Pittsburgh, Pittsburgh, PA 15260, USA

Address correspondence to Griffin E. Koch. Email: griffinkoch@pitt.edu

Abstract

After experiencing the same episode, some people can recall certain details about it, whereas others cannot. We investigate how common (intersubject) neural patterns during memory encoding influence whether an episode will be subsequently remembered, and how divergence from a common organization is associated with encoding failure. Using functional magnetic resonance imaging with intersubject multivariate analyses, we measured brain activity as people viewed episodes within wildlife videos and then assessed their memory for these episodes. During encoding, greater neural similarity was observed between the people who later remembered an episode (compared with those who did not) within the regions of the declarative memory network (hippocampus, posterior medial cortex [PMC], and dorsal Default Mode Network [dDMN]). The intersubject similarity of the PMC and dDMN was episode-specific. Hippocampal encoding patterns were also more similar between subjects for memory success that was defined after one day, compared with immediately after retrieval. The neural encoding patterns were sufficiently robust and generalizable to train machine learning classifiers to predict future recall success in held-out subjects, and a subset of decodable regions formed a network of shared classifier predictions of subsequent memory success. This work suggests that common neural patterns reflect successful, rather than unsuccessful, encoding across individuals.

Key words: episodic memory, encoding, hippocampus, MVPA, recall

Introduction

Imagine that two people witness a firefighter rescuing a cat from a tree. The following day, both remember the episode well. Were their brains functioning in the same manner to allow them to successfully encode this episode, and later retrieve it? Why might one person remember the steps taken by the firefighter to convince the cat to climb down, but not how many firefighters assisted, while her friend only remembers the cat's color? How can these friends differ in their particular memories for this episode yet both remember stopping for ice cream on the way home? This scenario highlights an important question: how do the brains of people who witness the same episode diverge during encoding to produce differences in what they later retrieve, while at other times converge to give similarities in what can later be remembered?

Successful episodic memory formation is associated with the modulation of activity levels in the medial temporal, prefrontal, and parietal lobes (Spaniol et al. 2009). More recently, a role in memory encoding has been identified for the multivariate patterns of activity found within memory regions. For instance, items that are subsequently remembered show more consistent activity patterns across repetitions (Xue et al. 2010). Similarly, successfully encoded words have greater neural similarity with each other than do words that are not encoded (Davis et al. 2014).

Findings of relationships between pattern similarity during encoding and subsequent memory performance have largely been restricted to patterns recorded within an individual brain's unique spatial pattern organization (LaRocque et al. 2013; Xue et al. 2013; with some exceptions noted below). Therefore, questions remain about the existence of a common "intersubject" (e.g., across-participant rather than across-item) neural pattern that indicates successful, or failed, encoding.

There is reason to believe that shared encoding signals might be observed across people. The univariate blood oxygenation level-dependent (BOLD) response of voxels in temporal regions, including the parahippocampal gyrus, has a more similar time-course between the pairs of subjects who subsequently remember a video, compared with subjects who do not (Hasson et al. 2008), suggesting that successful versus failed encoding have systematic differences in key regions. These regions might contain (time-invariant) multivariate signals across people that differentiate successful from failed encoding. Such intersubject signatures have provided insights into a variety of psychology phenomena, including social structures (Parkinson et al. 2018), psychological perspectives (Lahnakoski et al. 2014), and semantic memory (Shinkareva et al. 2011).

What forms might intersubject neural signatures of successful memory encoding take? On the one hand, neural representations evoked during encoding might be idiosyncratic to each individual, allowing within-subject, but not intersubject, similarity (Todd et al. 2013). This could arise from natural variation in how people process a video during memory encoding. For instance, actively relating stimuli to past personal experiences during an initial exposure might give distinct encoding signatures across people due to their differing past experiences. Similarly, alternate perceptual and attentive strategies during encoding might reduce commonalities in people's neural patterns. On the other hand, if successful encoding involves a region holding robust representations of observed episodes, the underlying neural representations might be common across people due to the shared episode. In this case, the generalizability of one person's neural representation to others' might index how well a particular episode is represented during encoding, which in turn will predict subsequent memory. Although the human memory system likely includes regions that fulfill each of the above scenarios, we have yet to identify the consequences of shared patterns of encoding activity for subsequent memory.

In an initial step toward understanding the role of shared patterns in memory success, Chen et al. (2017) investigated the representations of audio-visual information during encoding and recall using a novel "intersubject pattern correlation" framework. Neural representations during encoding and subsequent recall were highly similar within subjects in the posterior medial cortex (PMC), medial prefrontal cortex, and parahippocampal cortex (PHC). Moreover, subjects' activity patterns during recall were more similar to the recall patterns of other subjects, even to a greater degree than with their own encoding patterns. This pivotal study identified intersubject similarity during perception but did not consider how this relates to subsequent memory success (because the similarity reflected both encoded and not-encoded information). The question thus remains of whether (and how) intersubject similarity reflects encoding success. Furthermore, is intersubject similarity present for information that is not successfully encoded? These are important questions because, when memory outcomes are not considered, pattern-similarity can be driven by similar perceptual (but not necessarily memory) representations (Chen et al. 2017). This leaves open the question of which shared patterns are important for whether an episode is successfully encoded or not. It is also unclear if this similarity is based on shared patterns that are idiosyncratic to an observed episode or are common to all episodes.

We address these questions by directly comparing neural representations during movie-viewing for episodes that were subsequently remembered by some subjects, but not by others, in regions of the brain's memory systems and high-level cortical regions. Subjects viewed naturalistic videos during a functional magnetic resonance imaging (fMRI) scan and were then asked to recall information pertaining to specific episodes within the videos. Neural patterns for individual episodes (as well as those collapsed across episodes) were compared between subjects who later remembered, or did not remember, the details of each individual episode. This allowed us to identify memory signatures for encoding success, and for encoding failure, across individuals, as well as to examine whether these shared patterns contain information that is episode-specific or related to global encoding.

Materials and Methods

Participants

Participants were recruited until 20 contributed usable data. Twenty-three right-handed, native English speakers without a learning or attention disorder were recruited from the University of Pittsburgh community (12 females, mean [M] age = 24.2, standard deviation [SD] = 6.7). Three participants were excluded from further analyses (one for excessive head movement during the scan, one for a timing synchronization error with the scanner, and one for an identified anatomical brain abnormality). The remaining 20 participants' data were included in all the following analyses and results. The institutional review board at the University of Pittsburgh approved all measures prior to data collection. Participants were compensated for their participation.

Stimuli and Materials

The stimuli consisted of 24 nature documentary video clips of six unfamiliar animals (two fish, two birds, and two mammals; four clips per animal) without sound. Animals were intentionally chosen based on their unfamiliarity in order to minimize the effects of prior knowledge on memory retrieval (using familiarity ratings through Amazon Mechanical Turk, in which independent subjects rated familiarity on a scale of 1–5 [1 not familiar; 5 very familiar]; animals in the current study: M = 1.9; control familiar animals not included in the current study: M = 4.1). Each clip had a 45-s duration.

Experimental Design

During the first session, subjects completed a brief demographic questionnaire and safety procedures. Subjects then underwent an anatomical scan, followed by four functional runs of viewing the animal videos and three functional runs of tasks unrelated to the videos (not analyzed here).

Prior to the functional runs, subjects were given brief instructions to pay attention so they could answer a question after each silent clip. Subjects were not told that their memory would later be tested. Each of the four functional runs of video clips consisted of one video clip for each of the six animals,

interspersed with 4 s of a perceptual question (e.g., "Is this animal smaller than a car?"), followed by 4 s of fixation ("+"), before the next video clip began. All videos were presented using MATLAB (R2017a) and the Psychophysics Toolbox Version 3 (Brainard 1997; Kleiner et al. 2007), which synchronized stimulus onset with fMRI data acquisition. Upon completion of the scan, subjects were asked to respond to questions about half of the animals they had seen during the scan, with increasing degrees of specificity: the first set of questions (lowest level of specificity) asked subjects to freely recall as much information about each of the animals' videos as they could. Subjects were prompted to type their answers in response to a general question ("Can you tell me what you saw in the video?") corresponding to one of the animals they had seen during the scan. The second set of questions probed subjects' responses to more specific types of memory: episodic ("If you had to describe a day in the life of this animal, what would you say?"), semantic ("What do you now know about this animal after watching this video?"), or spatial ("Can you tell me about the places where this animal was?"). The third (most specific) set included four episodic questions (e.g., "What did the animal do while it yawned?"), eight semantic questions (e.g., "Describe this animal's ears"), and four spatial questions (e.g., "In which direction did the animal climb in the ree?"). This final set of questions always referred to episodes that were associated with specific time points from each of the video clips (total of 24 episodes), allowing timestamps to be created for each set of episodic questions, which is required for the fMRI analyses we conduct here. Three independent coders identified the specific video time points corresponding to each episodic question. The three coders discussed any discrepancies until a conclusion was reached on the video time window corresponding to each question. The episodes corresponding to each question ranged from 2 to 34 s ($M = 8.79$ s, SD $= 8.88$ s).

Subjects returned approximately 24 h after the first session for a subsequent behavioral session in which they answered the three types of questions for the other half of the animals (i.e., those that had not been tested in the first session), and completed cognitive surveys.

Individual Difference Measures

Subjects completed the Verbal-Visualizer Questionnaire, a survey that provides a measure of an individual's reliance on verbalizing and visualizing dimensions (Richardson 1977); the Survey of Autobiographical Memory (SAM), a self-report measure of a person's naturalistic mnemonic tendencies for episodic, semantic, and spatial autobiographical memory (Palombo et al. 2013); the Pittsburgh Sleep Quality Index, a survey which assesses an individual's quality of sleep in the past month (Buysse et al. 1989); and an animal familiarity survey, on which subjects rated their pre-experiment familiarity (on a 1–7 Likert scale) with each of the six animals. We verified that the SAM scores were not redundant with subjects' retrieval performance for the videos (correlation between the number of episodes retrieved and the episodic SAM scores: $r = -.28, P = 0.229$). Because these measures were not redundant, we included both as fixed effects terms within the regression models relating to memory in order to control for such individual differences.

Behavioral Analysis

To quantify memory performance in response to the specific set of episodic memory questions (e.g., "What did the animal

do while it yawned?"), an answer key was created by three independent coders, who listed all possible correct answers to each question while watching the videos. This answer key was then used to score subjects' responses as either entirely correct, partially correct, incorrect, or did not recall. Discrepancies were discussed by the coders until a final score was agreed upon, which was then used to index memory performance for each episode. For all following analyses, episodes were considered "retrieved" if the response to the corresponding question was scored as either entirely correct or partially correct. Episodes with responses scored as either incorrect or did not recall were considered "not-retrieved."

fMRI Acquisition

Subjects were scanned at the University of Pittsburgh's Neuroscience Imaging Center using a Siemens 3-T head only Allegra magnet and standard single channel radio-frequency coil equipped with mirror device to allow for fMRI stimulus presentation. The scanning session first consisted of a T_1-weighted anatomical scan (TR $= 1540$ ms, TE $= 3.04$ ms, voxel size $= 1.00 \times 1.00 \times 1.00$ mm) acquired in the straight Sagittal plane, followed by T_2-weighted functional scans which collected BOLD signals using a one-shot EPI pulse (TR $= 2000$ ms, TE $= 25$ ms, field of view $= 200$ mm, voxel size $= 3.125 \times 3.125 \times 3.125$ mm, 36 slices, flip angle $= 70°$, 168 volumes per functional run) acquired parallel to the anterior commissure-posterior commissure line. Whole brain coverage during the functional scans was achieved in most subjects; however, if whole brain coverage was not feasible, regions of the brainstem and cerebellum were clipped as they did not comprise any of the regions of interest (ROIs).

fMRI Preprocessing

Preprocessing was performed using the Analysis of Functional Neuroimages software (Cox 1996) and consisted of the following: motion correction registration, high-pass filtering, and scaling voxel activation values to have a mean of 100 (maximum limit of 200). To allow us to compare intersubject data, structural and functional images were warped to standardized space (Talairach 1988). Data were not smoothed. Following the preprocessing and standardizing steps, functional data were imported into MATLAB using the Princeton multivoxel pattern analysis toolbox (Detre et al. 2006) and custom MATLAB scripts.

Regions of Interest

We focused our analyses on three categorizations of brain regions (posterior, medial-temporal, and visual/attention) because of their proposed involvement in episodic memory, and perceiving dynamic stimuli. The posterior regions included dorsal (dDMN) and ventral Default Mode Networks (vDMN), as well as PMC and posterior cingulate cortex (PCC), while the medial-temporal regions included hippocampus and PHC. The visual/attention categorization consisted of anterior cingulate cortex (ACC), as well as a primary visual region, and the fusiform gyrus (FG), based on its response to biological motion and animacy (Shultz and McCarthy 2014). This region also contains multivariate patterns that reflect the properties of observed animals (Coutanche and Koch 2018) and higher level visual processing (Coutanche et al. 2016).

To isolate these ROIs, anatomical masks were defined for FG, hippocampus, PHC, ACC (consisting of rostral and caudal
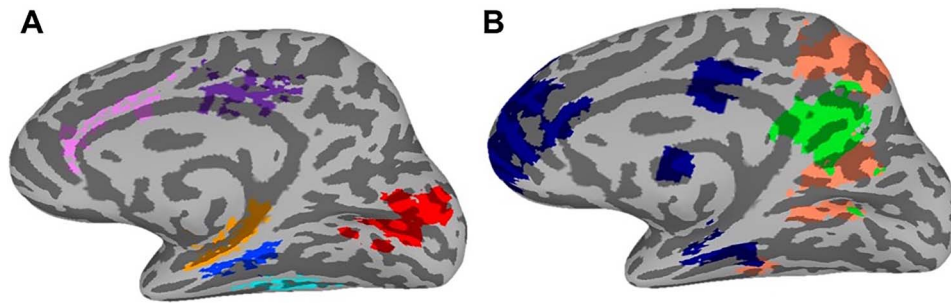
Figure 1. Regions of interest displayed on standard TT_N27 template brain. A (primary visual network: red; FG: turquoise; hippocampus: orange; PHC: blue; ACC: pink; PCC: purple). B (dDMN: navy; vDMN: salmon; PMC: green).

subregions), and PCC within each subject's native anatomical space using FreeSurfer's automated segmentation procedure (http://surfer.nmr.mgh.harvard.edu; Fischl et al. 2002; Fischl et al. 2004). Each subject's anatomical ROI masks were then standardized to Talairach space. Using these standardized masks, we created group-level masks for each of the aforementioned anatomical ROIs consisting of only voxels that were present in a minimum of 15 subjects. Additionally, ROIs for a primary visual network, dDMN, vDMN, and PMC were taken from an atlas defined using resting-state connectivity analyses (http://findlab.stanford.edu/functional_ROIs.html; Shirer et al. 2012) and warped to Talairach space. The PMC was identified as the posterior medial cluster within the dDMN ROI, as utilized by Chen et al. (2017). ROIs displayed on a standardized brain can be seen in Figure 1.

intersubject Memory Encoding Signature

Pattern similarity was measured across subjects (Chen et al. 2017). The BOLD response timecourse belonging to time points associated with every (subsequently tested) episode was shifted by three TRs to account for the hemodynamic delay. For each episode, we calculated an average activity pattern for these associated TRs (i.e., average of the z-scored BOLD activity pattern vectors collapsed across time), hereafter referred to as the episode's activity pattern. This procedure was conducted for each ROI, resulting in episode activity patterns for each subject, within each ROI (see Fig. 2). Each of the following pattern similarity analyses uses these episode activity patterns.

The first analysis involved comparing representations during the initial viewing of the animal video clips between subjects. We compared one subject's episode activity pattern with the average activity pattern for the same episode from the remaining 19 subjects using a Pearson correlation through a leave-one-subject-out cross-validation. The resulting Pearson correlation r-value was then Fisher-Z transformed. This procedure was conducted 24 times (once for each episode) across 20 iterations, where each leave-one-subject-out iteration compared a subject's episode activity pattern with the remaining group's average episode activity pattern using Pearson correlation.

After establishing a procedure to compare representations between subjects, we then asked how such representations might differ based on whether the particular episode had been retrieved or not. We followed the same procedure as above, but instead of correlating a subject's episode pattern with the entire group, we restricted the correlated group to only those other subjects who also successfully retrieved the episode (episodes scored as either entirely correct or partially correct).

We thus correlated each individual subject's episode activity pattern with a newly created "successful retrieval" group for each episode. Because every subject retrieved some episodes and not others, we could again conduct these correlations 24 times (once for each episode) across 20 iterations, where the subjects included in the successful retrieval group varied based on who had retrieved a particular episode.

Subjects' memories were tested immediately after scanning for half the animals and 24 h later for the other half. Because of this, we compared these encoding/retrieval delays by restricting the group average pattern to the same delay (i.e., either ∼ 15 min or 24 h after encoding).

Finally, we compared global encoding patterns associated with successful retrieval by averaging activity patterns across all the episodes that a subject retrieved (i.e., collapsing across episodes) and across all episodes that were not retrieved. We followed the same procedures as described above but conducted correlations between the individual's average retrieved episode pattern and the group's average retrieved episode pattern (and between an individual's average not-retrieved episode pattern and the group's average not-retrieved episode pattern).

Statistical Analysis

We report the following results based on linear mixed effects models (Baayen et al. 2008) that predict the Fisher-Z r-value between activity patterns within each ROI. We employed variables for subject, animal, and episode as random effects within each model (unless otherwise noted). We also included fixed effects terms for episodic behavioral accuracy and SAM episodic score (which had been mean-centered prior to inclusion in the model). Results pertaining to statistical significance for specific fixed effects factors were evaluated by comparing models with and without the factor of interest, and inferred based on the comparisons of AIC values (ΔAIC). When reporting mean and SDs, we first calculated average values at the subject-level, before calculating the group averages.

Machine Learning Decoding

We tested discriminability between the patterns of activity using a machine learning classifier within each ROI. We trained and tested a Gaussian Naïve Bayes (GNB) classifier on the average encoding patterns for subsequently retrieved and not-retrieved episodes for each subject, using a leave-one-subject-out cross-validation approach. This was repeated 20 times, so that each subject acted as the test subject once. Average classification accuracy for the 20 cross-validation iterations resulted in
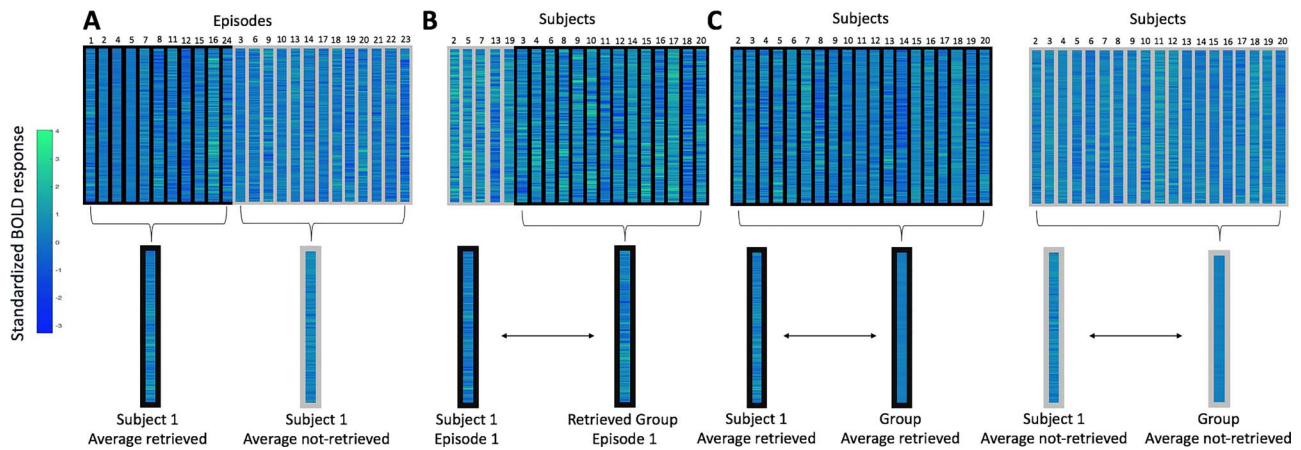
Figure 2. Overview of methodological approach for measuring the similarity of neural representations for individual episodes, and when collapsed across episodes. Columns outlined in black represent episodes that were subsequently retrieved; columns outlined in gray represent episodes that were not subsequently retrieved. Rows represent individual voxels within the ROI. Colors within the patterns represent z-scored BOLD response. A: Procedure for collapsing individual episodes within a subject to create average patterns based on subsequent memory (retrieved and not-retrieved). B: Procedure for correlating patterns for average individual episodes between each subject and other subjects showing successful retrieval. C: Procedure for correlating average patterns for collapsed episodes between individual subjects and the group.

a single classification accuracy. Statistical significance was calculated through permutation testing, in which a null distribution of 10 000 values was generated by randomizing the "retrieved" or "not-retrieved" labels for each subject. We also examined "decoding concordance" between ROIs by conducting a chi-square test of proportions of common classifier prediction across regions (e.g., Coutanche and Thompson-Schill 2015). This approach allowed us to determine whether the ROIs show concordance in the information they carry as a network (Anzellotti and Coutanche 2018) more than would be expected if the ROIs functioned independently. Statistical significance for the chi-square analysis was further supported by permutation testing in which a null distribution of 10 000 chi-square statistic values was generated by randomizing the classifier's predicted labels ("retrieved" or "not-retrieved"), and then calculating the observed and expected frequencies.

Results

This study compared the neural patterns that underlie successful or failed incidental encoding of episodes. Subjects' brain activity was recorded as they observed videos. Subsequent questions tested subjects' memory for episodes within videos. We examined neural signatures that were shared between subjects when the same episode was successfully encoded (reflected in success at answering the questions) versus unsuccessfully encoded.

Behavioral Memory Performance

Subjects differed in the number of questions they answered correctly ($M = 7.95$, $SD = 3.22$, range $= 4–16$).

intersubject Memory Encoding Signature

We first investigated neural activity recorded as subjects viewed the episodes, without considering subsequent memory success. This first analysis of intersubject patterns during encoding was consistent with previous studies by finding similar neural

Table 1 Representational similarity (Fisher-Z r-values) within regions showing statistically significant differences between retrieved and not-retrieved episodes
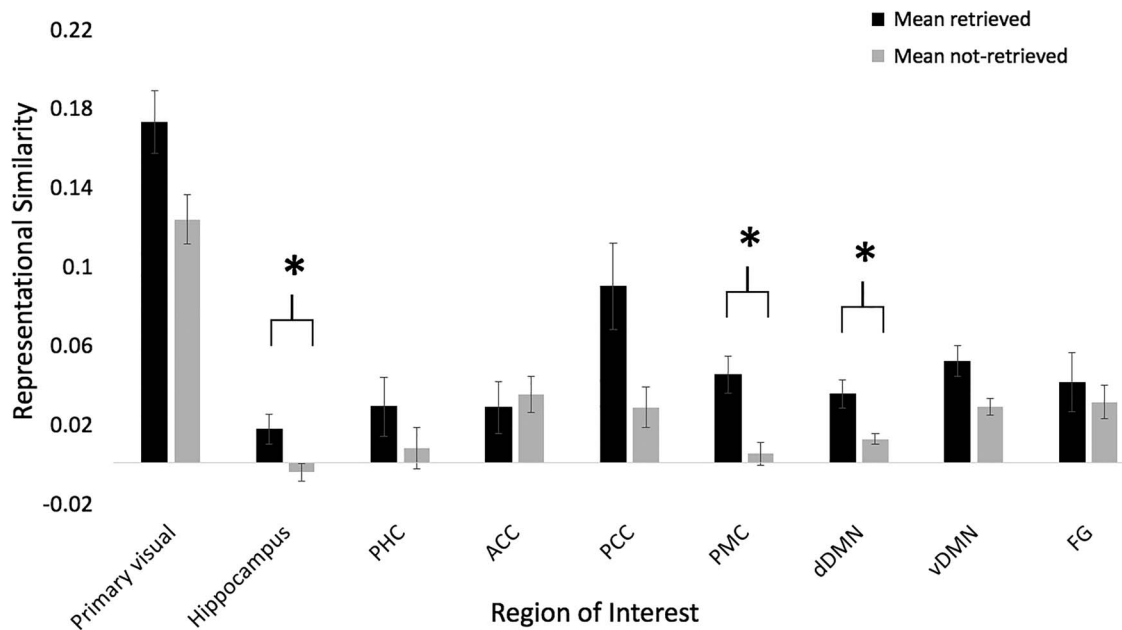
Region	Retrieved	Not-retrieved
Hippocampus	0.02 (0.03)	0.00 (0.02)
PMC	0.04 (0.04)	0.00 (0.03)
dDMN	0.03 (0.03)	0.01 (0.01)

Note. Values reflect mean. Values within parentheses reflect SD.

representations across subjects viewing the same dynamic stimuli (Chen et al. 2017). The primary visual network ($B = 0.19$, $P < 0.001$), FG ($B = 0.05$, $P = 0.015$), ACC ($B = 0.03$, $P = 0.024$), and PCC ($B = 0.06$, $P = 0.010$), as well as the dDMN ($B = 0.02$, $P = 0.017$) and vDMN ($B = 0.05$, $P = 0.011$), each showed significant similarity in their neural representations for each individual episode. The PMC showed a trend toward statistical significance ($B = 0.02$, $P = 0.062$). No other ROIs had neural representations with significant similarity (all $Ps > 0.420$).

In order to determine whether representations varied between episodes based on subsequent memory, we included a fixed effect within the regression models for each ROI to indicate whether the subject had successfully retrieved the episode or not. This factor for retrieval success was a significant predictor within the hippocampus (Δ AIC $= -3.04$, $\chi^2(1) = 5.04$, $P = 0.025$), PMC (Δ AIC $= -6.77$, $\chi^2(1) = 8.77$, $P = 0.003$), and dDMN (Δ AIC $= -3.00$, $\chi^2(1) = 5.01$, $P = 0.025$), see Figure 3. No other ROIs showed statistically significant differences (all $Ps > 0.194$). Within each statistically significant ROI, the retrieved episodes had greater similarity (Fisher-Z r-value) than not-retrieved episodes, see Table 1 for more details.

We next asked whether the above results might reflect a classic (univariate) subsequent memory effect (Brewer et al. 1998; Kim 2011; Wagner et al. 1998), rather than patterns of representations, using several different approaches. First, we used the same regression framework above to test whether each of the three ROIs (with multivariate effects) also had different univariate responses based on subsequent retrieval

Figure 3. Representational similarity, quantified as Fisher-Z *r*-values, of individual retrieved and not-retrieved episodes within each ROI. ∗indicates statistical significance ($P < 0.05$). Error bars reflect the standard error of the mean.

success. The factor for retrieval success was a significant predictor of univariate activity within the dDMN (Δ AIC $= -2.33$, $\chi^2(1) = 4.33$, $P = 0.037$), but not within the hippocampus nor PMC (both $Ps > 0.096$). Within the dDMN, retrieved episodes had less activity than not-retrieved episodes (dDMN: $M_{retrieved} = -0.03$, $SD_{retrieved} = 0.09$, $M_{not\text{-}retrieved} = 0.01$, $SD_{not\text{-}retrieved} = 0.05$). This suggests that a subsequent memory effect cannot explain the multivariate effects observed within the hippocampus and PMC, though might for the dDMN (a negative subsequent memory effect; though we note these are not mutually exclusive).

Second, we asked whether the patterns that predicted subsequent retrieval were episode-specific (i.e., represented content of the episodes), or if they reflected a global memory signal (such as that associated with the subsequent memory effect). We examined this by asking whether intersubject patterns of the same successfully recalled episode were more similar to each other than to intersubject patterns of other successfully recalled episodes. A global memory signal without content information would not distinguish between episodes that are all successfully retrieved (as all would contain the same global memory signal). For each subject, the pattern of each retrieved episode was correlated with the average pattern from other subjects who also successfully retrieved the same episode. This was compared with the correlation between each retrieved episode and a different retrieved episode's average pattern from the other subjects. We note that this was restricted to successfully retrieved episodes, so that any differences must reflect episode-specific information, beyond a general signal of retrieval success. These Fisher-Z *r*-values were analyzed using a similar regression framework as described above. Our key predictor was the difference between the same-episode comparisons (e.g., subject's episode 2 and group's episode 2) and the different-episode comparisons (e.g., subject's episode 2 and group's episode 5). Patterns of activity within the PMC and dDMN showed evidence of episode-specific

content (PMC: Δ AIC $= -11.90$, $\chi^2(1) = 13.89$, $P < 0.001$; dDMN: Δ AIC $= -14.20$, $\chi^2(1) = 16.20$, $P < 0.001$). In both regions, there was greater representational similarity between the same-episode comparisons than the different-episode comparisons (PMC: $M_{same} = 0.04$, $SD_{same} = 0.04$, $M_{different} = 0.01$, $SD_{different} = 0.02$; dDMN: $M_{same} = 0.03$, $SD_{same} = 0.03$, $M_{different} = 0.01$, $SD_{different} = 0.02$). Patterns of activity from the hippocampus did not reach statistical significance for this comparison ($P = 0.170$), although the numerical difference between the same- and different-episode comparisons was in the same direction as in the PMC and dDMN (Hippocampus: $M_{same} = 0.01$, $SD_{same} = 0.03$, $M_{different} = 0.00$, $SD_{different} = 0.01$). These findings suggest the PMC and dDMN contain episode-specific information in their patterns of subsequently retrieved episodes (and not simply a global memory signal), though patterns in the hippocampus were less episode-specific. It is worth noting that the earlier dDMN finding of a (negative) subsequent memory effect and these episode-specific patterns are not contradictory, because an overall univariate response and representational patterns can (and frequently do) coexist in a region's activity (Coutanche 2013).

We next asked if the shared similarity, in addition to predicting retrieval success, would reflect the amount of details that was spontaneously recalled in the free recall task. Independent coders identified and totaled each meaningful unit of information in every subject's set of free-recall responses. Linear regression models used these values to predict the strength of each subject's representational similarity with other subjects (for successfully retrieved episodes). Within the three regions showing multivariate differences between retrieved and not-retrieved episodes (as measured through the specific cued recall questions), we did not find a relationship with the number of details spontaneously given during free recall (all $Ps > 0.287$), suggesting that intersubject similarity was particularly sensitive to accuracy to questions about specific episodes, rather than an open-ended free recall cue.

Table 2 Representational similarity (Fisher-Z r-values) within hippocampus

Day	Delays included in group comparison	Retrieved	Not-retrieved
1	Both delays	0.00 (0.04)	0.00 (0.03)
2	Both delays	0.03 (0.05)	−0.01 (0.03)
1	Same delay	0.00 (0.05)	0.00 (0.04)
2	Same delay	0.05 (0.05)	0.00 (0.04)

Note. Values reflect mean. Values within parentheses reflect SD. Day 1 refers to episodes retrieved on the same day as initial movie viewing (∼15 min later) and Day 2 refers to episodes retrieved on the day following movie viewing (24 h later). Delays included in group comparison refer to the makeup of the group to which the individual subjects' episode patterns were compared. In the first analysis, we collapsed across all subjects to create the group average pattern (both delays), whereas in the second analysis, we restricted the group average pattern to only include subjects who had the same delay as the individual subject to which they were being compared (same delay).

In order to determine whether the encoding/retrieval delay moderates the relationship between the observed intersubject encoding signature and future retrieval success, we compared a model with a fixed effect term for the interaction between retrieval success and day of retrieval (same or following day), to the same model without this interaction, within the three regions that reflected subsequent retrieval success (hippocampus, PMC, and dDMN). The hippocampus showed a significant interaction between day and retrieval success (Δ AIC = −4.58, $\chi^2(1) = 6.57$, $P = 0.010$). The mean correlation values showed that hippocampal pattern similarity for the retrieved episodes was greater for the longer retrieval delay (i.e., between Days 1 and 2) than for the shorter retrieval delay, see Table 2 for details. Neither the PMC nor dDMN ROIs showed statistically significant differences between days ($Ps > 0.497$).

We then probed this effect further in the hippocampus by restricting our group average pattern to only include subjects from the same delay group (i.e., either ∼ 15 min or 24 h), to more directly hone in on effects of time delay between encoding and retrieval. Again, the hippocampus showed a significant interaction between day and retrieval success (Δ AIC = −9.66, $\chi^2(1) = 11.66$, $P < 0.001$), see Table 2 for means across both days. As before, the mean correlation values showed greater hippocampal pattern similarity for the retrieved episodes from Day 2, compared with Day 1.

We have shown that neural representations are more similar for episodes that are subsequently retrieved compared with those that are not retrieved; however, it is possible that such differences could be driven by the amount of variance within the temporal signal of retrieved versus not-retrieved episodes. Greater variance in signal for either the retrieved or not-retrieved episodes could influence the maximum possible similarity values for one of these categories. To test this possibility, we calculated the SD of the time course of all episodes that were subsequently retrieved, and separately, for all episodes subsequently not-retrieved within each subject. No statistically significant differences were observed in the variances of the average BOLD signal between retrieved and not-retrieved episodes, in any of the ROIs (all $Ps > 0.109$), thus systematic differences in temporal variance could not account for our results.

So far, we have shown that the encoding neural representations underlying individual episodes systematically vary based on whether they are subsequently remembered. One remaining question is whether this generalizes across all episodes (i.e.,

Table 3 Representational similarity (Fisher-Z r-values) within regions showing statistically significant differences between average retrieved and not-retrieved episodes

Region	Retrieved	Not-retrieved
Hippocampus	0.04 (0.08)	−0.03 (0.08)
PCC	0.18 (0.26)	0.00 (0.12)
dDMN	0.08 (0.10)	0.03 (0.05)

Note. Values reflect mean. Values within parentheses reflect SD.

whether this can be extracted after collapsing across episode-specific patterns). To address this, we calculated representational similarity using each individual's average pattern and the group's average pattern collapsed across episodes (rather than patterns for each individual episode). Because we averaged across episodes, subject was the only random effect in these models. The similarity between individuals and the group was significantly different between all episodes that were subsequently retrieved versus all those not retrieved within the hippocampus (ΔAIC = −4.22, $\chi^2(1) = 6.22$, $P = 0.013$), PCC (Δ AIC = −5.85, $\chi^2(1) = 7.85$, $P = 0.005$), and dDMN (Δ AIC = −2.62, $\chi^2(1) = 4.62$, $P = 0.032$) as well as trending toward statistical significance in PHC (ΔAIC = −1.79, $\chi^2(1) = 3.79$, $P = 0.052$), see Figure 4. No other ROIs showed statistically significant differences (all $Ps > 0.126$). Within the statistically significant ROIs, the average retrieved representations had greater similarity (mean Fisher-Z r-values) than the average not-retrieved representations (see Table 3 for more details). Interestingly, the PHC showed a trending effect in the reverse direction, where the average not-retrieved representations had greater similarity than retrieved representations.

In comparing results from the patterns of individual episodes (Fig. 3) and the patterns that were collapsed across episodes (Fig. 4), we found average representational similarity values that were numerically higher when collapsing across the episodes. We believe this is likely to be because (by design) the collapsed patterns include more time-points from the videos. Collapsing across idiosyncrasies that influence patterns for individual episodes (which have a small number of time-points) averages-out idiosyncrasies, leaving a more reliable signal (See Supplementary Material for more details).

To again ask whether the above results might include a classic (univariate) subsequent memory effect, we used the same regression framework as above to test whether the three ROIs that showed multivariate effects for averaged episodes also had average univariate activity that differed based on subsequent retrieval success. The factor for retrieval success was a significant predictor of activity within the dDMN (Δ AIC = −2.06, $\chi^2(1) = 4.06$, $P = 0.044$), but not within the hippocampus nor PCC (both $Ps > 0.344$). Within the dDMN, the average retrieved episodes had less activity than the average not-retrieved episodes (dDMN: $M_{retrieved} = −0.03$, $SD_{retrieved} = 0.09$, $M_{not-retrieved} = 0.01$, $SD_{not-retrieved} = 0.05$). This suggests that (in a similar manner to individual episodes) intersubject similarity for averaged episodes in the hippocampus and PCC cannot be attributed to the subsequent memory effect, while the dDMN shows this effect.

Machine Learning Decoding

We trained GNB classifiers in a between-subject classification to decode the average retrieved and not-retrieved patterns from
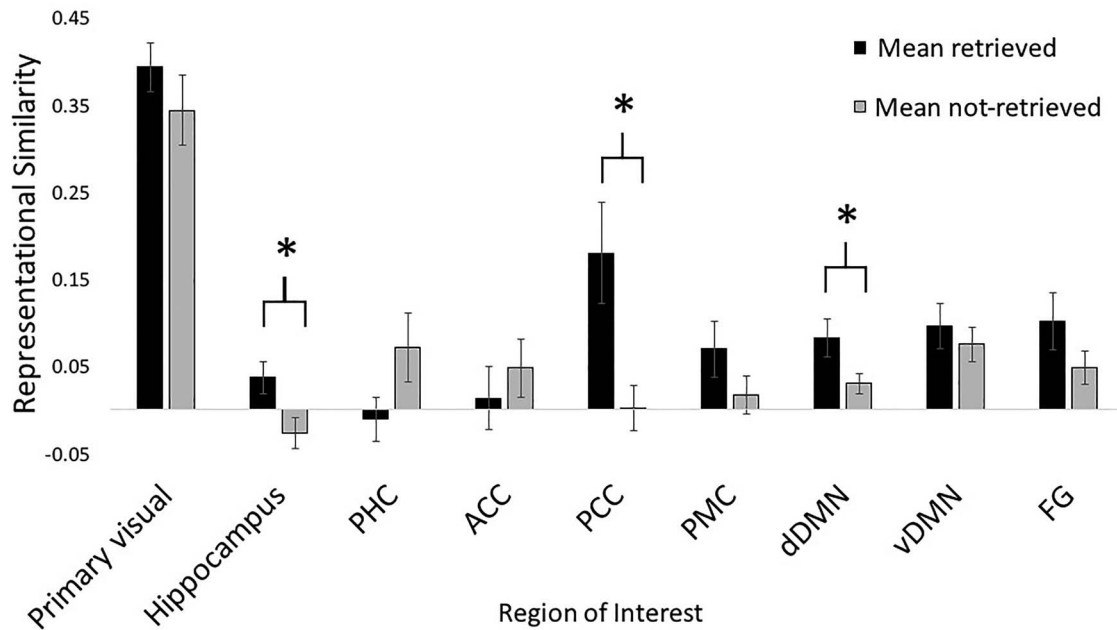
Figure 4. Representational similarity, quantified as Fisher-Z r-values, of average retrieved and not-retrieved episodes within each ROI, collapsed across episodes. ∗indicates statistical significance (P < 0.05). Error bars reflect the standard error of the mean.

each subject (used in the representational similarity analyses above). Classifiers trained on the average patterns from each of seven regions were able to decode subsequently retrieved from not-retrieved encoding patterns (statistical significance calculated from permutation testing): FG ($M = 0.68$, $P = 0.033$), ACC ($M = 0.75$, $P = 0.002$), PCC ($M = 0.80$, $P = 0.001$), PMC ($M = 0.70$, $P = 0.013$), dDMN ($M = 0.72$, $P = 0.007$), and vDMN ($M = 0.75$, $P = 0.003$). No other ROIs showed statistically significant differences (all Ps > 0.095).

Finally, to complement these classification results, we asked if there was contingency in the classifiers' predictions of patterns (as either subsequently retrieved or not-retrieved) between the ROIs with decoding success. Significant contingency would reflect a shared basis for discriminating encoding patterns. We used a chi-square analysis to statistically evaluate the concordance of ROI classification performance by comparing the actual (i.e., observed) concordance to that expected by independence. Here, concordance means that a pair of ROI classifiers predicts the same label (i.e., in agreement for predicting subsequent memory outcome as either retrieved or not-retrieved), regardless of the prediction's accuracy (since ROIs were only included if they were able to successfully classify better than chance). We perform this analysis using four of the isolated ROIs that could successfully classify between the average retrieved and not-retrieved patterns (FG, ACC, PCC, and PMC). We excluded the dorsal and ventral DMN from this analysis as they encompass multiple anatomical regions and overlap with the other ROIs to be analyzed.

Examining the four ROIs together, classifier predictions were more concordant than would be expected if the ROIs' classification varied independently ($\chi^2(5) = 14.69$, $P = 0.012$). Results were further supported through permutation testing which formed a null distribution of Chi-Square statistics resulting from the observed and expected frequencies for each of 10 000 iterations of random permuting ($P = 0.019$). Because the four ROIs together displayed evidence for a possible network of concordance, we next examined concordance between pairs of regions, again using a Chi-Square analysis. Two of the six pairwise comparisons resulted in greater concordance than would be expected if independent: ACC & PMC and PCC & PMC, as displayed in Table 4. As before, these results were further supported through permutation testing involving 10 000 iterations. Figure 5 depicts these regions in a network of concordance for predicting subsequent memory outcome.

Discussion

We have asked how the similarity of neural representations across people during memory encoding reflects encoding success (as indicated by subsequent retrieval), and encoding failure. First, we identified a neural pattern that is common across subjects regardless of subsequent memory within posterior (PCC and DMN) regions, visual regions, and the ACC, supporting findings from previous research (Chen et al. 2017; Hasson et al. 2004) that neural representations in attention-modulated regions of the brain (Davis et al. 2000) are similar across subjects viewing individual episodes. We extend our understanding by relating a common spatial organization during initial perception to behavioral outcomes of memory. Importantly, shared information is present in neural signals during encoding that can be related to successes and failures in retrieval. Notably, more than a general global memory signal (without episodic content), intersubject patterns within the dDMN and PMC reflect the content of retrieved episodes. This suggests that the neural processing underlying these shared patterns includes unique episode-specific information. It was not previously known that intersubject neural markers specific to encoding success (versus failure) contained this level of information, thus shedding light on this common neural processing. An alternative organization was for the underlying encoding processing to be more idiosyncratic to each person, such as through relating an episode to one's

Table 4 Frequency table of classifier performance

Classifier predictions	Agreement	Disagreement	$\chi^2(1)$	P	P (permutation)
ACC & PCC	30 (26)	10 (14)	1.76	0.185	0.155
ACC & PMC	**30 (24)**	**10 (16)**	**3.75**	**0.053**	**0.039**
ACC & FG	27 (23.5)	13 (16.5)	1.26	0.261	0.206
PCC & PMC	**32 (24.8)**	**8 (15.2)**	**5.50**	**0.019**	**0.021**
PCC & FG	29 (24.2)	11 (15.8)	2.41	0.121	0.139
PMC & FG	23 (22.8)	17 (17.2)	0.00	0.949	0.978

Note. Pair-wise observed frequencies of classifier agreement. Expected frequencies for independence are displayed in parentheses. Bold indicates statistically significant classifier concordance from permutation testing.
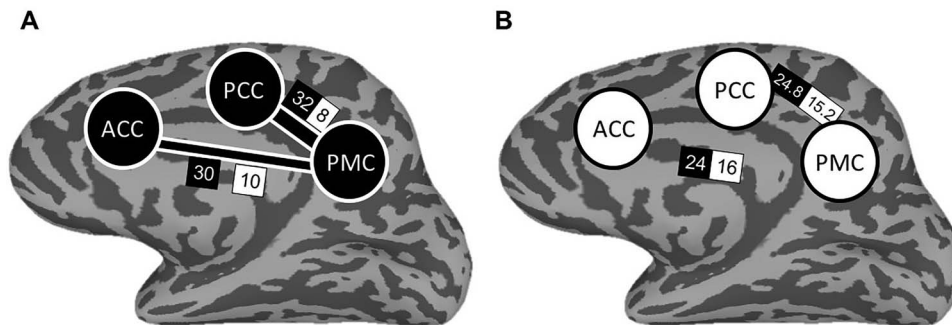
Figure 5. Depiction of the discrimination network of subsequently retrieved and not-retrieved outcomes based on encoding activity. The network consists of ACC, PCC, and PMC. Panel A depicts the observed frequencies of classifier concordance. Width of connecting bars between regions depicts the magnitude of Chi-Square statistic for classifier predictions. Panel B depicts the expected frequencies of classifier concordance if the regions were independent. Values within black boxes indicate the frequency of classifier agreement; values within white boxes indicate the frequency of disagreement.

own experiences or unique set of knowledge. This sort of individualized processing is unlikely to be accompanied by similarity across a group of people. Without comparing patterns across people (as we do here), it would not have been possible to determine if neural patterns were consistent or individualized. For instance, multivariate analyses conducted within subjects can give significant results at the group level despite differing neural principles at the individual level (Todd et al. 2013). In contrast, intersubject similarity reflects a shared neural basis—in this case for the successful encoding of episodes.

Chen and colleagues showed that regions within the DMN have event-specific activity patterns that are common across subjects during recollection (Chen et al. 2017). Our evidence suggests that these regions also have common encoding patterns that predict future retrieval success. Thus, more than reactivating earlier perceptually driven activity, the intersubject activity patterns evoked during recollection in Chen et al. (2017) likely are also impacted by the encoding processes taking place during an initial experience. Retrieval will then be affected by both the earlier encoding patterns, and the shared patterns evoked during retrieval itself. This is in line with theories that point to the importance of cognitive and neural processing during the encoding phase for the ability to later retrieve a memory (Hebscher et al. 2019; McDaniel et al. 1986; Uncapher and Wagner 2009). On the other hand, other regions with reactivated shared patterns during retrieval, such as high-level visual areas (Chen et al. 2017), did not show a relationship between intersubject similarity and encoding, suggesting that their influence on encoding might be more individualized.

For the first time, we examined intersubject neural patterns of encoding by comparing subjects for whom the same episode

is subsequently retrieved successfully versus unsuccessfully. Within the hippocampus, greater intersubject similarity predicted future memory retrieval based on patterns for individual episodes, and collapsed across episodes. Hippocampal patterns for the same successfully retrieved episodes were not more similar to each other (across subjects), than with patterns for different episodes. This suggests that common hippocampal representations are not merely driven by stimulus-specific features during encoding but rather reflect a fundamental shared pattern of neural activity for encoding (Richter et al. 2016)—a global encoding signal for subsequent memory success. The involvement of the hippocampus in episodic memory encoding is not surprising (Tulving and Markowitsch 1998), but we report a novel finding in identifying a common pattern of neural activity between people that relates to future memory success or failure. Additionally, greater similarity of intersubject hippocampal patterns for following-day (rather than same-day) retrieval is in line with research implicating the hippocampus as serving a role in the durability of memory (Sneve et al. 2015). This finding extends prior work by showing that the importance of the length of the delay extends to intersubject representational similarity for dynamic episodes (Ranganath and Hsieh 2016).

In contrast to the greater similarity for retrieved episodes within hippocampal patterns, we observed a trending result in the opposite direction (greater similarity for not-retrieved episodes) for intersubject patterns of PHC. The PHC includes regions that engage in scene and place processing (Epstein 2005), so a possible driver is greater attention to the location and context of not-retrieved episodes, rather than the temporal component that is necessary for episodic memory (Howard and Kahana 2002). This shared attention to the scene might lead to greater similarity across people.

The multivariate approach we used allowed us to investigate the discriminability of neural patterns for episodes that were remembered and those that were not. Classifiers trained on activity from both the dorsal and ventral DMNs could successfully predict subsequent memory outcomes in held-out individuals. Additionally, we identified a network of regions related to memory outcomes. Our concordance analysis of machine learning predictions suggested a network of three key regions: ACC, PCC, and PMC, where the PMC is a hub in which the representations of subsequent memory outcome are similar to those of ACC and PCC (resulting in more congruent classifier predictions than would be expected if the regions were independent). Note also that concordance between these regions is not merely a result of having similar classification performance, as shown by the absence of the FG in the network, despite its similar overall performance. Our work adds to a growing literature of group-level similarity and decoding, notably the use of leave-one-subject-out cross-validation (Wang et al. 2020), as a viable alternative to traditional methods of within-subject decoding, which can inform us about commonalities in neural activity across subjects.

While this study is an initial foray into tracking differences in subsequent memory success via shared patterns during encoding, there is room for us to expand our knowledge based on the methodology and approaches used in this study. Here, we categorized subsequent memory performance as either a success or a failure. This binary classification, while useful for identifying whether a particular episode was remembered or not, does not provide a comprehensive description of a person's memory. Within only those people who had retrieved an episode, there may still be differences in memory performance that are not quantified here. In an analysis of the number of details spontaneously free recalled, we did not find a relationship with intersubject similarity, though future studies could be designed to optimally examine this. Such investigations might wish to ask whether similarity in neural representations falls along a type of continuum that tracks other behavioral indicators of memory. Additionally, our analysis provides evidence that common neural similarity is not simply driven by a global large-scale indicator of a person's memory, but moreover can reflect specific episodes. Furthermore, the current study instructed subjects to watch the videos and respond to basic perceptual questions, but future studies might investigate how neural representations differ depending on intentionality. For instance, the intention to remember items can sometimes result in worse memory performance than when not intending to remember (Storm et al. 2007), but it is yet unknown how intentions might impact common neural representations across people. Relatedly, it can be asked whether the neural signatures we have identified can be connected to whether or not a subject attempts to remember a specific episode, in addition to whether or not they were ultimately successful.

What are the possible reasons for greater intersubject similarity for successfully encoded episodes? One possibility is that subjects who later successfully retrieved an episode used similar (effective) encoding strategies, reflected in similar encoding neural activity. Episode-specific information was detected in the common patterns of some regions, so any such encoding strategy likely involves processing the content of the episode, rather than being a more general global signal. This content might reflect narrative or situational elements of episodes, which have been linked to the shared DMN activity patterns (Chen et al. 2017). Future studies might wish to probe or manipulate subjects' encoding strategies (e.g., trying to focus solely on the animal, building a narrative story to connect the actions into a coherent plot, etc.) to examine the role these might play on intersubject similarity. Alternatively, a drop in intersubject similarity for episodes that are not later retrieved might reflect a failure to deeply engage with the episode when it is presented, due to an attentional shift or interfering process. These possibilities could be directly tested in future research.

To summarize, we have identified shared neural signatures of memory encoding that differ based on future retrieval performance. This relatively novel approach of investigating neural representations for not only encoding success, but also failures of encoding, shows that common neural signatures go beyond basic perceptual information and instead reflect memory-encoding processes that are linked to memory formation.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Funding

Notes

References

Anzellotti S, Coutanche MN. 2018. Beyond functional connectivity: investigating networks of multivariate representations. *Trends Cogn Sci*. 22(3):258–269. doi: 10.1016/j.tics.2017.12.002.

Baayen RH, Davidson DJ, Bates DM. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang*. 59(4):390–412. doi: 10.1016/j.jml.2007.12.005.

Brainard DH. 1997. The psychophysics toolbox. *Spat Vis*. 10(4):433–436.

Brewer JB, Zhao Z, Desmond JE, Glover GH, Gabrieli JDE. 1998. Making memories: brain activity that predicts how well visual experience will be remembered. *Science*. 281(5380):1185–1187. doi: 10.1126/science.281.5380.1185.

Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. 1989. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 28(2):193–213.

Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, Hasson U. 2017. Shared memories reveal shared structure in neural activity across individuals. *Nat Neurosci*. 20(1):115–125. doi: 10.1038/nn.4450.

Coutanche MN. 2013. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cogn Affect Behav Neurosci*. 13(3):667–673. doi: 10.3758/s13415-013-0186-2.

Coutanche MN, Koch GE. 2018. Creatures great and small: real-world size of animals predicts visual cortex representations

beyond taxonomic category. *NeuroImage*. 183:627–634. doi: 10.1016/j.neuroimage.2018.08.066.

Coutanche MN, Solomon SH, Thompson-Schill SL. 2016. A meta-analysis of fMRI decoding: quantifying influences on human visual population codes. *Neuropsychologia*. 82:134–141. doi: 10.1016/j.neuropsychologia.2016.01.018.

Coutanche MN, Thompson-Schill SL. 2015. Creating concepts from converging features in human cortex. *Cereb Cortex*. 25(9):2584–2593. doi: 10.1093/cercor/bhu057.

Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 29(3):162–173.

Davis KD, Hutchison WD, Lozano AM, Tasker RR, Dostrovsky JO. 2000. Human anterior cingulate cortex neurons modulated by attention-demanding tasks. *J Neurophysiol*. 83(6):3575–3577. doi: 10.1152/jn.2000.83.6.3575.

Davis T, Xue G, Love BC, Preston AR, Poldrack RA. 2014. Global neural pattern similarity as a common basis for categorization and recognition memory. *J Neurosci*. 34(22):7472–7484. doi: 10.1523/JNEUROSCI.3376-13.2014.

Detre G, Polyn SM, Moore C, Natu V, Singer B, Cohen J, Haxby JV, Norman KA. 2006. *The Multi-Voxel Pattern Analysis (MVPA) Toolbox*. Florence, Italy: Presented at the Annual Meeting of the Organization for Human Brain Mapping.

Epstein R. 2005. The cortical basis of visual scene processing. *Vis Cogn*. 12(6):954–978. doi: 10.1080/13506280444000607.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Dale AM. 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 33(3):341–355.

Fischl B, Salat DH, van der Kouwe AJW, Makris N, Ségonne F, Quinn BT, Dale AM. 2004. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*. 23(Suppl 1):S69–S84. doi: 10.1016/j.neuroimage.2004.07.016.

Hasson U, Furman O, Clark D, Dudai Y, Davachi L. 2008. Enhanced Intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*. 57(3):452–462. doi: 10.1016/j.neuron.2007.12.009.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. 2004. Intersubject synchronization of cortical activity during natural vision. *Science (New York, NY)*. 303(5664):1634–1640. doi: 10.1126/science.1089506.

Hebscher M, Wing E, Ryan J, Gilboa A. 2019. Rapid cortical plasticity supports long-term memory formation. *Trends Cogn Sci*. 23(12):989–1002. doi: 10.1016/j.tics.2019.09.009.

Howard MW, Kahana MJ. 2002. A distributed representation of temporal context. *J Math Psychol*. 46(3):269–299. doi: 10.1006/jmps.2001.1388.

Kim H. 2011. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *NeuroImage*. 54(3):2446–2461. doi: 10.1016/j.neuroimage.2010.09.045.

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C. 2007. What's new in psychtoolbox-3. *Perception*. 36(14):1–16.

Lahnakoski JM, Glerean E, Jääskeläinen IP, Hyönä J, Hari R, Sams M, Nummenmaa L. 2014. Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*. 100:316–324. doi: 10.1016/j.neuroimage.2014.06.022.

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(13):5466–5474. doi: 10.1523/JNEUROSCI.4293-12.2013.

McDaniel MA, Einstein GO, Dunay PK, Cobb RE. 1986. Encoding difficulty and memory: toward a unifying theory. *J Mem Lang*. 25(6):645–656. doi: 10.1016/0749-596X(86)90041-0.

Palombo DJ, Williams LJ, Abdi H, Levine B. 2013. The survey of autobiographical memory (SAM): a novel measure of trait mnemonics in everyday life. *Cortex*. 49(6):1526–1540. doi: 10.1016/j.cortex.2012.08.023.

Parkinson C, Kleinbaum AM, Wheatley T. 2018. Similar neural responses predict friendship. *Nat Commun*. 9(1):332. doi: 10.1038/s41467-017-02722-7.

Ranganath C, Hsieh L-T. 2016. The hippocampus: a special place for time. *Ann N Y Acad Sci*. 1369(1):93–110. doi: 10.1111/nyas.13043.

Richardson A. 1977. Verbalizer-visualizer: a cognitive style dimension. *J Ment Imag*. 1(1):109–125.

Richter FR, Cooper RA, Bays PM, Simons JS. 2016. Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *elife*. 5. doi: 10.7554/eLife.18260.

Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA. 2011. Commonality of neural representations of words and pictures. *NeuroImage*. 54(3):2418–2425. doi: 10.1016/j.neuroimage.2010.10.042.

Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD. 2012. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex*. 22(1):158–165. doi: 10.1093/cercor/bhr099.

Shultz S, McCarthy G. 2014. Perceived animacy influences the processing of human-like surface features in the fusiform gyrus. *Neuropsychologia*. 60:115–120. doi: 10.1016/j.neuropsychologia.2014.05.019.

Sneve MH, Grydeland H, Nyberg L, Bowles B, Amlien IK, Langnes E et al. 2015. Mechanisms underlying encoding of short-lived versus durable episodic memories. *J Neurosci*. 35(13):5202–5212. doi: 10.1523/JNEUROSCI.4434-14.2015.

Spaniol J, Davidson PSR, Kim ASN, Han H, Moscovitch M, Grady CL. 2009. Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia*. 47(8):1765–1779. doi: 10.1016/j.neuropsychologia.2009.02.028.

Storm BC, Bjork EL, Bjork RA. 2007. When intended remembering leads to unintended forgetting. *Q J Exp Psychol*. 60(7):909–915. doi: 10.1080/17470210701288706.

Talairach J. 1988. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging*. 1st ed. Stuttgart; New York: Thieme.

Todd MT, Nystrom LE, Cohen JD. 2013. Confounds in multivariate pattern analysis: theory and rule representation case study. *NeuroImage*. 77:157–165. doi: 10.1016/j.neuroimage.2013.03.039.

Tulving E, Markowitsch HJ. 1998. Episodic and declarative memory: role of the hippocampus. *Hippocampus*. 8(3):198–204. doi: 10.1002/(SICI)1098-1063(1998)8:3<198::AID-HIPO2>3.0.CO;2-G.

Uncapher MR, Wagner AD. 2009. Posterior parietal cortex and episodic encoding: insights from fMRI subsequent mem-

ory effects and dual-attention theory. *Neurobiol Learn Mem.* 91(2):139–154. doi: 10.1016/j.nlm.2008.10.011.

Wagner A, Poldrack R, Eldridge L, Desmond J, Glover G, Gabrieli JE. 1998. Material-specific lateralization of prefrontal activation during episodic encoding and retrieval. *Neuroreport.* 9(16):3711–3717.

Wang Q, Cagna B, Chaminade T, Takerkart S. 2020. inter-subject pattern analysis: a straightforward and powerful scheme for group-level MVPA. *NeuroImage.* 204:116205. doi: 10.1016/j.neuroimage.2019.116205.

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 (New York, NY).* 330(6000):97–101. doi: 10.1126/science.1193125.

Xue G, Dong Q, Chen C, Lu Z-L, Mumford JA, Poldrack RA. 2013. Complementary role of frontoparietal activity and cortical pattern similarity in successful episodic memory encoding. *Cereb Cortex.* 23(7):1562–1571. doi: 10.1093/cercor/bhs143.