Cortical representations of visual stimuli shift locations with changes in memory states

Running Title: Stimulus representations and memory states

Nicole M. Long[1], Brice A. Kuhl[2]

Affiliations:

1: Department of Psychology, University of Virginia 22904

2: Department of Psychology, University of Oregon 97403

Corresponding Author: Nicole Long (email: niclong@virginia.edu; twitter: @dorsolateralpfc); Brice Kuhl (email: bkuhl@uoregon.edu; twitter: @KuhlLab)

Lead Contact: Nicole Long (email: niclong@virginia.edu)

Number of Figures: 4

Number of Tables: 0

Word Counts

Summary: 246

Results: 2217

Discussion: 1184

Summary

Episodic memory retrieval is thought to rely on reactivation of the same content-sensitive neural activity patterns initially expressed during memory encoding[1–6]. Yet, there are emerging examples of content representations expressed in different brain regions during encoding versus retrieval[7–14]. While these differences have been observed by comparing encoding and retrieval tasks that differ in terms of perceptual experience and cognitive demands, there are many real-world contexts–e.g., meeting a new colleague who reminds you of an old acquaintance–where the memory system may be intrinsically biased either toward encoding (the new colleague) or retrieval (the old acquaintance)[15,16]. Here, we test whether intrinsic memory states, independent of task demands, determine the cortical location of content representations. In a human fMRI study, subjects (n = 33) viewed object images and were instructed to either encode the current object or retrieve a similar object from memory. Using pattern classifiers, we show that biases toward encoding versus retrieval were reflected in large-scale attentional networks[17–19]. Critically, memory states decoded from these networks – even when entirely independent from task instructions – predicted shifts of object representations from visual cortex (encoding) to ventral parietal cortex (retrieval). Finally, visual versus ventral parietal cortices exhibited differential connectivity with the hippocampus during memory encoding versus retrieval, consistent with the idea that the hippocampus mediates cortical shifts in content representations. Collectively, these findings demonstrate that intrinsic biases toward memory encoding versus retrieval determine the specific cortical locations that express content information.

Results

The main experiment was divided into two phases: List 1 (prior to fMRI scanning) and List 2 (during fMRI scanning; Figure 1). In List 1, subjects learned a set of 24 object images (e.g., bench, fan, etc.). In List 2, subjects saw a new set of 24 object images. Critically, each List 2 image was from the same category as a List 1 image (e.g., a new bench, a new fan) and was preceded by a cue instructing subjects to either encode the current stimulus (encode instruction) or retrieve the corresponding List 1 stimulus (retrieve instruction; Figure 1). Each List 2 image

was presented 16 times across 8 fMRI scan runs and always with the same instruction cue (encode or retrieve). No behavioral response was required for either trial type.

Cortical location of object representations shifts with changes in memory states

We first tested whether the cortical location of stimulus-specific object representations was influenced by encode vs. retrieve instructions. We measured stimulus representations in visual and parietal cortices, using eight previously-described regions of interest (ROIs)[11]. These eight ROIs corresponded to visual cortical regions [early visual cortex (V1), lateral occipital cortex (LO), and ventral temporal cortex (VTC)], ventral parietal regions [the ventral aspect of lateral intraparietal sulcus (vLatIPS), angular gyrus (AnG), ventral intraparietal sulcus (vIPS)] and dorsal parietal regions [posterior intraparietal sulcus (pIPS) and the dorsal aspect of lateral intraparietal sulcus (dLatIPS)]. Based on prior findings[11,20], we predicted that ventral parietal regions (in particular, angular gyrus) would exhibit a relative bias toward representing past experiences whereas visual regions would exhibit a relative bias toward representing current perceptual experience. In other words, we predicted that encode versus retrieve instructions would differentially influence object representations in ventral parietal versus visual regions. While we did not have specific predictions about dorsal parietal regions, we included these ROIs for consistency with prior work[11] and as a point of contrast given evidence of functional dissociations between ventral and dorsal parietal regions[11,21,22]. To test for stimulus-specific representations, we correlated activity patterns corresponding to the same object (within-object correlations) and, from this value, subtracted the mean correlation for activity patterns corresponding to different object (across-object correlations; Figure 2A). All correlations were performed using data from independent runs (see Methods). The difference between within-stimulus and between-stimulus correlations was our critical dependent measure, with values greater than 0 constituting positive evidence for stimulus-specific representations.

Figure 2B shows stimulus-specific representations as a function of instruction across the eight ROIs. We first focused on visual versus ventral parietal regions. Consistent with our prediction, a repeated measures ANOVA with factors of instruction (encode, retrieve) and cortical location (visual, ventral parietal) revealed a significant interaction between instruction and cortical loca-

tion ($F_{1,32} = 6.185$, $p = 0.0183$). In other words, instructions differentially influenced stimulus representations in visual regions versus ventral parietal regions (also see Supplementary Figure 1A). Within ventral parietal regions, significant stimulus-specific representations were observed during retrieve trials (mean $z = 0.0136$, SD $= 0.0225$; $t_{32} = 3.4654$, $p = 0.0015$), but not during encode trials (mean $z = 0.0018$, SD = 0.0221; $t_{32} = 0.4701$, $p = 0.641$) and representations were significantly stronger during retrieve than encode trials (t-test: $t_{32} = 2.1585$, $p = 0.0385$; Figure 2C). In contrast, within visual regions stimulus-specific representations were significant both during encode trials ($t_{32} = 8.7788$, $p < 0.001$) and retrieve trials ($t_{32} = 9.9331$, $p < 0.001$) with no difference between encode and retrieve trials ($t_{32} = 0.2430$, $p = 0.8096$, Figure 2C).

A separate repeated measures ANOVA also revealed that instructions differentially influenced stimulus representations in ventral versus dorsal parietal regions ($F_{1,32} = 6.185$, $p = 0.0046$). Within dorsal parietal regions, stimulus-specific representations were significant both during encode trials ($t_{32} = 2.5078$, $p = 0.0174$) and retrieve trials ($t_{32} = 2.6561$, $p = 0.0122$), with no difference between encode and retrieve trials ($t_{32} = 0.1361$, $p = 0.8926$, Figure 2C). Thus, there was a clear dissociation between ventral and dorsal parietal regions, with ventral regions selectively exhibiting a bias toward stronger stimulus representations during retrieve than encode trials. Notably, when considering univariate effects of instruction, an opposite dissociation was observed: there was a significant difference (retrieve $>$ encode) in dorsal, but not ventral parietal regions (Supplementary Figure 2B).

Memory states are decodable from large-scale attentional networks

We next tested whether memory states (encoding vs. retrieval) are reflected in patterns of activity within three large-scale attentional networks: the frontoparietal control network (FPCN), the dorsal attention network (DAN) and the ventral attention network (VAN)[23]. These attentional networks have previously been shown to carry information about task states[24], to be involved in establishing internal attention[25], and to reflect distinct memory states in broadly distributed patterns of activation[17,18,26]. To test whether the attentional networks carried unique information about intrinsic memory states, we also included the visual network (VisN)[23] as a 'control' network. We predicted that while VisN would be sensitive to superficial differences between

encode versus retrieve trials (in particular, the instruction cue), it would not reflect higher-level, intrinsic memory states[24].

To determine whether networks carried information about memory states, we used cross-validated pattern classification analyses and permutation procedures to establish chance performance (see Methods). Classification accuracy was above chance for each network (FPCN: M = 57.81%, SD = 7.466%, t_{32}= 5.907, $p < 0.001$; DAN: M = 62.58%, SD = 8.618%, t_{32}= 8.278, $p < 0.001$; VAN: M = 54.49%, SD = 4.375%, t_{32}= 5.817, $p < 0.001$; VisN, M = 62.76%, SD = 6.751%, t_{32}= 10.67, $p < 0.001$; Figure 3). Because we did not have *a priori* predictions about differences between the attentional networks, all subsequent analyses combined data across the three attentional networks but included network as a factor. For additional characterization of networks and sub-networks, see Supplementary Figure 3.

Decoded memory states predict cortical location of stimulus representations

We next tested whether memory states decoded from the attentional networks predicted the relative strength of stimulus-specific representations in ventral parietal versus visual regions. Among the ventral parietal regions, here we specifically focused on angular gyrus (AnG) because, in contrast to the other ventral parietal ROIs (vLatIPS, vIPS), AnG was largely non-overlapping with the attentional networks (see Methods). Among the visual regions, we specifically focused on VTC given prior evidence of functional dissociations between VTC and angular gyrus[20,27,28].

We first grouped objects according to evidence derived from the pattern classifiers ('encoding state' or 'retrieval state') irrespective of the actual instruction (encode, retrieve) that objects were associated with (see Methods). A repeated measures ANOVA with factors of cortical location (AnG, VTC), decoded memory state (encoding, retrieval), and attentional network (FPCN, DAN, VAN) revealed a significant two-way interaction between cortical location and decoded memory state ($F_{1,32} = 9.481$, $p = 0.0042$). This interaction was driven by relatively stronger stimulus representations in AnG during retrieval than encoding states and relatively stronger stimulus representations in VTC during encoding than retrieval states (Supplementary Figure 1B). The

interaction between cortical location and decoded memory state did not significantly vary by attentional network (three-way interaction: $F_{2,64} = 2.476$, $p = 0.0921$). Critically, the finding that AnG and VTC representations were differentially influenced by decoded memory states mirrors the effects we observed across ventral parietal and visual regions when considering actual trial instructions (Figure 2B and Supplementary Figure 1A). A significant and qualitatively similar interaction between cortical location and decoded memory state was also observed when memory states were decoded from VisN ($F_{1,32} = 5.841$, $p = 0.0215$).

While the preceding analyses establish a critical relationship between decoded memory states and the cortical location of object representations, the approach does not establish that the classifiers indexed memory states that were *independent from* the actual instructions. In other words, to the extent that the classifiers were generally accurate, then decoded memory states may simply be redundant with trial instructions. However, prior behavioral studies have demonstrated that fluctuations between encoding versus retrieval states can be dissociated from current task demands[15,16] and prior neuroimaging studies have shown that memory states decoded from patterns of neural activity predict behavioral expressions of memory even when controlling for instructions[17,18] or in the absence of instructions altogether[17,19]. Motivated by these prior findings, we next tested whether decoded memory states predicted the cortical location of object representations *even when controlling for the instruction on each trial*. To do this, we again generated encoding state and retrieval state groups, but we did so *within* each instruction condition, thereby dissociating decoded memory states from task instructions (see Figure 3B and Methods).

A repeated measures ANOVA with factors of cortical location (AnG, VTC), decoded memory state (encoding, retrieval), instruction (encode, retrieve), and attentional network (FPCN, DAN, VAN) again revealed a significant two-way interaction between cortical location and decoded memory state ($F_{1,32} = 5.711$, $p = 0.0229$). Qualitatively, the interaction mirrored the pattern from the preceding analysis: stimulus-specific representations in AnG were relatively stronger during retrieval than encoding states whereas representations in VTC were relatively stronger during encoding than retrieval states (Figure 3C, D; Supplementary Figure 1C). The interaction

between cortical location and decoded memory state did not differ according to the attentional network from which memory states were decoded ($F_{2,64} = 0.399$, $p = 0.673$; Figure 3C). Moreover, the interaction between cortical location and decoded memory state also did not differ as a function of actual instructions ($F_{1,32} = 0.26$, $p = 0.613$; Figure 3D). Bayes Factor analysis revealed that a model without instruction as a factor is preferred to a model with instruction by a factor of 2.64. Thus, decoded memory states and instructions each had an influence on shifts in cortical representations, but these influences were additive. Qualitatively, this was particularly evident for stimulus representations in AnG which were numerically highest when the instruction was to retrieve *and* the decoded memory state was retrieval and numerically lowest when the instruction was to encode *and* the decoded memory state was encoding (Supplementary Figure 1D).

Notably, when controlling for instructions, the interaction between cortical location and decoded memory state was not significant when memory states were decoded from VisN ($F_{1,32} = 0.051$, $p = 0.822$). Moreover, an ANOVA with factors of network (attentional vs. VisN), decoded memory state, and cortical location revealed a significant three-way interaction ($F_{1,32} = 7.453$, $p = 0.0102$; Figure 3C). This interaction indicates that shifts in the cortical location of stimulus representations were better predicted by memory states decoded from attentional networks than by memory states decoded from VisN despite the fact that overall decoding accuracy was quite high in VisN.

Medial temporal lobe stimulus representations and connectivity

Although we had an *a priori* interest in stimulus-specific representations in parietal vs. visual regions[11], recent evidence also suggests that representations in the medial temporal lobe (MTL) are stronger during retrieval than during perception[12]. Accordingly, we assessed stimulus-specific representations in two hippocampal regions (CA1; CA23/DG) and three cortical MTL regions [entorhinal cortex (ERC), perirhinal cortex (PRC), parahippocampal cortex (PHC)]. For the hippocampal ROIs, a repeated measures ANOVA with factors of instruction (encode, retrieve) and region (CA1, CA23/DG) revealed a significant main effect of instruction ($F_{1,32} = 5.796$, $p = 0.022$; Figure 4A), with greater stimulus representations during retrieve compared

to encode instruction trials. Similarly, for the MTL cortical ROIs, a repeated measures ANOVA with factors of instruction and region (ERC, PRC, PHC) also revealed a main effect of instruction ($F_{1,32} = 4.726$, $p = 0.0372$; Figure 4A), again with stronger stimulus representations during retrieve compared to encode instruction trials.

Interestingly, instructions did not have any influence on univariate responses in the hippocampus or cortical MTL ROIs, nor did any of these ROIs support successful classification of memory states (Supplementary Figure 4A, B). However, given the putative role of the hippocampus in establishing memory states[15,16,29] and given the proposal that parietal regions exhibit a bias toward representing retrieved content due to their connectivity with the hippocampus[14], we conducted a beta series correlation[30] to test whether connectivity between the hippocampus and cortical targets (visual vs. ventral parietal regions) differed for encode versus retrieve trials. Specifically, we correlated trial-level beta values in CA1 and CA23/DG with each visual and ventral parietal ROI, separately for encode and retrieve instruction trials. A repeated measures ANOVA with factors of hippocampal subfield (CA1, CA23/DG), instruction (encode, retrieve), and cortical target (V1, LO, VTC, vLatIPS, AnG, vIPS) revealed a significant interaction between instruction and cortical target ($F_{5,160} = 3.792$, $p = 0.0028$; Figure 4B). Qualitatively, this interaction was driven by greater connectivity between the hippocampus and ventral parietal regions during retrieve compared to encode trials whereas connectivity with visual regions showed little difference across instruction conditions. A more targeted analysis which averaged across the visual ROIs (V1, LO, VTC) and the ventral parietal ROIs (vLatIPS, AnG, vIPS) again revealed an interaction between instruction and cortical target ($F_{1,32} = 5.467$, $p = 0.0258$), with ventral parietal regions exhibiting significantly greater connectivity with the hippocampus during retrieve than encode trials ($t_{32} = 2.2634$, $p = 0.0305$) while connectivity between visual regions and the hippocampus did not significantly differ for encode versus retrieve trials ($t_{32} = 0.1154$, $p = 0.9082$). Finally, considering AnG, specifically, connectivity with the hippocampus was markedly stronger during retrieve than encode trials ($t_{32} = 3.1896$, $p = 0.0032$).

Discussion

Here we show that the cortical representations of visual objects shift with changes in memory states. This effect was most pronounced in ventral parietal cortex, where stimulus-specific representations were markedly stronger during memory retrieval compared to memory encoding. The shift in content representations was evident when comparing stimulus representations as a function of task instructions (encode versus retrieve) despite a lack of perceptual or behavioral differences between these trials. Moreover–and critically–the same shift was also evident when comparing stimulus representations as a function of memory states (encoding, retrieval) that were decoded from large-scale attentional networks–even when decoded memory states were entirely independent from task instructions.

The fact that ventral parietal cortex contained content representations during memory retrieval but not during memory encoding (Figure 2B) is a striking aspect of our findings. This finding bears a strong resemblance to recent evidence that rodent parietal cortex preferentially represents sensory information from past environmental states over sensory information from current environmental states[10]. Importantly, the idea that some brain regions preferentially, or even selectively, represent past experiences over present experience is not accounted for by the phenomenon of reactivation which, by definition, explains representations at retrieval as a re-expression of encoding-related activity patterns[1–6]. While our findings do not argue against the idea that reactivation occurs – and our experimental design did not directly test for reactivation – our findings support the emerging idea of systematic differences in the cortical location of content representations during memory encoding versus memory retrieval[7–14]. In particular, our findings are consistent with recent arguments of a spatial transformation wherein representations initially encoded by visual cortex are systematically re-expressed in parietal cortex during memory retrieval[9–11]. More broadly, our findings are also consistent with the idea that angular gyrus plays an important role in processing internally-generated information[31], which may include memories[11,28], thoughts[32] or even simulations of future events[33]. However, whereas traditional views proposed that these contributions reflect content-general processes[34,35], our findings underscore that angular gyrus maintains detailed (stimulus-specific) representations of

internally-generated content[11,28,36].

Why does ventral parietal cortex preferentially represent content retrieved from memory as compared to content encoded from the current environment? Although there are multiple potential accounts of such biases (for detailed consideration of these accounts, see[14]), one account that is particularly consistent with the present results is that a bias toward retrieved content is the result of strong connectivity with–or drive from–the hippocampus[7,14]. This account is motivated by evidence that the default mode network, of which angular gyrus is a core component, is functionally coupled with the hippocampus[37,38], particularly during memory retrieval[39,40]. Here, we provide direct and unique support for this account by demonstrating, within a single experimental paradigm, that (a) patterns of activity in angular gyrus exhibited a bias toward representing retrieved content over encoded content and (b) evoked responses in angular gyrus were more strongly correlated with responses in the hippocampus during memory retrieval than during memory encoding. More generally, given the diversity of connections between the hippocampus and neocortex[41], and given the putative role of the hippocampus in establishing biases between encoding versus retrieval states[42–44], the hippocampus is well positioned to mediate transformations in the cortical expressions of mnemonic content.

In designing our experimental paradigm, a point of emphasis was to minimize differences between encode and retrieve trials so as to constrain potential accounts of why content representations might shift across these conditions. In particular, to support our argument that content representations shift with intrinsic memory states, it was important to rule out the possibility that the shift in content representations was an artifact of differences in the specific *tasks* that were used. To this end, encode and retrieve trials were perceptually matched (cf.[9,11]) and neither trial type required a behavioral response. The lack of behavioral response is notable in that retrieval tasks often involve a stronger decision-making component than encoding tasks, which could explain a greater involvement of parietal cortex during retrieval[4,45,46]. Even in the absence of behavioral responses, however, it is still possible that retrieve trials were more effortful than encode trials. Yet, it is notable that effortful memory retrieval decisions have specifically been associated with dorsal parietal regions and not ventral parietal regions[21],

whereas the bias toward retrieved content that we observed was significantly stronger and selectively present in ventral parietal regions. However, the most compelling evidence in support of an account based on intrinsic memory states comes from our use of pattern classification analyses to index fluctuations between encoding and retrieval states. Strikingly, the pattern of results we observed when memory states were decoded from attentional networks–even when these states were fully independent from task instructions–was qualitatively identical to the pattern of results we observed when considering explicit trial instructions (Figure 3D). The fact that decoded memory states predicted shifts in content representations that were independent of task instructions provides critical support for our argument that shifts in the cortical location of content representations were not a product of differences in encoding versus retrieval tasks, but of differences in intrinsic memory states. This point is important and relevant when considering that biases toward memory encoding versus retrieval are often independent of any explicit task demands[15,16].

The fact that attentional networks, in particular, carried information about intrinsic memory states is notable for several reasons. First, this finding complements recent evidence that internal attention specifically involves interactions between attentional networks and the default mode network (of which angular gyrus is a core component)[25,47]. Second, although we did not observe differences across the attentional networks (FPCN, DAN, VAN) in terms of the degree to which these networks predicted shifts in the cortical location of stimulus representations (Figure 3C), we did observe a statistical dissociation between the attentional networks and the visual network. Namely, memory states decoded from the visual network (when controlling for trial instructions) did not predict cortical shifts in content representations (Figure 3C). Our interpretation is that although the visual network supported robust classification of encode versus retrieve trials (Figure 3A), this classification was at least partly driven by superficial differences between the conditions (e.g., the visual word form of the instruction cue). In contrast, activity patterns in the attentional networks were, putatively, less sensitive to superficial differences between encode versus retrieve trials and more sensitive to intrinsic memory states. Third, and relatedly, the fact that memory states decoded from attentional networks predicted cortical shifts in content representations that were independent from actual trial instructions demonstrates that

these networks were not simply tracking explicit task demands. This point is important in light of evidence, in other contexts, that memory states decoded from frontoparietal regions can be largely driven by explicit task demands[26].

Taken together, our findings indicate that the cortical location of content representations is fundamentally determined by whether attention is internally oriented to memories or externally oriented to current perceptual experience. These findings have important implications for understanding how the memory system orchestrates the encoding of new experience with the retrieval of past experience.

Acknowledgments

This work was supported by the Lewis Family Endowment to the University of Oregon, which supports the Robert and Beverly Lewis Center for NeuroImaging and by grants from the National Institutes of Health (NINDS R01 NS107727, PI: B.A.K.) and National Science Foundation (CAREER BCS-1752921, PI: B.A.K.). We thank Alex Tremblay-McGaw for assistance with data collection.

Author contributions

N.M.L and B.A.K. designed the experiments. N.M.L. ran the experiments. N.M.L. analyzed the data. N.M.L. and B.A.K. wrote the paper.

Declaration of Interests

The authors declare no competing interests.

Figure 1. Task Design. During List 1, subjects studied images of individual objects (e.g. bench, fan). All List 1 objects were studied four times, across four unscanned runs. During List 2, subjects saw novel objects that were from the same categories as the items shown in List 1 (e.g., a new bench, a new fan). Preceding each List 2 object was an instruction cue: either "NEW" or "OLD." The NEW cue signaled that subjects were to *encode* the current item (e.g., the new bench). The OLD cue signaled that subjects were to *retrieve* the corresponding item from List 1 (e.g., the old fan). We refer to these two trial types as 'encode instruction' and 'retrieve instruction,' respectively. Each List 2 object was presented twice in each of eight scanned runs. The instruction cue associated with each List 2 object remained consistent throughout the experiment (always encode or always retrieve).

Figure 2. Stimulus-specific representations across cortical regions as a function of instruction. **(A)** Object representations were indexed by performing Pearson correlations between spatial patterns of activity (beta values) from odd and even scan runs. Stimulus-specific representations were calculated by subtracting the mean correlation between different objects (across-object correlations) from the mean correlation between the same objects (within-object correlations). Across-object correlations were always performed within instruction condition. **(B)** Stimulus-specific representations for each cortical location as a function of instruction condition (encode = orange, retrieve = teal). **(C)** Difference in the strength of stimulus-specific representations for encode vs. retrieve instruction trials, for three broad cortical regions (visual, dorsal parietal, ventral parietal). Values toward the left (< 0) reflect stronger stimulus representations during retrieve trials; values toward the right (> 0) reflect stronger stimulus representations during encode trials. Stimulus representations in ventral parietal regions were significantly stronger during retrieve than encode trials and exhibited a significantly stronger bias toward retrieval than did visual regions (ventral parietal vs. visual: p = 0.0183) or dorsal parietal regions (ventral parietal vs. dorsal parietal: p = 0.0046). * $p < 0.05$. Error bars are standard error of the mean. See also Supplementary Figure S2.

Figure 3. Memory state decoding and stimulus-specific representations. (A) Cross-validated classification of encode vs. retrieve instruction trials in three attentional networks (FPCN, DAN, VAN) and the visual network (VisN). Decoding was significantly above chance, as determined by permutation procedures, in each network. **(B)** To assess stimulus-specific representations as a function of decoded state (encoding, retrieval) while controlling for actual instruction (encode, retrieve), the twelve objects within each instruction condition were median-split into 'encoding state' and 'retrieval state' groups according to the relative strength of classifier evidence (see Methods). Note: although illustrated separately here, encode instruction trials (top) and retrieve instruction trials (bottom) were randomly intermixed in the experiment. **(C)** Biases in stimulus-specific representations (encoding state $>$ retrieval state) in AnG and VTC as a function of the network from which memory states were decoded, controlling for instruction condition as described in **(B)**. A shift in stimulus representations from VTC (relatively stronger during encoding state) to AnG (relatively stronger during retrieval state) was observed when memory states were decoded from the attentional networks (FPCN, DAN, VAN), but not when they were decoded from the visual network (VisN). **(D)** Stimulus-specific representations for encoding states $>$ retrieval states are shown averaged across the three attentional networks (FPCN, DAN, VAN) and are separated by actual instruction (encode, orange; retrieve, green). We found a significant interaction between cortical location (AnG, VTC) and decoded memory state ($p = 0.0229$), indicating that the cortical location of object representations was predicted by memory states decoded from attentional networks. Error bars are standard error of the mean. *** $p < 0.001$. See also Supplementary Figures S1 and S3.

Figure 4. Medial temporal lobe stimulus representations and connectivity. (A) Difference in the strength of stimulus-specific representations for encode vs. retrieve instruction trials in the hippocampus (CA1, CA23/DG) and medial temporal lobe cortical regions (entorhinal cortex, ERC; perirhinal cortex, PRC; parahippocampal cortex, PHC). Values toward the left (< 0) reflect stronger stimulus representations during retrieve trials; values toward the right (> 0) reflect stronger stimulus representations during encode trials. For the hippocampus and medial temporal lobe cortical regions, stimulus representations were significantly stronger during retrieve than encode trials. **(B)** Difference in correlation (encode - retrieve trials) between trial-level univariate responses in the hippocampus (CA1 and CA23DG) and in visual and ventral parietal regions. Values toward the left (<0) reflect stronger correlations with the hippocampus during retrieve than encode trials. Values toward the right (> 0) reflect stronger correlations with the hippocampus during encode than retrieve trials. Error bars are standard error of the mean. See also Supplementary Figure S4.

STAR Methods

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Nicole Long (niclong@virginia.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The raw, de-identified data and the associated experimental and analysis codes used in this study can be accessed via the Open Science Foundation (https://osf.io/fdnh7/).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects

35 (22 female; mean age = 21 years, range: 18 - 28 years) right-handed, native English speakers from the University of Oregon community participated. All subjects had normal or corrected-to-normal vision. Informed consent was obtained in accordance with the University of Oregon Institutional Review Board. Two subjects were excluded from the final dataset for excessive head motion during scanning. Thus, data are reported for the remaining 33 subjects.

METHOD DETAILS

Stimuli and Design

Overview. The experiment was divided into three phases: List 1, List 2, and Recognition Test. List 1 and List 2 were completed while subjects were in the MRI scanner and the Recognition Test was completed after subjects exited the scanner.

Stimuli. Stimuli consisted of 96 object images drawn from a database of categorized images[48]. Four exemplars were drawn from each of 24 object categories (e.g., 4 images of benches). For each subject, one exemplar from each object category served as a List 1 item, one as a List 2 item, and the two remaining exemplars served as lures for the recognition phase. Object condition assignment was randomly generated for each subject.

List 1. On each trial, subjects saw a single object presented for 2000 ms followed by a 500 ms inter-stimulus interval (ISI; Figure 1). Subjects were instructed to study the presented object in anticipation for a later memory test. Subjects completed 4 runs of List 1 trials with 24 objects per run, yielding a total of 96 List 1 trials. List 1 was completed during the high-resolution anatomical T1 scan (see below).

List 2. Each List 2 trial began with an instruction cue – either "OLD" or "NEW" – which was presented for 1000 ms. The cue was followed by an object image for 2000 ms. All objects in List 2 were non-identical exemplars drawn from identical categories as the objects presented in the immediately preceding List 1. For example, if a subject saw a bench and a fan during List 1, a different bench and a different fan would be presented during List 2. On trials with a NEW instruction subjects were asked to encode the presented object. We refer to these trials as 'encode instruction' trials throughout the manuscript. On trials with an OLD instruction, subjects were asked to retrieve the categorically-identical item from the preceding List 1 (e.g.,

the previously-encoded fan). We refer to these trials as 'retrieve instruction' trials throughout the manuscript. Importantly, the design prevented subjects from completely ignoring List 2 items following retrieve instructions because subjects could only identify the to-be-retrieved object by processing the current List 2 item. Note: neither the encode or retrieve instructions required a behavioral response from subjects. After each object image, there was a 5000 ms ISI during which subjects made odd/even judgments, via button press, for two individually-presented single-digit numbers.

The List 2 trials were distributed across 8 fMRI scan runs with 2 presentations of each of the 24 List 2 objects in each run, yielding a grand total of 16 repetitions for each List 2 object. Within each run, each of the 24 objects was presented once, in random order, before the 2nd presentation of any of the objects (which was also in random order). The instruction cue associated with each object was fixed across the entire List 2 phase (e.g. bench would always be presented with the encode instruction and fan would always be presented with the retrieve instruction). Of the 24 List 2 objects, half were associated with a retrieve instruction and half with an encode instruction. Object-instruction pairings were randomly assigned for each subject.

Recognition Test. After exiting the scanner, subjects completed a forced-choice recognition test. On each trial, subjects saw two exemplars from the same object category (e.g. two benches). One object had previously been encountered either during List 1 or 2. The other object was a lure and had not been presented during the experiment. Subjects selected (via mouse click) the previously presented object. Subjects had 4000 ms to respond and all responses were made within this time window. Trials were separated by a 1000 ms ISI. There were a total of 48 recognition trials (corresponding to the 24 List 1 items and 24 List 2 items). Note: List 1 and List 2 items never appeared in the same trial together, thus subjects never had to choose between two previously presented items. List 1 and List 2 items were randomly intermixed in the recognition test. Accuracy on the recognition test was 100% for every subject and is therefore not analyzed further.

fMRI Data Acquisition

Imaging data were collected on a Siemens 3T Skyra scanner at the Robert and Beverly Lewis Center for NeuroImaging at the University of Oregon. Before the functional imaging, a whole-brain high-resolution anatomical scan was conducted for each subject using a T1-weighted protocol (grid size 256 $\times$ 256; 176 sagittal slices; voxel size 1 $\times$ 1 $\times$ 1 mm). Next, a custom anatomical T2 coronal scan was conducted for each subject (TR = 13,520 ms; TE = 88 ms; flip angle = 150°; grid size 512 $\times$ 512; 65 contiguous slices oriented perpendicularly to the main axis of the hippocampus; interleaved acquisition; FOV=220 mm; voxel size=0.4$\times$0.4$\times$2 mm; GRAPPA factor=2;[49]). Prior to the List 2 phase, a resting state scan was conducted (not analyzed here). Functional images were collected using a T2*-weighted multi-band accelerated EPI sequence (TR = 2s; TE = 36ms; flip angle = 90°; grid size 124 $\times$ 124; 72 contiguous slices oriented parallel to the hippocampus; voxel size 1.7 $\times$ 1.7 $\times$ 1.7 mm). Note: because our primary focus was on visual and parietal cortices, we used a high-resolution protocol for functional images that prioritized spatial resolution in posterior regions over whole brain coverage. In particular, for most subjects a small portion of superior frontal cortex was omitted from the functional scans. Eight functional scans were collected, each consisting of 198 volumes. Following the eight functional scans, we collected a second resting state scan (not analyzed here).

fMRI Data Analysis

All fMRI preprocessing was performed using *fMRIPrep* 1.4.0[50,51], which is based on *Nipype* 1.2.0[52,53].

Anatomical data preprocessing. The T1-weighted (T1w) image was corrected for intensity non-uniformity with *N4BiasFieldCorrection*[54], distributed with ANTs 2.2.0[55], and used as the T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the *antsBrainExtraction.sh* workflow (from ANTs), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF),

white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using *fast* (FSL 5.0.9,[56]). Brain surfaces were reconstructed using *recon-all* (FreeSurfer 6.0.1,[57]) and the brain mask generated from the T1w-reference was further refined using a custom variation of Mindboggle's method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter[58]. Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with *antsRegistration*, using brain-extracted versions of both the T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* ([59]; TemplateFlow ID: MNI152NLin2009cAsym).

Functional data preprocessing. For each of the 8 functional runs per subject, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD reference was then co-registered to the T1w reference using *bbregister* (FreeSurfer). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using *mcflirt* (FSL,[60]).

BOLD runs were slice-time corrected using *3dTshift* from AFNI 20160207[61]. The BOLD time-series were resampled to surfaces on the following spaces: *fsnative*, *fsaverage*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS were calculated for each functional run using their implementations in *Nipype* (following the definitions by[62]). The three global signals were extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise

correction (*CompCor*,[63]). Principal components were estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components were then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask was obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components were calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components were also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values were retained, such that the retained components' time series were sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components were dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each[64]. All resamplings were performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using *antsApplyTransforms* (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels[65]. Non-gridded (surface) resamplings were performed using *mri_vol2surf* (FreeSurfer).

Network and region of interest selection

We decoded memory states in four previously-defined resting-state networks[23]. We focused on three attentional networks (frontoparietal control network, FPCN; dorsal attention network, DAN; ventral attention network, VAN) and the visual network (VisN). These networks have been shown to represent both stimulus features and task goals[24]. The resting-state networks were generated for each subject using their high-resolution anatomical image and the FreeSurfer

cortical parcellation scheme (http://surfer.nmr.mgh.harvard.edu). The networks were then co-registered to the functional data.

We assessed stimulus representations in eight regions of interest (ROIs): V1, lateral occipital cortex (LO), ventral temporal cortex (VTC), posterior intraparietal sulcus (pIPS), dorsolateral intraparietal sulcus (dLatIPS), ventral intraparietal sulcus (vIPS), angular gyrus (AnG), and ventrolateral intraparietal sulcus (vLatIPS). These ROIs were generated in each subject's native space, following procedures described in our previous work[11]. Of particular note, the AnG ROI was comprised of subcomponents of the default mode network, which was not one of the networks from which memory states were decoded.

As in our prior study[11], for some analyses we grouped the eight ROIs according to three broad cortical regions: visual (V1, LO, VTC), dorsal parietal (pIPS, dLatIPS), and ventral parietal (vIPS, AnG, vLatIPS). ROIs were first defined on the FreeSurfer average cortical surface and then reverse-normalized to each subjects' native anatomical surface. They were then projected into the volume at the resolution of the functional data to produce binary masks.

For analyses that measured the cortical location of stimulus representations as a function of decoded memory states, we focused on two specific cortical regions: AnG and VTC. We selected these regions based on our previous work contrasting these specific regions[20,27,28]. These regions were also minimally overlapping with the attentional networks from which memory states were decoded. For AnG, 3.0% of the voxels in the ROI overlapped with FPCN, 6.5% with DAN, and 0.06% with VAN. For VTC, 0.005% of the voxels overlapped with FPCN, 21.0% with DAN, and 0% with VAN.

Using the automatic segmentation of hippocampal subfields (ASHS) machine learning toolbox[66] applied to the T2 images, we extracted two hippocampal ROIs (CA1, CA23/DG) and three extra-hippocampal medial temporal lobe ROIs (entorhinal cortex, ERC; perirhinal cortex, PRC;

parahippocampal cortex, PHC).

Univariate Analyses

Univariate data analyses were conducted under the assumptions of the general linear model (GLM) using SPM12 (http://www.fil.ion.ucl.ac.uk/spm). The model included regressors for every List 2 trial (N = 384) as well as regressors for scan run and six motion parameters for each run. Resulting trial-level beta values were used for both pattern similarity and pattern classification analyses.

Pattern Similarity Analyses

Stimulus-specific representations were measured by pattern similarity analyses, following the general procedures from a prior study[11]. From the GLM, we obtained a beta value for every presentation (16 each) of the 24 List 2 items. These betas were separated according to odd vs. even run numbers. We then averaged the betas for each object across all the even runs and, separately, across all the odd runs. This resulted in a single, mean beta value for each object for the odd runs and the even runs. We then computed the Fisher *z*-transformed Pearson correlation between the spatial pattern of beta values, within a given ROI, for each pair of objects across odd and even runs (i.e., a correlation matrix). Correlations between the odd/even runs were divided into two groups: within-stimulus correlations and across-stimulus correlations. Within-stimulus correlations ('on-diagonal correlations') refer to correlations between the same stimulus [e.g., *r*(odd run bench, even run bench)]. Across-stimulus correlations ('off-diagonal correlations') refer to correlations between different stimuli [e.g., *r*(odd run bench, even run suitcase)]. Importantly, across-stimulus correlations were restricted to objects from the same instruction condition (encode or retrieve). The average across-stimulus correlation within each instruction condition functioned as a baseline and was subtracted from the average within-stimulus correlation within each instruction condition to produce a measure of stimulus-specific information. Values greater than 0 constituted positive evidence for stimulus-specific representations. Stimulus-specific in-

formation was separately computed for each subject, ROI, and instruction cue (encode, retrieve).

Pattern classification analyses

Pattern classification analyses were performed using penalized (L2) logistic regression (penalty parameter = 1), implemented via the sklearn module in Python and custom Python code. For each subject, and for individual networks/ROIs, leave-one-run-out cross validation was performed, using the 8 List 2 scan runs, to test whether encode/retrieve instruction trials could be reliably decoded. Classifiers were trained/tested using the spatial pattern of beta values within a specified network or ROI and classification was performed on a trial-by-trial basis. Classifier performance was assessed in two ways: classification accuracy and classifier evidence. Classification accuracy represents the percentage of trials for which the correct label (encode or retrieve) was assigned. Classification accuracy was used for general assessment of classifier performance (i.e., whether instructions could be decoded). Classifier evidence was a continuous value reflecting the logit-transformed probability that the classifier assigned to the correct label for each trial. Classifier evidence was used as a trial-specific, continuous measure of memory states, and was only used for testing whether stimulus-specific representations varied as a function of decoded memory states (as described in the following section).

Stimulus-specific representations as a function of decoded memory states

To test whether stimulus-specific representations (as measured by pattern similarity analyses) varied as a function of decoded memory states (indexed by pattern classifiers), we divided the 24 objects into two groups according to the strength of classifier evidence (for an encoding vs. retrieval state). This was performed in two ways. First, all of the 24 objects (regardless of whether they were presented with an encode or retrieve instruction) were median split according to the mean classifier evidence for an encoding state that they generated (averaged across the 16 presentations of each object). Objects with mean encoding evidence greater than the median were comprised the ‘encoding state’ group and objects with mean evidence less than

the median comprised the 'retrieval state' group. Note: the rationale for using a median-split of classifier evidence (as opposed to using the categorical label generated by the classifier) is that the median-split approach ensured an equal number of trials in the encoding state and retrieval state groups.

The second way in which we performed this analysis was identical except that the median split was performed within each instruction condition (encode, retrieve) in order to control for the actual instruction on each trial. Specifically, the 12 objects that were presented with an encode instruction were median-split into 'encoding state' and 'retrieval state' groups according to the relative strength of classifier evidence for an encoding state (above or below the median, respectively). Likewise, the 12 objects presented with a retrieve instruction were also median-split into 'encoding state' and 'retrieval state' groups. The encoding state groups were then averaged across the two instruction conditions to create a single encoding state group. Likewise, the retrieval state groups were averaged across the two instruction conditions to create a single retrieval state group. The critical feature of this analysis approach is that the objects assigned to each group contained an equal number of trials from each instruction condition. Thus, the encoding state and retrieval state groups differed with respect to the relative strength of classifier evidence for encoding vs. retrieval, but they did not differ with respect to the proportion of trials associated with an encode vs. retrieve instruction.

For both versions of this analyses, within-stimulus and across-stimulus pattern similarity analyses were performed for each group (encoding state and retrieval state), following the same procedures described above (see Pattern Similarity Analyses). This yielded a measure of stimulus-specific representations (within-stimulus $-$ across-stimulus) as a function of decoded memory state.

Cross-region correlation analyses

To test whether activation in hippocampal regions was differentially correlated with activation in visual and parietal cortical regions as a function of encode versus retrieve instructions, we extracted trial-level univariate beta values from CA1 and CA23/DG and from the three visual cortical ROIs (V1, LO, VTC) and the three ventral parietal ROIS (vIPS, AnG, vLatIPS). The beta values from each of the hippocampal ROIs were then correlated with beta values from each visual/parietal ROI, separately for encode and retrieve trials. The resulting Pearson's *r* values were Fisher Z transformed and subtracted (encode - retrieve), resulting in a single zRho value for each hippocampal-cortical ROI pair and for each subject.

QUANTIFICATION AND STATISTICAL ANALYSIS

Repeated measures ANOVAs were used to assess stimulus-specific representations as a function of stimulus location (AnG, VTC) and as a function of either instruction or decoded state (encoding, retrieval). Paired samples *t*-tests were also used for follow-up analyses of these data. Classification accuracy was compared to chance performance using permutation tests. Specifically, for each subject and network/ROI, the condition labels (encode or retrieve) were shuffled and classification accuracy was then computed. This was repeated 1000 times for each subject and network/ROI. The mean of these 1000 values was used as an empirically-derived, subject-specific measure of chance performance. Paired samples *t*-tests were used to compare the true (unshuffled) accuracies to the means of the shuffled accuracies.

References

[1] Wheeler, M. E., Petersen, S. E., and Buckner, R. L. Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20):11125–11129, 2000.

[2] Polyn, S. M., Natu, V. S., Cohen, J. D., and Norman, K. A. Category-specific cortical activity precedes retrieval during memory search. *Science*, 310:1963–1966, 2005.

[3] Danker, J. F. and Anderson, J. R. The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, 136(1):87, 2010.

[4] Kuhl, B. A., Rissman, J., Chun, M., and Wagner, A. Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences of the United States of America*, 108(14):5903–5908, 2011.

[5] Rissman, J. and Wagner, A. D. Distributed representations in memory: Insights from functional brain imaging. *Annual Review of Psychology*, 63:101–128, 2012.

[6] Levy, B. J. and Wagner, A. D. Measuring memory reactivation with functional mri implications for psychological theory. *Perspectives on Psychological Science*, 8(1):72–78, 2013.

[7] Baldassano, C., Esteva, A., Fei-Fei, L., and Beck, D. M. Two distinct scene-processing networks connecting vision and memory. *eNeuro*, 3(5), 2016.

[8] Long, N. M., Lee, H., and Kuhl, B. A. Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *Journal of Neuroscience*, 36(50):12677–12687, 2016.

[9] Xiao, X., Dong, Q., Gao, J., Men, W., Poldrack, R. A., and Xue, G. Transformed neural pattern reinstatement during episodic memory retrieval. *Journal of Neuroscience*, 37(11):2986–2998, 2017.

[10] Akrami, A., Kopec, C. D., Diamond, M. E., and Brody, C. D. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, 554:368–372, 2018.

[11] Favila, S. E., Samide, R., Sweigart, S. C., and Kuhl, B. A. Parietal representations of stimulus features are amplified during memory retrieval and flexibly aligned with top-down goals. *Journal of Neuroscience*, 38(36):7809–7821, 2018.

[12] Lee, S.-H., Kravitz, D. J., and Baker, C. I. Differential representations of perceived and retrieved visual information in hippocampus and cortex. *Cerebral Cortex*, 29(10):4452–4461, 2019.

[13] Silson, E. H., Gilmore, A. W., Kalinowski, S. E., Steel, A., Kidder, A., Martin, A., and Baker, C. I. A posterior–anterior distinction between scene perception and scene construction in human medial parietal cortex. *Journal of Neuroscience*, 39(4):705–717, 2019.

[14] Favila, S. E., Lee, H., and Kuhl, B. A. Transforming the concept of memory reactivation. *Trends in Neurosciences*, 43(12), 2020.

[15] Duncan, K., Sadanand, A., and Davachi, L. Memory's penumbra: Episodic memory decisions induce lingering mnemonic biases. *Science*, 337(6093):485–487, 2012.

[16] Patil, A. and Duncan, K. Lingering cognitive states shape fundamental mnemonic abilities. *Psychological Science*, 29(1):45–55, 2018.

[17] Richter, F. R., Chanales, A. J., and Kuhl, B. A. Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *NeuroImage*, 124:323–335, 2016.

[18] Long, N. M. and Kuhl, B. A. Decoding the tradeoff between encoding and retrieval to predict memory for overlapping events. *NeuroImage*, 201, 2019.

[19] Chanales, A. J. H., Dudukovic, N. M., Richter, F. R., and Kuhl, B. A. Interference between overlapping memories is predicted by neural states during learning. *Nature Communications*, 10(5363), 2019.

[20] Lee, H., Samide, R., Richter, F. R., and Kuhl, B. A. Decomposing parietal memory reactivation to predict consequences of remembering. *Cerebral Cortex*, 29(8):3305–3318, 2019.

[21] Hutchinson, J. B., Uncapher, M. R., Weiner, K. S., Bressler, D. W., Silver, M. A., Preston, A. R., and Wagner, A. D. Functional heterogeneity in posterior parietal cortex across attention and episodic memory retrieval. *Cerebral Cortex*, 24(1):49–66, 2014.

[22] Sestieri, C., Shulman, G. L., and Corbetta, M. The contribution of the human posterior parietal cortex to episodic memory. *Nature Reviews Neuroscience*, 18(3):183–192, 2017.

[23] Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165, 2011.

[24] Long, N. M. and Kuhl, B. A. Bottom-up and top-down factors differentially influence stimulus representations across large-scale attentional networks. *Journal of Neuroscience*, 38(10):2495–2504, 2018.

[25] Kam, J. W. Y., Lin, J. J., Solbakk, A.-K., Endestad, T., Larsson, P. G., and Knight, R. T. Default network and frontoparietal control network theta connectivity supports internal attention. *Nature Human Behavior*, 3:1263–1270, 2019.

[26] Uncapher, M. R., Boyd-Meredith, J. T., Chow, T. E., Rissman, J., and Wagner, A. D. Goal-directed modulation of neural memory patterns: Implications for fMRI-based memory detection. *The Journal of Neuroscience*, 35(22):8531–8545, 2015.

[27] Lee, H., Chun, M. M., and Kuhl, B. A. Lower parietal encoding activation is associated with sharper information and better memory. *Cerebral Cortex*, 27(4):2486–2499, 2017.

[28] Kuhl, B. A. and Chun, M. M. Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *The Journal Of Neuroscience*, 34(23):8051–8060, 2014.

[29] Duncan, K., Tompary, A., and Davachi, L. Associative encoding and retrieval are predicted by functional connectivity in distinct hippocampal area ca1 pathways. *The Journal of Neuroscience*, 34(34):11188–11198, 2014.

[30] Rissman, J., Gazzaley, A., and D'Esposito, M. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, 23:752–763, 2004.

[31] Wagner, A., Shannon, B., Kahn, I., and Buckner, R. Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Science*, 9(9):445–453, 2005.

[32] Buckner, R. L. and DiNicola, L. M. The brain's default network: updated anatomy, physiology and evolving insights. *Nature Neuroscience*, 20:593–608, 2019.

[33] Thakral, P. P., Madore, K. P., and Schacter, D. L. A role for the left angular gyrus in episodic simulation and memory. *Journal of Neuroscience*, 37(34):8142–8149, 2017.

[34] Buckner, R. L. and Wheeler, M. E. The cognitive neuroscience of remembering. *Nature Reviews Neuroscience*, 2(9):624–634, 2001.

[35] Cabeza, R., Ciaramelli, E., Olson, I. R., and Moscovitch, M. The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, 9(8):613–625, 2008.

[36] Ester, E. F., Sprague, T. C., and Serences, J. T. Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*, 87(4):893–905, 2015.

[37] Kahn, I., Andrews-Hanna, J. R., Vincent, J. L., Snyder, A. Z., and Buckner, R. L. Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, 100:129–139, 2008.

[38] Ritchey, M. and Cooper, R. A. Deconstructing the posterior medial episodic network. *Trends in Cognitive Sciences*, 24(6):451–465, 2020.

[39] Huijbers, W., Pennartz, C. M., Cabeza, R., and Daselaar, S. M. The hippocampus is coupled with the default network during memory retrieval but not during memory encoding. *PLoS One*, 6(4):e17463, 2011.

[40] Higgins, C., Liu, Y., Vidaurre, D., Kurth-Nelson, Z., Dolan, R., Behrens, T. E. J., and Woolrich, M. W. Replay bursts coincide with activation of the default mode and parietal alpha network. 2020.

[41] Lavenex, P. and Amaral, D. G. Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus*, 10:420–430, 2000.

[42] Hasselmo, M. E., Schnell, E., and Barkai, E. Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region ca3. *Journal of Neuroscience*, 15:5249–5262, 1995.

[43] Bein, O., Duncan, K., and Davachi, L. Mnemonic prediction errors bias hippocampal states. *Nature Communications*, 11(3451), 2020.

[44] Tarder-Stoll, H., Jayakumar, M., Dimsdale-Zucker, H. R., Günseli, E., and Aly, M. Dynamic internal states shape memory retrieval. *Neuropsychologia*, 138, 2020.

[45] Gonzalez, A., Hutchinson, J. B., Uncapher, M. R., Chen, J., LaRocque, K. F., Foster, B. L., Rangarajan, V., Parvizi, J., and Wagner, A. D. Electrocorticography reveals the temporal dynamics of posterior parietal cortical activity during recognition memory decisions. *Proceedings of the National Academy of Sciences*, 112(35):11066–11071, 2015.

[46] Tibon, R., Fuhrmann, D., Levy, D. A., Simons, J. S., and Henson, R. N. A. Multimodal integration and vividness in the angular gyrus during episodic encoding and retrieval. *Journal of Neuroscience*, 39(22):4365–4374, 2019.

[47] Fornito, A., Harrison, B. J., Zalesky, A., and Simons, J. S. Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceedings of the National Academy of Sciences*, 109(31):12788–12793, 2012.

[48] Konkle, T., Brady, T. F., Alvarez, G. A., and Oliva, A. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3):558, 2010.

[49] Bowman, C. R. and Zeithamova, D. Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *Journal of Neuroscience*, 38(10):2605–2614, 2018.

[50] Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16:111–116, 2019.

[51] Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, James D. andGoncalves, M., DuPre, E., et al. fmriprep. Technical report, 2019.

[52] Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., and Ghosh, S. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, 5(13), 2011.

[53] Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., et al. Nipype. Technical report, 2019.

[54] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010.

[55] Avants, B., Epstein, C., Grossman, M., and Gee, J. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.

[56] Zhang, Y., Brady, M., and Smith, S. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.

[57] Dale, A. M., Fischl, B., and Sereno, M. Cortical surface-based analysis I: Segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194, 1999.

[58] Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., et al. Mindboggling morphometry of human brains. *PLOS Computational Biology*, 13(2), 2017.

[59] Fonov, V., Evans, A., McKinstry, R., Almli, C., and Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, Supplement 1:S102, 2009.

[60] Jenkinson, M., Bannister, P., Brady, M., and Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841, 2002.

[61] Cox, R. W. and Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10:171–178, 1997.

[62] Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. Methods to detect, characterize, and remove motion artifact in resting state fmri. *NeuroImage*, 84:320–341, 2014.

[63] Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fmri. *NeuroImage*, 37:90–101, 2007.

[64] Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64:240–256, 2013.

[65] Lanczos, C. Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, 1(1):76–85, 1964.

[66] Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S.-L., Gertje, E. C., Mancuso, L. E., Kliot, D., Das, S. R., and Wolka, D. A. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, 36:258–287, 2014.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw data, experiment codes, analysis codes	Open Science Foundation	https://osf.io/fdnh7/
Software and Algorithms		
fMRIPrep 1.4.0	[50,51]	SCR_016216
Nipype 1.2.0	[52,53]	SCR_002502
ANTs 2.2.0	[55]	SCR_004757
FSL 5.0.9	[56]	SCR_002823
FreeSurfer 6.0.1	[57]	SCR_001847
MindBoggle	[58]	SCR_002438
ICBM 152 Nonlinear Asymmetrical template version 2009c	[59]	SCR_008796
AFNI 20160207	[61]	SCR_005927
SPM 12	Wellcome Department of Cognitive Neurology, London, United Kingdom	https://www.fil.ion.ucl.ac.uk/spm/

Figure1

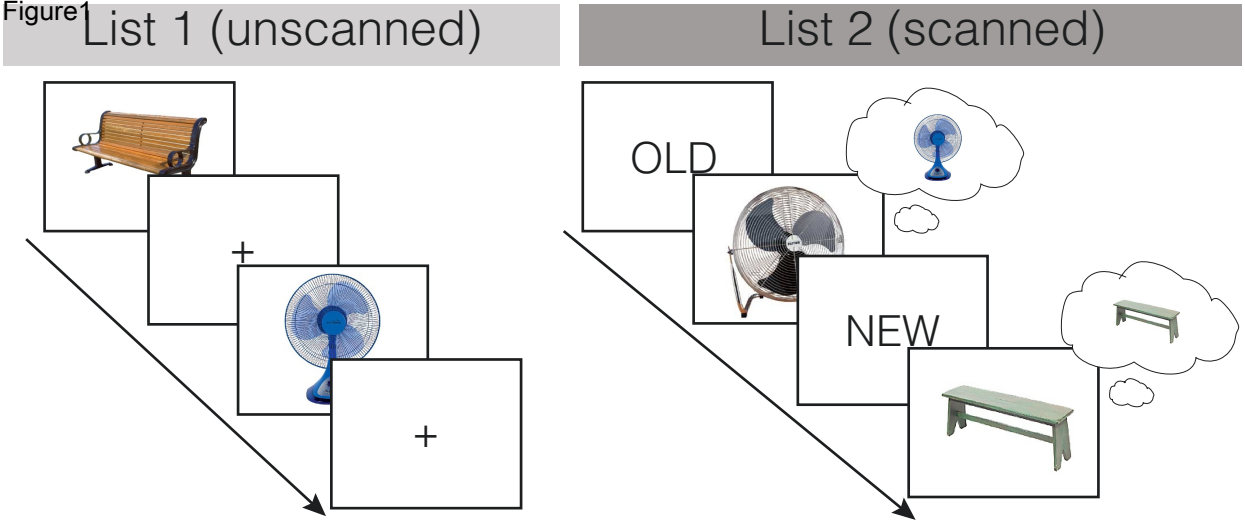


Figure2

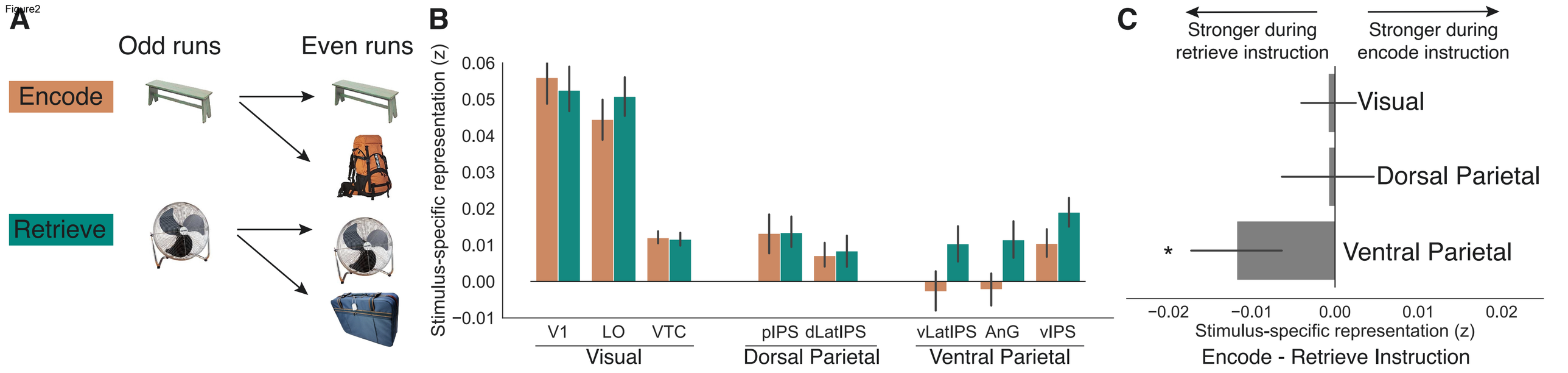
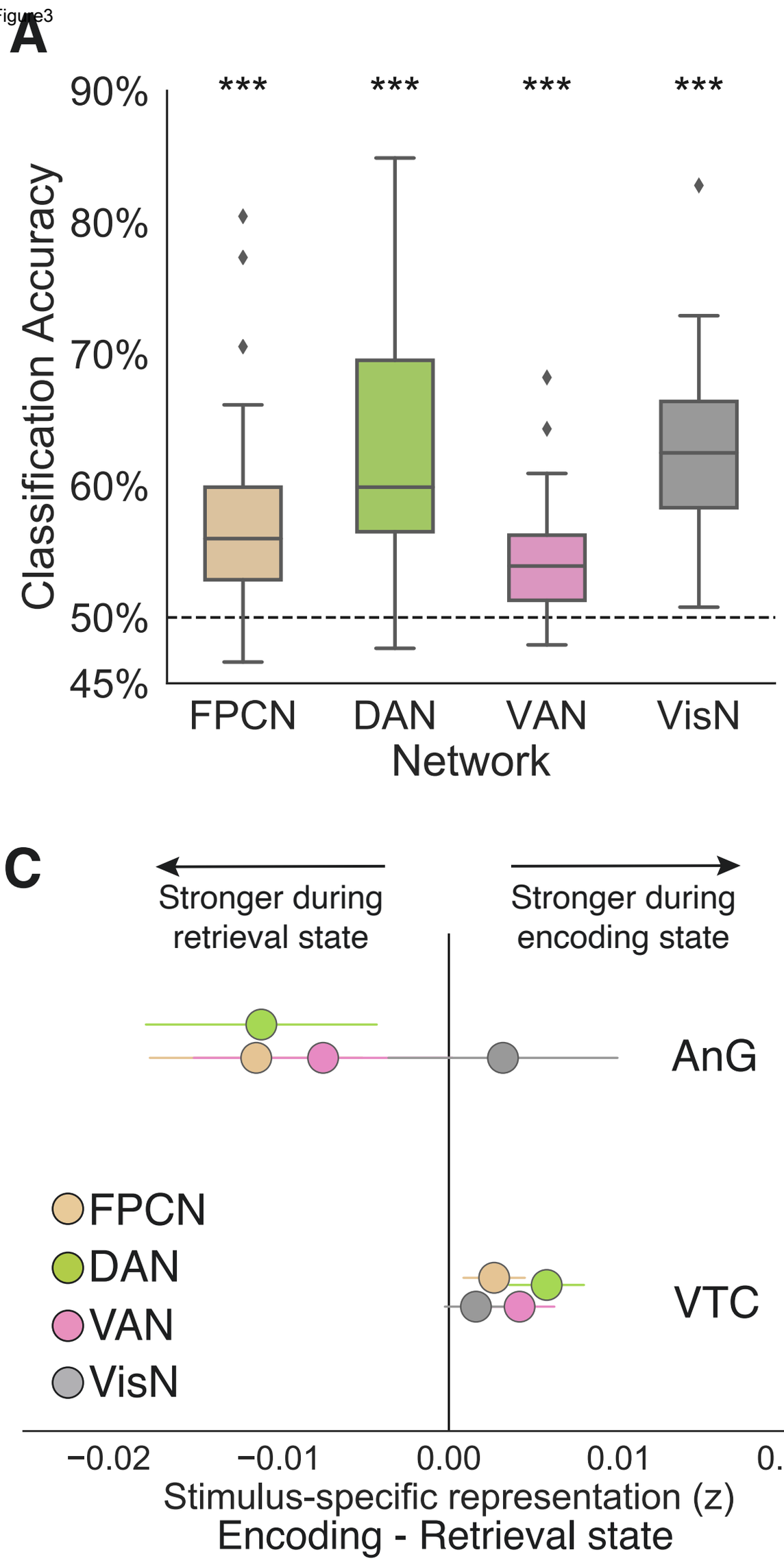
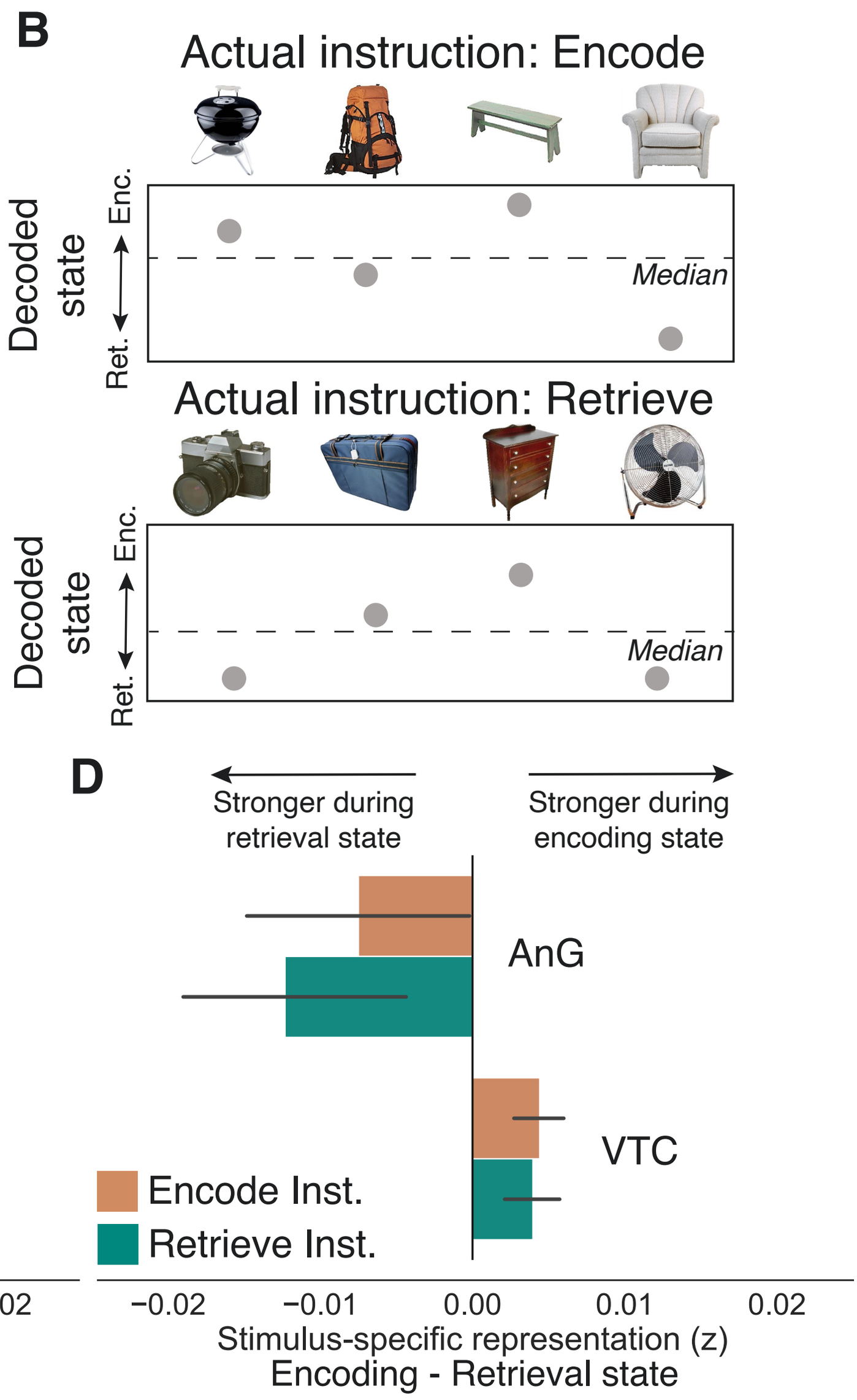

A
Odd runs
Even runs
Encode
Retrieve
B
Stimulus-specific representation (z)
0.06
0.05
0.04
0.03
0.02
0.01
0.00
−0.01
V1
LO
VTC
Visual
pIPS
dLatIPS
Dorsal Parietal
vLatIPS
AnG
vIPS
Ventral Parietal
C
Stronger during retrieve instruction
Stronger during encode instruction
Visual
Dorsal Parietal
Ventral Parietal
*
−0.02
−0.01
0.00
0.01
0.02
Stimulus-specific representation (z)
Encode - Retrieve Instruction

Figure3

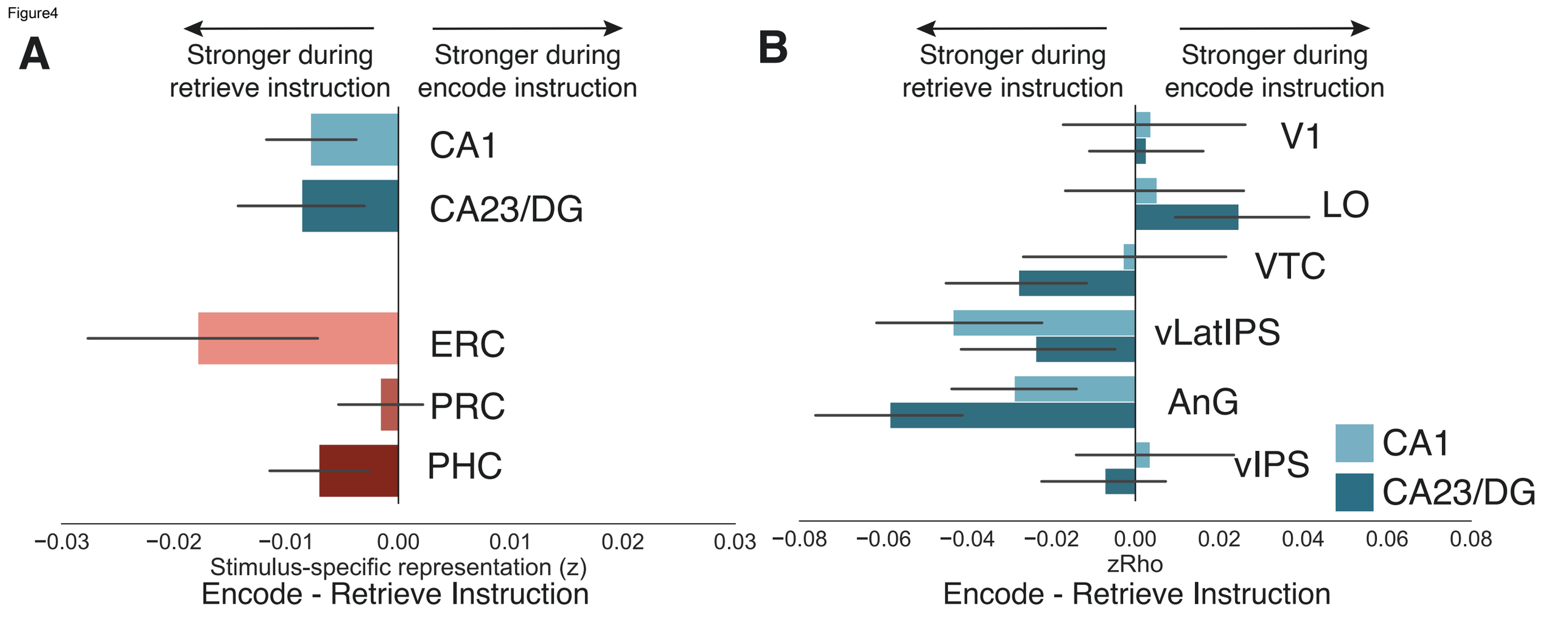

A
90%
80%
70%
60%
50%
45%
Classification Accuracy
***
***
***
***
FPCN
DAN
VAN
VisN
Network


B
Actual instruction: Encode
Decoded state
Enc.
Ret.
Median
Actual instruction: Retrieve
Decoded state
Enc.
Ret.
Median

C
Stronger during retrieval state
Stronger during encoding state
AnG
VTC
FPCN
DAN
VAN
VisN
−0.02
−0.01
0.00
0.01
0.02
Stimulus-specific representation (z)
Encoding - Retrieval state

D
Stronger during retrieval state
Stronger during encoding state
AnG
VTC
Encode Inst.
Retrieve Inst.
−0.02
−0.01
0.00
0.01
0.02
Stimulus-specific representation (z)
Encoding - Retrieval state

Figure4


A
Stronger during retrieve instruction
Stronger during encode instruction
CA1
CA23/DG
ERC
PRC
PHC
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03
Stimulus-specific representation (z)
Encode - Retrieve Instruction
B
Stronger during retrieve instruction
Stronger during encode instruction
V1
LO
VTC
vLatIPS
AnG
vIPS
CA1
CA23/DG
−0.08
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
0.08
zRho
Encode - Retrieve Instruction

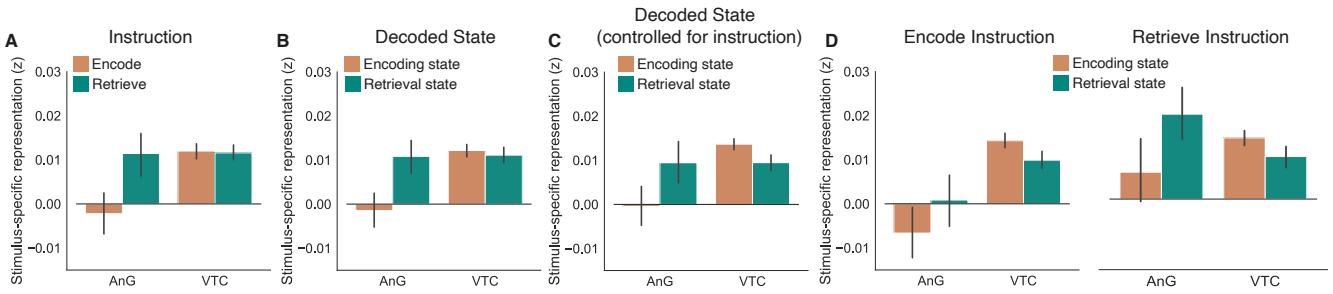


Figure S1. Stimulus-specific representations across angular gyrus and ventral temporal cortex. Related to Figure 3. (A) Stimulus-specific representations for each cortical location (angular gyrus, AnG; ventral temporal cortex, VTC) as a function of instruction (encode instruction = orange, retrieve instruction = teal). The interaction between cortical location and instruction was significant ($F_{1,32} = 5.022$, $p = 0.0321$). **(B)** Stimulus-specific representations for each cortical location as a function of memory state decoded from the attentional networks (encoding state = orange, retrieval state = teal), not controlling for actual instruction. As reported in the main text, the interaction between cortical location and decoded state was significant ($p < 0.01$) **(C)** Stimulus-specific representations for each cortical location as a function of memory state decoded from the attentional networks (encoding state = orange, retrieval state = teal), controlling for actual instruction. As reported in the main text, the interaction between cortical location and decoded state was significant ($p < 0.05$) **(D)** Stimulus-specific representations for each cortical location as a function of memory state decoded from the attentional networks (encoding state = orange, retrieval state = teal), separated by actual instruction (encode, retrieve). The interaction between cortical location and decoded memory state did not further interact with instruction condition ($p = 0.613$). Error bars are standard error of the mean.

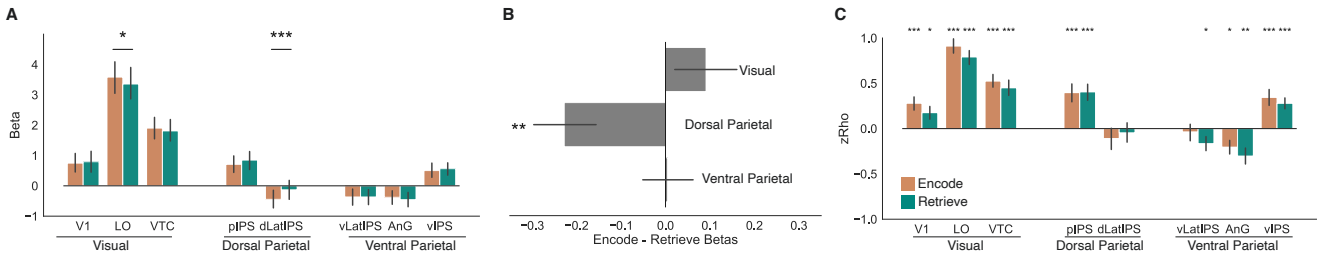


Figure S2. Univariate effects across cortical locations. Related to Figure 2. (A) Mean beta values for each of the eight visual and parietal ROIs as a function of instruction condition (encode instruction = orange, retrieve instruction = teal). A significant difference between instruction conditions was observed in LO (encode $>$ retrieve: $t_{32} = 2.142$, $p = 0.0399$) and dLatIPS (retrieve $>$ encode: $t_{32} = 3.166$, $p = 0.0034$). **(B)** Difference in mean beta values as a function of instruction condition (encode retrieve) for each of three broad cortical regions (visual [V1, LO, VTC]; dorsal parietal [pIPS, dLatIPS]; ventral parietal [vLatIPS, AnG, vIPS]). Beta values did not differ for encode vs. retrieve trials in either the visual or ventral parietal regions (t's < 1.3, p's > 0.20). Beta values were significantly greater for retrieve compared to encode trials in the dorsal parietal region ($t_{32} = 3.166$, $p = 0.0034$). **(C)** Correlation between the mean beta value and the strength of stimulus-specific representation for each stimulus within each of the eight visual and parietal ROIs, separated by instruction condition (encode instruction = orange, retrieve instruction = teal). Significant correlations indicate that the strength of the univariate response for a given stimulus was related to the strength of the stimulus-specific representation. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

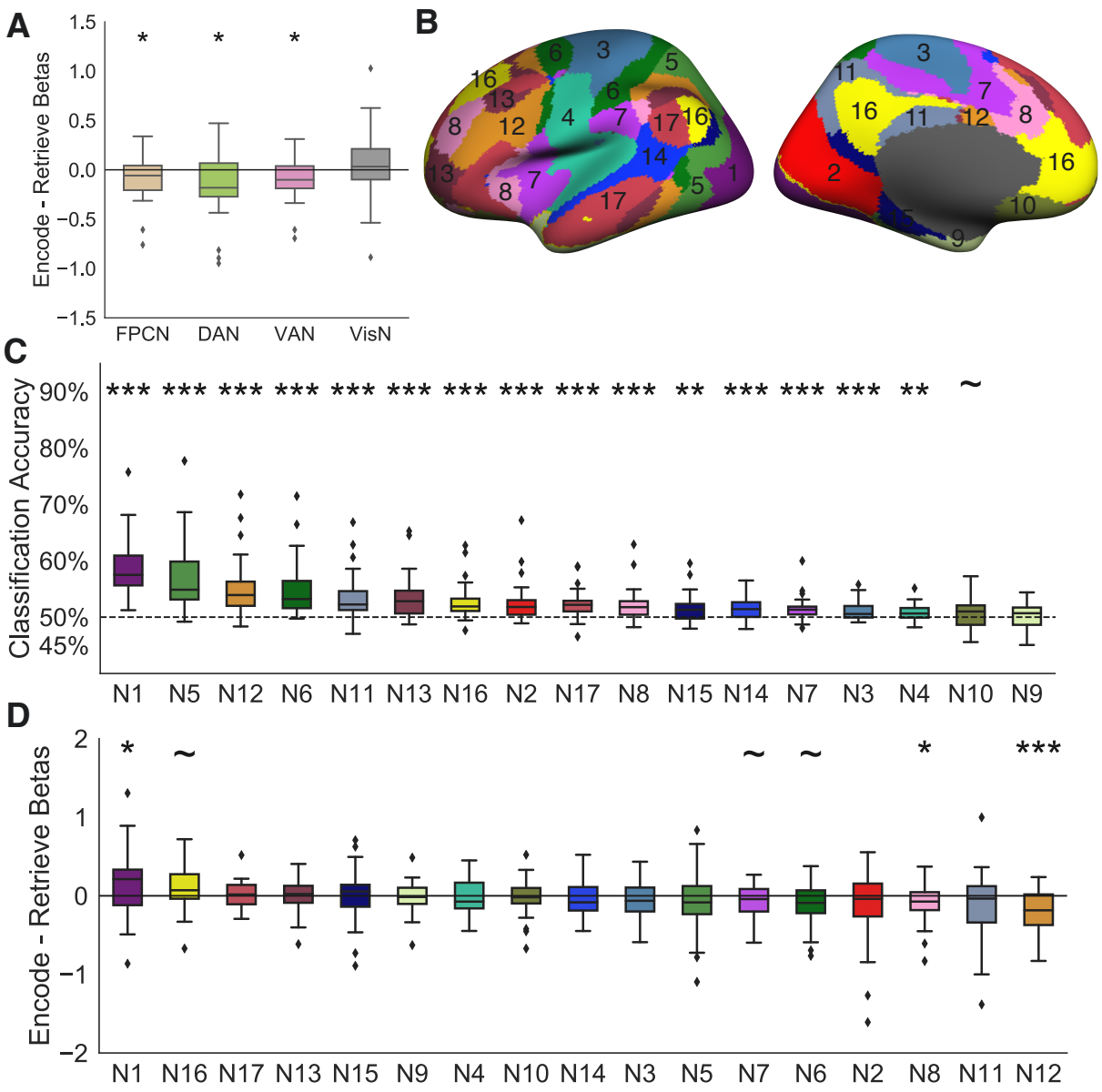


Figure S3. Effects of encode versus retrieve trials across cortical networks. Related to Figure 3. **(A)** Difference in univariate response (mean beta values) for encode minus retrieve instruction trials in four *a priori* cortical networks (frontoparietal control network, FPCN; dorsal attention network, DAN; ventral attention network; VAN, visual network VisN). Beta values were significantly greater for retrieve compared to encode instruction trials in FPCN, DAN, and VAN (t's > 2, p's < 0.05). **(B)** 17 cortical sub-networks from Yeo et al., 2011 **(C)** Cross-validated classification accuracy of encode vs. retrieve instruction trials across the 17 cortical sub-networks. 1000 voxels from each network were randomly sampled over 100 iterations to match the number of voxels across networks. Decoding accuracy was significantly above chance (50%), without correction for multiple comparisons, in all networks (t's > 3, p's < 0.01) except Network 10 ($t_{32} = 1.6965$, $p = 0.0995$) and Network 9 ($t_{32} = 0.6672$, $p = 0.5094$). **(D)** Difference in beta values for encode minus retrieve instruction trials across the 17 cortical sub-networks. Beta values were significantly greater for encode compared to retrieve trials in N1 (p = 0.0334) and significantly greater for retrieve compared to encode instruction trials in N8 (p = 0.02) and N12 (p = 0.0004). $\sim p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

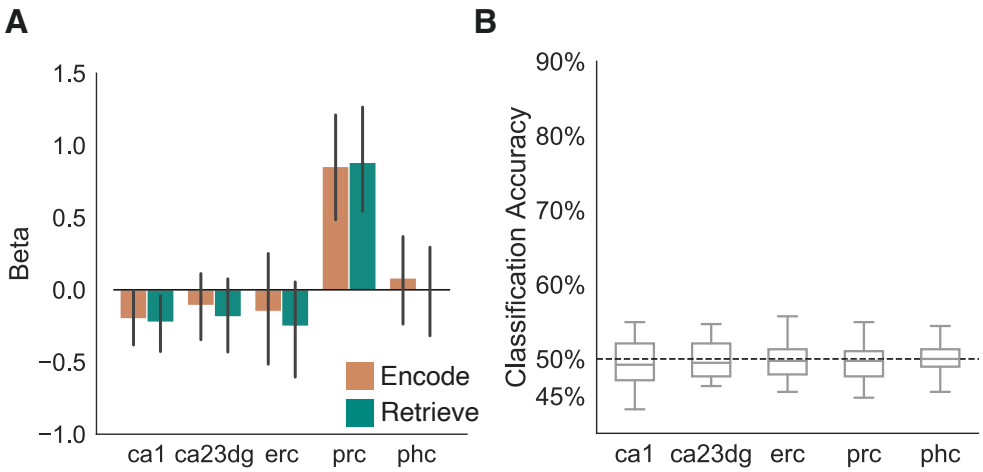


Figure S4. Effects of encode versus retrieve trials in the medial temporal lobes. Related to Figure 4. (A) Univariate responses (mean beta values) for two hippocampal regions (CA1, CA23DG) and three medial temporal lobe cortical regions [entorhinal cortex (ERC), perirhinal cortex (PRC), parahippocampal cortex (PHC)] as a function of instruction condition (encode instruction = orange, retrieve instruction = teal). None of the ROIs exhibited a significant difference between encode and retrieve trials (t's < 1.65, p's > 0.10). **(B)** Cross-validated classification of encode vs. retrieve instruction trials for the hippocampal and medial temporal lobe cortical regions. Classification accuracy was not above chance (50%) for any of the ROIs (t's < 1.2, p's > 0.30).