



Published in final edited form as:

*Hippocampus*. 2014 October ; 24(10): 1248–1260. doi:10.1002/hipo.22310.

## CA<sub>1</sub> subfield contributions to memory integration and inference

Margaret L. Schlichting<sup>1,2</sup>, Dagmar Zeithamova<sup>1,2,3</sup>, and Alison R. Preston<sup>1,2,3,\*</sup>

<sup>1</sup>Department of Psychology, The University of Texas at Austin, 1 University Station, Austin, TX 78712, Phone: (512) 232-5145, Fax: (512) 475-8000

<sup>2</sup>Center for Learning and Memory, The University of Texas at Austin, 1 University Station, Austin, TX 78712, Phone: (512) 232-5145, Fax: (512) 475-8000

<sup>3</sup>Institute for Neuroscience, The University of Texas at Austin, 1 University Station, Austin, TX 78712, Phone: (512) 232-5145, Fax: (512) 475-8000

### Abstract

The ability to combine information acquired at different times to make novel inferences is a powerful function of episodic memory. One perspective suggests that by retrieving related knowledge during new experiences, existing memories can be linked to the new, overlapping information as it is encoded. The resulting memory traces would thus incorporate content across event boundaries, representing important relationships among items encountered during separate experiences. While prior work suggests that the hippocampus is involved in linking memories experienced at different times, the involvement of specific subfields in this process remains unknown. Using both univariate and multivariate analyses of high-resolution functional magnetic resonance imaging (fMRI) data, we localized this specialized encoding mechanism to human CA<sub>1</sub>. Specifically, right CA<sub>1</sub> responses during encoding of events that overlapped with prior experience predicted subsequent success on a test requiring inferences about the relationships among events. Furthermore, we employed neural pattern similarity analysis to show that patterns of activation evoked during overlapping event encoding were later reinstated in CA<sub>1</sub> during successful inference. The reinstatement of CA<sub>1</sub> patterns during inference was specific to those trials that were performed quickly and accurately, consistent with the notion that linking memories during learning facilitates novel judgments. These analyses provide converging evidence that CA<sub>1</sub> plays a unique role in encoding overlapping events and highlight the dynamic interactions between hippocampal-mediated encoding and retrieval processes. More broadly, our data reflect the adaptive nature of episodic memories, in which representations are derived across events in anticipation of future judgments.

### Keywords

associative inference; episodic memory; high-resolution fMRI; integrative encoding; pattern similarity

---

\*Corresponding author: Alison R. Preston, Ph.D., Center for Learning & Memory, The University of Texas at Austin, 1 University Station, C7000, Austin, TX 78712, apreston@mail.clm.utexas.edu.

The authors declare no conflicts of interest.

## Introduction

Much memory research to date has focused on how our brains encode, store, and recall memories for individual experiences. However, in the real world, it is rare that an appropriate decision can be made on the basis of a single event memory alone; rather, the vast majority of decision and action requires us to draw upon knowledge derived across multiple events. For instance, we might infer a relationship between two people not because we have seen them together on a single occasion, but because we have seen them both with a common third element (e.g., their pet Poodle) on separate occasions. Existing theories suggest that the ability to extract new information from prior memories is supported by the hippocampus via relational memory networks, in which individual memory representations are connected to one another in terms of the people, places, and things they have in common (Eichenbaum, 1999). These networks could then enable the extraction of novel, never directly experienced information. From such flexible behaviors, it is clear that episodic memory serves a much more powerful function than simple recordkeeping. Rather, research suggests that our memories are prospectively oriented (Klein et al., 2002; Buckner, 2010; Addis and Schacter, 2012), formed in anticipation of future decisions in novel situations.

Memory network formation may rely on dynamic interactions that occur between old memories and new information during encoding. It has been suggested that related memories are retrieved during new experiences containing overlapping content, which may in turn impact the way the new information is encoded (Bartlett, 1932; Tolman, 1948; O'Keefe and Nadel, 1978; Cohen and Eichenbaum, 1995). Specifically, it has been proposed that encountering a new event that shares content with a previous experience would lead to the reactivation of the existing memory through pattern completion mechanisms supported by the hippocampus (McClelland et al., 1995; Eichenbaum, 2000; O'Reilly and Rudy, 2001). This learning-phase memory retrieval would allow new, externally available information to be encoded in relation to these internally generated (i.e., reactivated) memories, thereby facilitating subsequent inferences about the relationships among memories.

Despite the everyday importance of behaviors like inference, the underlying neural substrates are not well understood. One open question is the role of specific hippocampal subfields in encoding experiences containing elements that overlap with prior memories. The hippocampus is a heterogeneous structure, with subfields that differ in cellular organization, anatomical connectivity and mnemonic functionality (Manns and Eichenbaum, 2006). Despite this heterogeneity, the degree to which specific hippocampal subfields contribute to encoding and retrieval operations that support inference has not been studied in either animals or humans. However, based on its hypothesized function, we suggest that area CA<sub>1</sub> might play a particularly important role when experiencing events that overlap with prior knowledge. While theoretical and computational models (Marr, 1971; McNaughton and Morris, 1987) attribute binding of elements *within* individual episodes to area CA<sub>3</sub>, CA<sub>1</sub> may be important for relating information *across* episodes. In particular, CA<sub>1</sub> is hypothesized to serve as a comparator, signaling when new experiences deviate from memory-based expectations (Lisman and Grace, 2005; Chen et al., 2011; Duncan et al., 2012). The detection of differences between reactivated memories and current events may

trigger a specialized encoding process, leading to the formation of links between current experience and existing knowledge (Shohamy and Wagner, 2008; Wang and Morris, 2010; van Kesteren et al., 2012). Consistent with this idea, recent rodent work has demonstrated increases in CA<sub>1</sub> activity and plasticity in the presence of novel stimuli or familiar stimuli in novel locations (Larkin et al., 2014). The authors of that study propose that CA<sub>1</sub> may signal the existence of novelty, while at the same time increasing plasticity to allow prior memories to be updated with new information. Thus, the present study aims to assess the idea that CA<sub>1</sub> might play a unique role in encoding new content that relates to existing memories.

We employed an associative inference task (Preston et al., 2004; Zeithamova and Preston, 2010) in combination with high-resolution functional magnetic resonance imaging (fMRI) and neural pattern similarity analysis (NPSA; Kriegeskorte, Mur, & Bandettini, 2008). During study phases (Fig. 1a), participants were first presented with a series of AB object pairs (e.g., clipboard-truck) followed by the corresponding BC object pairs (e.g., truck-binoculars), where the B item (truck) was common to both associations. After each study phase, participants completed a two-alternative forced choice test of both the directly learned (AB, BC) and inference (AC; e.g., clipboard-binoculars) associations (Fig. 1b). The goals of the present study were to examine (1) how processes engaged during presentation of the overlapping memories are predictive of subsequent inference; as well as (2) how encoding responses are reinstated during successful inference judgments. Specifically, we addressed the hypothesis that the degree to which study patterns were reinstated during subsequent test would be associated with faster and more accurate inferences (Fig. 2).

## Materials and Methods

### Participants

Twenty-five right-handed volunteers from the Stanford University community participated in this study. Participants were in good general health and were screened for contraindications to MRI. Consent was obtained in accordance with an experimental protocol approved by the Stanford University and The University of Texas at Austin Institutional Review Boards. Participants received monetary compensation for their involvement in the study. Data from four participants were excluded for the following reasons: failure to achieve above chance accuracy on directly learned associations (3 participants) and loss of anatomical data (1 participant). Data from the remaining 21 participants (12 females, ages 18-31, median = 22 years) were included in all reported analyses.

### Materials

Stimuli consisted of 360 grayscale images of common objects organized into 144 overlapping pairs (72 AB pairs and 72 BC pairs, forming a total of 72 ABC triads; Fig. 1a) and 72 non-overlapping pairs (XY). Overlapping pairs were those for which two objects (A and C) were each associated with a third overlapping object (B). Non-overlapping XY pairs consisted of two unique objects not paired with any other items. To control for the viewing order and pair type of the object stimuli, participants were assigned to one of six randomization groups.

## Procedures

Participants completed an associative inference task (Preston et al., 2004; Zeithamova and Preston, 2010) during fMRI scanning. The task consisted of six alternating study and test phases; both study and test phases were scanned. During study phases, participants intentionally encoded object pairs (AB, BC, XY) and single objects (X) (Fig. 1a). Participants saw each pair only once during study, requiring rapid acquisition of associative information. During the test phases, participants were assessed on their memory for associations learned during the immediately preceding study phase. Memory for both the directly learned (AB, BC, XY) and inference (AC) associations was tested using a two-alternative forced choice procedure (Fig. 1b).

Study materials were presented in a mixed fMRI design in which stimuli were blocked by type. Study scans consisted of four cycles during which each of four condition blocks (AB, BC, X, XY) was presented exactly once. Condition blocks were presented such that AB blocks immediately preceded BC blocks and X blocks immediately preceded XY blocks; the order was counterbalanced both within and across participants. Each condition block consisted of 12 study trials lasting a total of 72 s. For each trial, stimuli (a pair of images or a single object) were presented on the screen for 3 seconds with a 1 second response period, during which time participants made a judgment of learning (1, will remember; 2, may remember; 3, will forget). These responses were collected solely to ensure participants' attention during the study task and were not considered in the data analysis. Trial onsets were jittered within each condition block. Between study trials, participants completed a variable number of 2 s baseline trials (range 0-3 baseline trials) in which a single digit ranging from 1 to 8 was presented on the screen; participants indicated with a button press whether the digit was odd or even (Stark and Squire, 2001). Baseline task blocks (12 s) were also presented at the beginning and end of each study phase scan and in between each condition block.

An event-related test scan occurred after each study scan. Tested materials included directly learned object pairs (AB, BC, XY) and inference associations (AC) that were viewed during the immediately preceding study phase. Each test trial lasted 4 s during which time 3 objects were presented on the screen (Fig. 1b): a cue object at the top (e.g., object A from triad 1, denoted  $A_1$ ) with two options at the bottom (the correct stimulus from the same triad, denoted  $B_1$ , and a studied stimulus from a different triad, e.g.,  $B_2$ ). Participants indicated which of these two choice objects was associated with the cue object by pressing a button. For inference (AC) judgments, participants were told that the relationship between the cue (e.g.,  $A_1$ ) and the correct choice ( $C_1$ ) was mediated through their common association with a third item ( $B_1$ ). For both direct and inference test trials, incorrect choices were familiar objects from a different triad in the same study scan. The order of test trials was pseudo-random such that the inference test trial for a given triad was presented before the corresponding direct test trials (i.e.,  $A_1C_1$  was tested prior to testing  $A_1B_1$  and  $B_1C_1$ ) to prevent additional learning of the direct associations during the test phases. Odd/even digit baseline trials (range 0-5) were intermixed with test trials. For both study and test phases, the order of conditions was determined by a sequencing program to optimize the efficiency of event-related fMRI design (Dale, 1999). Because study and test phases were presented in

alternation, participants were aware during learning that they would be tested on both direct and inference associations.

### **fMRI data acquisition and preprocessing**

Imaging data were acquired on a 3.0 T GE Signa MRI system (GE Medical Systems). Functional images were acquired using gradient echo spiral in/out pulse sequence (TR = 2 s; TE = 30 ms; 2 shot; flip angle = 61;  $128 \times 128$  matrix;  $2 \times 2 \times 2$  mm voxels; 14 oblique axial slices oriented parallel to the main axis of the hippocampus) (Glover and Law, 2001). While the slices were prescribed to cover hippocampus proper in all participants, we did not achieve full coverage of the parahippocampal gyrus in the majority of participants. Thus, the analyses reported here are restricted to hippocampal subfields. A T2-weighted inplane structural image was acquired in the same prescription as the functional images (TR = 3 s, TE = 68 ms,  $512 \times 512$  matrix,  $0.43 \times 0.43$  mm in-plane resolution). To delineate anatomical regions of interest, we also acquired a high-resolution T2-weighted coronal structural image (TR = 3 s, TE = 68 ms,  $512 \times 512$  matrix,  $0.43 \times 0.43$  mm in-plane resolution, 30 3-mm thick slices). A T1-weighted 3D SPGR structural volume ( $256 \times 256 \times 156$  matrix,  $0.86 \times 0.86 \times 1$  mm) was also collected to facilitate coregistration of the inplane and coronal images. Foam padding was used to minimize head motion.

To obtain a field map for correction of magnetic field heterogeneity, the first volume of each functional scan was collected with an echo time 2 ms longer than all subsequent volumes. For each slice, the map was calculated from the phase of the first two time frames and applied as a first order correction during reconstruction of the functional images. In this way, blurring and geometric distortion were minimized on a per-slice basis. In addition, correction for off-resonance due to breathing was applied on a per-time-frame basis using phase navigation (Pfeuffer et al., 2002). This initial volume was then discarded as well as the following five volumes of each scan (for a total of six discarded volumes, or 12 s) to allow for T1 stabilization.

Data were preprocessed and analyzed using SPM5 (Wellcome Department of Imaging Neuroscience, University College London, London, UK) and custom Matlab (MathWorks, MA) routines. Functional images were corrected to account for differences in slice acquisition times by interpolating the voxel time series using sinc interpolation and resampling the time series using the middle slice as a reference point. Functional images were then realigned to the first volume to correct for motion. The inplane and SPGR images were registered to the mean functional image created during the realignment step. Timeseries data were high-pass filtered with a 128 s cutoff and converted to percent signal.

Both neural pattern similarity and univariate analyses were performed under the assumptions of the general linear model (GLM). With the exception of group-level GLMs, all analyses were done in the native space of each participant. For all GLMs described below, regressor functions were constructed by modeling stimulus-related activation as a stick function convolved with a canonical hemodynamic response function. Study and test scans were modeled in separate GLMs for all analyses. X items and XY pairs served no purpose for the current study and thus were treated as events of no interest in all models. Additionally, test

trials for which no response was generated in the 4 s time window were excluded from all performance-based analyses.

Several participants in the present study had a small minimum number of trials per condition (e.g., due to few incorrect responses; see below for trial counts for each analysis). To ensure the present results were not substantially impacted by inclusion of these participants, we repeated all analyses in a subset of participants who had a minimum of eight trials per condition of interest. Imposing this restriction did not significantly change the pattern of the reported results; thus, the remainder of the paper focuses on the full sample.

## Univariate analysis

### Participant-level analysis

If existing knowledge is retrieved and linked to new experiences during learning, it is necessarily the case that this must occur during the second, overlapping associations (i.e., BC) in the present paradigm. We hypothesized that memory integration would be reflected in study-phase neural engagement that was predictive of later inference during acquisition of the overlapping—but not initially experienced—associations. Such a signature would reflect the unique opportunity afforded by BC study trials to retrieve and link prior AB memories with new experience, thereby supporting later inference decisions. Thus, we searched for regions whose encoding activation was more predictive of subsequent inference success during BC than AB study trials to examine the role of hippocampal subfields in this specialized encoding process (for a similar approach see Zeithamova & Preston, 2010). AB and BC study trials were sorted according to later success on the corresponding inference judgment; study events were then modeled separately according to trial type (AB, BC) and subsequent inference (inference correct, inference incorrect) to generate one model per participant. This procedure resulted in four regressors of interest (with trial counts for each condition reported in parentheses): AB-inference correct (range: 28-64 trials; mean  $\pm$  SEM =  $42.19 \pm 2.42$ ), AB-inference incorrect (4-37;  $22.29 \pm 2.31$ ), BC-inference correct (28-64;  $42.19 \pm 2.42$ ) and BC-inference incorrect (4-37;  $22.29 \pm 2.31$ ) study trials. We tested the study trial type by inference interaction (BC > AB  $\times$  inference correct > inference incorrect) to isolate those regions whose activation was more predictive of subsequent inference during BC than AB study trials.

### Spatial normalization

A custom template was generated using Advanced Normalization Tools (ANTS) (<http://picsl.upenn.edu/software/ants/>; Avants et al., 2011). The T2-weighted inplane (i.e., oblique axial) images from a subset of ten participants with canonical hippocampi were selected for template generation. After template generation, each participant's inplane image was normalized to the group template image as follows. To maximize alignment of hippocampi across participants, a bilateral hippocampal ROI was delineated by hand on each participant's inplane image as well as on the group template. Hippocampal ROIs were used as labels to guide the spatial normalization from each participant's inplane image to the group inplane template.

## Group-level analysis

The contrast images representing the study trial type by inference interaction from the participant-level GLMs were transformed to template space for group-level univariate analyses. Specifically, the transformations calculated for each participant's inplane image during the above-described spatial normalization step were applied to each participant's statistics image. The normalized contrast images were then smoothed with a 2.5 mm FWHM Gaussian kernel and submitted to a group-level GLM. We anticipated more neural evidence for this specialized encoding process among participants with good relative to poor inference ability. Thus, we were interested in those regions for which the interaction term was significantly modulated by AC performance across participants. An uncorrected voxelwise threshold of  $p < 0.025$  was applied to the group statistic image. Correction for multiple comparisons was performed using small volume correction to establish the cluster size corresponding to a cluster-corrected threshold of  $p < 0.05$ . This calculation was carried out using 3dClustSim, part of AFNI (Cox, 1996). 3dClustSim uses a Monte Carlo simulation approach to take into account the size and shape of the volume as well as the smoothness of the data in determining a critical cluster size. Simulations were performed separately for the right and left hippocampal ROIs. Cluster sizes that occurred with a probability of less than 0.05 across 5000 simulations were considered statistically significant. This procedure yielded a critical cluster size of 13 voxels for both the left and right hippocampus.

## Neural pattern similarity analysis (NPSA)

Numerous theories suggest that successful memory depends on reinstating the activation patterns engaged during the initial encoding experience (Tulving and Thomson, 1973; Alvarez and Squire, 1994; McClelland et al., 1995; Wheeler et al., 2000; Norman and O'Reilly, 2003; Moscovitch et al., 2005; Polyn et al., 2005). We reasoned that in the present paradigm, the degree of similarity between neural study and test patterns should also relate to inference performance. We used NPSA to quantify the degree of similarity between patterns of activation evoked during study with those evoked during inferential test trials for each of the five anatomical regions of interest (ROIs): anterior hippocampus, CA<sub>1</sub>, a combined dentate gyrus/CA<sub>2,3</sub> region (DG/CA<sub>2,3</sub>), subiculum, and posterior hippocampus. ROIs were manually segmented on the high-resolution T2-weighted coronal images using established guidelines (Amaral and Insausti, 1990; Insausti et al., 1998; Pruessner et al., 2000, 2002; Zeineh et al., 2000, 2003; Preston et al., 2010), registered to the SPGR and downsampled to functional resolution. Anterior and posterior hippocampal ROIs were defined as those portions of the hippocampus for which subfields could not be reliably delineated. This analysis was performed in the native functional space of each participant; no spatial normalization was performed. After resampling to functional resolution, the number of voxels in each hippocampal subfield ROI were as follows (range, mean  $\pm$  SEM): anterior hippocampus: 205-544, 312.43  $\pm$  20.73; CA<sub>1</sub>: 159-393, 282.43  $\pm$  12.32; DG/CA<sub>2,3</sub>: 267-487, 362.86  $\pm$  13.60; subiculum: 148-330, 235.57  $\pm$  11.03; posterior hippocampus: 269-734, 495.67  $\pm$  25.19. All voxels within a given anatomical region were included in the analysis (i.e., no voxel selection was performed).

Three sets of GLMs were performed for the purposes of the NPSA: one based on AC performance (performance-based analysis), one based on response time (RT) on the correct

AC triads (RT-based analysis), and one based on event type, irrespective of performance (cross-participant analysis).

For all three analyses, voxelwise parameter estimates were extracted for each participant, condition of interest, and region, resulting in a vector of parameter estimates (one per voxel) associated with each condition in each hippocampal subfield ROI. The resulting study and test patterns were then used to address our hypotheses about the nature of the neural patterns reinstated during the inference test. In all analyses, Pearson correlations were computed to assess the similarity between study and inference test patterns. Correlation coefficients were then Fisher transformed to more closely conform to the assumptions of normality underlying standard statistical tests.

### Performance-based NPSA

The performance-based analysis assessed how study-test pattern similarity related to inference within each participant. One perspective predicts that reinstatement of both AB and BC encoding patterns will relate to inference performance, as memories for both associations are needed for successful inference. On the other hand, learning of overlapping BC associations provides a unique opportunity to link two related memories. Thus, this perspective makes a different prediction: namely, that test-phase reinstatement of overlapping BC study patterns should be more related to inference than should reinstatement of initial AB study patterns (Fig. 2). We reasoned that for subfields supporting this BC-specific learning mechanism, AC test patterns should be more similar to the neural patterns evoked during BC learning than to those evoked during the encoding of initial AB associations (Fig. 2b, bidirectional arrows). In particular, we predicted that these subfields would show a study trial type by inference success interaction in study-test pattern similarity measures, such that BC-AC similarity would be more predictive of performance within participant than would AB-AC similarity. We also anticipated greater study-test pattern similarity for BC than AB study trials, particularly for triads on which the AC inference was correct. We limited this analysis to those triads for which both corresponding direct associations (AB and BC) were correctly remembered, henceforth referred to as “direct correct triads.” This was done to investigate the processes involved specifically in inference, i.e., those processes that go above and beyond individual pair encoding.

Study and test trials associated with direct correct triads were split based on inference performance (AC correct or AC incorrect). For the study-phase GLMs, regressors were constructed representing the following conditions of interest (all limited to direct correct triads): AB trials for which AC was later correct (range: 10-56 trials; mean  $\pm$  SEM =  $28 \pm 3.23$  trials); AB trials for which AC was later incorrect (3-24;  $11.33 \pm 1.15$ ); BC trials for which AC was later correct (10-56;  $28 \pm 3.23$ ); and BC trials for which AC was later incorrect (3-24;  $11.33 \pm 1.15$ ). AB and BC study trials for which one or both of the corresponding direct test trials were incorrect (i.e., direct incorrect triads) were modeled as events of no interest. For the test-phase GLMs, AC trials were first limited to those for which the corresponding direct pairs were correct and then further split based on AC performance. This procedure resulted in two regressors of interest: AC correct test trials and



AC incorrect test trials, both limited to direct correct triads. AB test trials, BC test trials, and AC trials corresponding to direct incorrect triads were modeled as regressors of no interest.

Voxelwise parameter estimates were extracted for each participant, ROI and condition of interest. The following measures of neural similarity between study and test patterns were then calculated using Pearson correlation and Fisher transformed: AB-AC study-test similarity for AC correct triads; AB-AC study-test similarity for AC incorrect triads; BC-AC study-test similarity for AC correct triads; BC-AC study-test similarity for AC incorrect triads. We then performed a  $2 \times 2$  repeated measures analysis of variance of study-test pattern similarity with study trial type (AB, BC) and inference success (AC correct, AC incorrect) as factors for each region. For regions demonstrating significant interactions, follow-up comparisons were performed using paired t-tests. Pattern similarity measures for this analysis were also compared across ROIs using a  $5 \times 4$  repeated measures ANOVA, with region (anterior hippocampus, CA<sub>1</sub>, DG/CA<sub>2,3</sub>, subiculum, posterior hippocampus) and condition (AB-AC correct, AB-AC incorrect, BC-AC correct, BC-AC incorrect) as factors.

### RT-based NPSA

We further hypothesized that in regions supporting subsequent inference through a specialized encoding process that links related memories, reinstating study patterns at test would be associated with faster inference judgments. Accordingly, we constructed RT-based models to further interrogate those regions showing a significant study trial type  $\times$  inference success interaction from the performance-based analysis described above. We tested the hypothesis that the interaction between study trial type and inference would be driven specifically by the fast inferences. Moreover, as this encoding process is engaged only during overlapping events, reinstatement of BC study patterns in particular may be associated with a speed advantage.

Regressors were constructed for these GLMs as described in the performance-based analysis above with a single exception: for each participant, study and test trials associated with AC correct triads were further split by median RT on the critical AC inference judgment. For the study scans, this resulted in GLMs with the following four conditions of interest: AB study trials for which AC was later correct, fast RT; AB study trials for which AC was later correct, slow RT; BC study trials for which AC was later correct, fast RT; BC study trials for which AC was later correct, slow RT (all limited to direct correct triads). AB and BC study trials associated with AC incorrect triads were modeled as regressors of no interest, regardless of RT. GLMs for the test phases included regressors for two conditions of interest: AC correct test trials, fast RT; and AC correct test trials, slow RT. Each condition of interest included an average of 14 trials (range 5-28 trials per condition; mean  $\pm$  SEM =  $13.86 \pm 1.59$ ). AC incorrect trials were modeled regardless of RT as a separate regressor of no interest. Additional regressors of no interest for study- and test-phase GLMs were identical to those described in the performance-based GLMs above.

After extracting patterns associated with each condition of interest and ROI for every participant, we computed four pattern similarity values: AB-AC study-test similarity for fast AC correct triads; AB-AC study-test similarity for slow AC correct triads; BC-AC study-test

similarity for fast AC correct triads; BC-AC study-test similarity for slow AC correct triads. Comparisons of interest were then performed using paired t-tests.

### Cross-participant NPSA

The cross-participant NPSA was performed to explore individual differences in the relationship between study-test pattern similarity and inference ability. Events were modeled irrespective of performance to avoid introducing bias into the across-participant correlations that could result from including varying numbers of trials across participants. Thus, all trials were included in this analysis. AB study trials and BC study trials were the conditions of interest for the study phase (modeled irrespective of subsequent memory or inference; 72 trials per condition). For the test scans, AC trials served as the single regressor of interest (72 trials); AB and BC test trials were modeled as regressors of no interest.

We then compared the distributed activation patterns associated with AC test to those evoked during AB and BC encoding. These calculations were performed separately for each participant, resulting in AB-AC study-test and BC-AC study-test similarity measures, computed irrespective of performance. These study-test pattern similarity measures (Fisher's  $z$ ) were then related to AC inference performance (proportion correct) across participants using Pearson correlation. For those ROIs demonstrating a significant relationship between pattern similarity and AC performance for either AB or BC study trials, AB-AC and BC-AC correlations with inference performance were compared using the Hotelling-Williams test (Hotelling, 1940; Williams, 1959).

### Study-test lag calculations

In the present task, it is necessarily the case that BC study pairs will be presented closer in time to the AC test trials than will AB study pairs. To ensure that any observed differences between BC and AB study-test pattern similarity were not simply due to differences in study-test temporal distance, we calculated mean lag in seconds between study and test trials separately for each study trial type (AB, BC) and inference performance (AC correct, AC incorrect) for each participant. As in the performance-based pattern similarity analysis described above, we limited our comparisons to direct correct triads. This resulted in an average lag for AB-AC correct, AB-AC incorrect, BC-AC correct, and BC-AC incorrect trial types, which we compared using both a  $2 \times 2$  repeated measures ANOVA with study type and inference performance as factors and follow-up paired t-tests.

## Results

### Behavioral results

Performance on both directly learned AB and BC pairs (range: 63.8%-97.1% correct; mean  $\pm$  SEM = 79.2%  $\pm$  2.5% correct;  $t(20) = 11.49$ ,  $p < 0.001$ ) and inference associations (AC; 43.1%-93.8%; 65.2  $\pm$  3.4%;  $t(20) = 4.48$ ,  $p = 0.002$ ) was above chance. Memory for AB pairs was significantly higher than memory for BC pairs (AB: 63.9%-97.2%; 81.6%  $\pm$  2.5%; BC: 59.4%-96.9%; 76.8%  $\pm$  2.7%;  $t(20) = 4.42$ ,  $p = 0.003$ ). There were large individual differences in accuracy for inference judgments, which allowed us to determine how hippocampal activation related to inference performance across participants.

### **CA<sub>1</sub> activation during learning of overlapping events predicts subsequent inference**

We conducted a univariate analysis to investigate the relationship between hippocampal encoding activation and later inference. Importantly, this specialized encoding process can only occur in the present task during the second, overlapping (i.e., BC) experiences, when current experience may be compared with prior memories. The detection of novelty in the environment then allows for the formation of links between related memories. Thus, we reasoned that in subfields supporting this mechanism, trial-by-trial encoding responses should be more predictive of subsequent inference during study of BC than AB pairs (Zeithamova and Preston, 2010). To test the hypothesis that CA<sub>1</sub> supports this process, we interrogated the study trial type by inference interaction (i.e., BC > AB × inference correct > inference incorrect) across the entire hippocampus. We were specifically interested in those regions for which the interaction term was modulated by AC performance. Significant activation was observed exclusively in the right CA<sub>1</sub> body (Fig. 3), consistent with this region's hypothesized role in detecting novelty and integrating current experience with prior memories.

### **Neural patterns during inference reflect reinstatement of encoding patterns in CA<sub>1</sub>**

We next employed a more sensitive multivariate approach to directly compare neural engagement during study and test. Successful memory retrieval is thought to occur via reinstatement of the neural patterns engaged during encoding (Tulving and Thomson, 1973; Alvarez and Squire, 1994; McClelland et al., 1995; Wheeler et al., 2000; Norman and O'Reilly, 2003; Moscovitch et al., 2005; Polyn et al., 2005). We predicted that reinstating study patterns might also relate to inference performance. Moreover, as BC study trials provide a unique opportunity to link current information with existing memories, we hypothesized that this relationship may be specific to BC-AC study-test similarity for subfields supporting this specialized encoding mechanism.

### **Performance-based analysis**

To examine how patterns of hippocampal activation evoked during AB and BC study were later reinstated during inference judgments, we directly compared BC-AC and AB-AC study-test pattern similarities as a function of inference success. Because BC study trials present a unique opportunity to link related memories, we hypothesized that the degree to which BC study patterns were accessed during inference would be related to performance, with greater BC-AC study-test similarity for correct relative to incorrect inferences. We anticipated that AB-AC study-test similarity would be less related to performance. Moreover, we predicted that in cases of correct inference, AC test patterns would be more similar to BC study than AB study patterns, reflecting reinstatement of encoding patterns unique to memory integration.

Parameter estimates associated with each condition were extracted for every voxel within the anatomically defined hippocampal subfield ROIs, resulting in average patterns of distributed activation for AB study, BC study, and AC inference trials, all for which AC was either correct or incorrect (for a total of six conditions). Importantly, this analysis was limited to only those triads for which the corresponding AB and BC pairs were both remembered. Study-test pattern similarities were then computed for each participant as a

function of study trial type and inference performance. Within each hippocampal subfield, we tested the following predictions: (1) that there would be a significant study trial type  $\times$  inference interaction and (2) that this interaction would be driven by greater BC-AC than AB-AC study-test similarity when AC was correct. Because these statistical tests were performed on all five hippocampal subfields, Bonferroni correction was performed to yield a critical p-value of 0.01 (corrected  $p < 0.05$ ).

We observed the predicted study trial type  $\times$  inference interaction exclusively in CA<sub>1</sub> ( $F(1, 20) = 12.26, p = 0.002$ ) (Fig. 4a). Moreover, a follow-up paired t-test revealed that these effects were driven by significantly greater BC-AC than AB-AC study-test pattern similarity for AC correct triads ( $t(20) = 3.69, p < 0.001$ ). Pattern similarity also differentiated correct from incorrect inference performance for BC ( $t(20) = 3.33, p = 0.003$ ) but not AB ( $t(20) = 0.25, p = 0.81$ ) study trial types. Main effects of study trial type ( $F(1, 20) = 5.40, p = 0.03$ ) and inference performance ( $F(1, 20) = 4.43, p = 0.05$ ) did not reach our corrected significance threshold in this region.

Other ROIs did not show the same relationship between study and test patterns as CA<sub>1</sub> (all other study trial type  $\times$  inference interactions  $F(1,20) < 1, p > 0.33$ ; Fig. 5a-d, top charts). Rather, in anterior hippocampus, subiculum, and posterior hippocampus, we observed a main effect of inference performance that survived Bonferroni correction, demonstrating greater study-test similarity for correct vs. incorrect trials (all  $F(1, 20) > 12.87, p < 0.01$ ). There were no significant main effects of study trial type at our corrected threshold (all  $F(1, 20) < 6.86, all p > 0.01$ ).

To assess the differences in NPSA measures across subfields, we performed a  $5 \times 4$  repeated measures ANOVA with region and condition (AB-AC correct, AB-AC incorrect, BC-AC correct, BC-AC incorrect) as factors. The region  $\times$  condition interaction was significant ( $F(12,240) = 4.80, p < 0.001$ ), indicating that the relationship between pattern similarity and subsequent inference differed significantly across hippocampal subfields.

### RT-based analysis

We reasoned that test-phase reinstatement of study patterns would be associated with facilitated inference. Here, we hypothesized that specifically for fast inferences, AC test patterns would more closely reflect BC than AB study patterns in regions identified above as important for linking memories during encoding. Moreover, we predicted that evidence for reinstatement of hippocampal patterns evoked during BC study would be greater for fast compared with slow inferences. In this analysis, study and test trials associated with correct inference judgments were median split based on AC test RT. As in the performance-based analysis described above, all study and test trials were limited to only those triads for which AB and BC were both correct. Parameter estimates associated with the conditions of interest were extracted for every voxel in CA<sub>1</sub> for each participant. We then computed study-test pattern similarity as described previously.

Consistent with our predictions, we observed greater BC-AC than AB-AC study-test pattern similarity for fast ( $t(20) = 2.58, one-tailed p = 0.009$ ) but not slow ( $t(20) = 0.68, p = 0.50$ ; Fig. 4b) inferences. We also found higher BC-AC study-test pattern similarity for fast

relative to slow AC correct triads ( $t(20) = 2.13$ , one-tailed  $p = 0.02$ ). This difference was not observed for AB study measures ( $t(20) = -0.67$ ,  $p = 0.51$ ). Together, these results suggest that the observed study trial type  $\times$  inference interaction was driven specifically by the fast inferences, i.e., those likely facilitated by retrieval during learning. Notably, this effect was found only for BC study trials.

### Cross-participant analysis

We performed the cross-participant NPSA to assess the relationship between study-test pattern similarity and individual differences in inference performance. Parameter estimates were extracted for each event type (AB study, BC study, AC inference test trials). We did not limit this analysis to particular trials in the experiment (e.g., based on subsequent memory); all study and test trials were included to avoid introducing bias into the across-participant correlations. As we performed two correlations in each of five hippocampal subfield ROIs, our Bonferroni-corrected critical  $p$ -value was 0.005.

We related BC-AC and AB-AC study-test pattern similarity in hippocampal subfields to inference performance across participants using Pearson correlation. We observed a significant positive relationship between BC-AC study-test pattern similarity and AC performance across participants in CA<sub>1</sub> ( $r = 0.62$ ,  $p = 0.003$ ; Fig. 4c). In contrast, there was no relationship between AB-AC study-test pattern similarity and AC performance ( $r = 0.03$ ,  $p = 0.89$ ). A follow-up direct comparison between these two correlation values revealed that they were significantly different from one another (Hotelling-Williams test;  $t(20) = 3.73$ ,  $p < 0.01$ ). No significant relationship was observed between study-test pattern similarity (BC-AC or AB-AC) and inference performance in any other hippocampal subfield (all  $r < 0.43$ ,  $p > 0.05$ ) (Fig. 5a-d, bottom scatterplots).

### Pattern similarity measures do not parallel differences in study-test lag

In the current task design, AB study trials are by definition presented before BC study trials. Therefore, the temporal lag between AB study trials and AC test trials is necessarily longer than the lag between BC study trials and AC test trials. To address the possibility that observed pattern similarity differences might simply reflect discrepancies in intervening time between study and test among our conditions of interest, we performed a control study-test lag analysis. We calculated the average lag in seconds between study and test trials for each participant and condition included in the performance-based analysis (not considering the lag between the study run and the test run). The results demonstrated that pattern similarity did not track with this temporal distance measure. Specifically, despite the large differences in pattern similarity as a function of performance (see Fig. 4a; Fig. 5a-d, top charts), study-test lags did not differ between AC correct and AC incorrect trials ( $F(1, 20) = 0.61$ ,  $p = 0.44$ ). In contrast, large differences in average study-test lag for BC and AB study trial types, dictated by our task design, did not substantially impact our similarity measures in any subfield.

## Discussion

Prevailing views suggest that novel inference relies on the ability to link distinct events across time (Cohen and Eichenbaum, 1995). One process that might support this ability is retrieval of related content during new learning. It has been speculated that overlapping events elicit associative novelty signals (Shohamy and Wagner, 2008; van Kesteren et al., 2012) in CA<sub>1</sub> (Wang and Morris, 2010; Duncan et al., 2012; Larkin et al., 2014), which in turn enable new experiences to be encoded into existing memory traces. Our data provide converging evidence from both univariate and multivariate approaches in support of this idea. We found evidence of a specialized encoding mechanism engaged during learning of overlapping events that supports later inference exclusively in right CA<sub>1</sub>. Moreover, we used neural pattern similarity analysis to demonstrate that CA<sub>1</sub> study patterns were reinstated during the inference test, particularly when inference judgments were fast and accurate.

In the present paradigm, events that overlap with prior memories (i.e., BC pairs) provide the unique opportunity to combine current and previous experiences. Thus, we reasoned that associative novelty signals would be possible only during BC study trials. Moreover, such signatures may be associated with superior inference performance on a trial-by-trial basis. Our findings are consistent with the hypothesized role of CA<sub>1</sub> in associative novelty detection. That is, CA<sub>1</sub> was the only region to demonstrate a greater subsequent inference effect during encoding of new events that partially overlapped with prior experiences (BC pairs) relative to the initially experienced events themselves (AB pairs). We underscore that this signature was observed only in the right hemisphere. Interestingly, this converges with a previous standard-resolution study using this paradigm (Zeithamova and Preston, 2010), in which this unique encoding signature related to subsequent inference was also specific to right hippocampus. Right-lateralized hippocampal effects have also been reported in studies of prospection (Addis et al., 2007, 2011; Weiler et al., 2010 a; b; Martin et al., 2011), which require similar cognitive processes—recombining prior experiences—in service of imagining future scenarios.

Using neural pattern similarity analysis, we also showed that CA<sub>1</sub> is similarly engaged during overlapping event encoding and successful subsequent inferences. This relationship held across participants, with greater BC, but not AB, study-test pattern similarity relating to superior inference performance. Moreover, this was driven by those inference decisions made quickly. Interestingly, these results were also found exclusively in CA<sub>1</sub>, consistent with hypotheses regarding this area's special role in comparing prior memories with current experience.

On a mechanistic level, our neural pattern similarity analysis findings might reflect test-phase reinstatement of memory representations evoked during BC study trials (Zeithamova et al., 2012 a). In other words, A, B, and C information is represented in the brain during BC learning and later reactivated during inference to answer the novel AC questions. This interpretation comes with the caveat that the present findings do not provide evidence for reinstatement of specific items at any point in time, as all pattern similarity measures pooled across multiple trials. Accordingly, a more likely interpretation of the present findings is that similarity of BC encoding patterns with AC inference within CA<sub>1</sub> reflects processes

common to BC study and AC test that do not occur during initial AB study. Two likely candidates for such processes are retrieval (i.e., recalling task-relevant memories) and memory-based comparison (i.e., signaling deviations between stored memories and current experience). Critically, if the present results do reflect engagement of such processes, they are processes associated with important behavioral benefits in inference speed and accuracy. Thus, regardless of whether our findings reflect reinstatement of integrated memories or engagement of common cognitive processes, these data demonstrate that CA<sub>1</sub> plays a unique and specialized role in encoding mechanisms that support the extension of memory beyond direct experience.

Existing research suggests that there are at least two general mechanisms that may work in a complementary fashion to support flexible memory behaviors such as inference (Zeithamova et al., 2012 b). First, information acquired at different times may be combined to make novel inferences during the inferential judgment itself. Under this hypothesized mechanism, individual pattern-separated memory representations are retrieved, manipulated and recombined to address novel questions. This intuitive explanation has received support from a number of fMRI studies (Acuna et al., 2002; Heckers et al., 2004; Preston et al., 2004; Zalesak and Heckers, 2009; Zeithamova and Preston, 2010), generally on the basis of greater activation in medial temporal and prefrontal regions during inferential judgments requiring consideration of multiple memories relative to memory judgments about single events. In the present paradigm, this framework might suggest that AB memories are retrieved and separately strengthened during BC study, thus leading to better inference.

However, recent work suggests that in some cases, effortful retrieval-based inference can be bypassed through engagement of a different mechanism that operates solely during encoding. In this process, alternatively known as retrieval-mediated learning (Holland, 1981; Hall, 1996; Iordanova et al., 2011; Zeithamova et al., 2012 a) or integrative encoding (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010), it has been suggested that existing memory networks are updated with new information *while that new content is being experienced*. By initially encoding new information into existing memories, this process is thought to result in hippocampal memory traces that bridge temporally disparate events. Importantly, integrated memories code the relationships among elements in our environment—even those relationships we have not experienced firsthand. Such integrated representations would thus support inferences directly, requiring no further recombination of knowledge during the inference judgment itself. In the present study, an integrative encoding framework would predict that retrieval of AB information during BC study would lead to the formation of an integrated ABC memory representation, thus enabling direct extraction of the novel AC association.

Our data are compatible with both of these accounts. However, we argue that taken in the context of prior work, the results presented here more likely reflect engagement of an integrative encoding mechanism. Recent evidence implicates hippocampal encoding processes in the integration of related memories during learning (Shohamy and Wagner, 2008; Zeithamova and Preston, 2010; Zeithamova et al., 2012 a). For example, one study demonstrated that hippocampal activation specifically during overlapping event encoding was predictive of inference success (Zeithamova and Preston, 2010). Moreover, another

study (Shohamy and Wagner, 2008) provided behavioral evidence consistent with the idea that memories are integrated during learning, enabling direct extraction of inferential information at test. The authors showed that in successful participants, inference judgments were just as fast as retrieval of directly learned associations. These findings suggest that inferential trials required minimal extra processing during test, presumably because the link between indirectly related items was already established during encoding. While these studies suggest an integrative encoding process supported by the hippocampus, no study to date has localized this signature to a particular hippocampal subfield. Thus, through the combined use of univariate and multivariate analysis approaches, the present work represents the first empirical evidence highlighting the role of CA<sub>1</sub> in memory integration and inference.

Notably, our results are not readily explained by either memory strength or temporal context accounts. It is true that BC study trials are closer in time to AC test trials than are AB study trials, which might influence the neural pattern similarity analysis in a number of ways. One possible consequence of this design is that BC memories may be stronger (i.e., easier to retrieve) because they were encoded more recently than were AB memories. However, we found significantly better memory for AB than BC associations, suggesting that BC memories are actually *weaker* by comparison. This finding rules out a memory strength account. A second possible consequence relates to differences in temporal context. In other words, neural patterns evoked during AC test may more closely reflect BC than AB study patterns simply because BC information was encoded in a more similar temporal context. To this point, it is important to note that study and test phase data were collected in separate scans, thus minimizing the effects of low-level factors such as high within-scan temporal autocorrelation. Nevertheless, it is conceivable that differences in trial timing could give rise to greater BC-AC than AB-AC study-test similarity due to higher-level factors like *cognitive* context. To rule out this possibility, we performed a control study-test lag analysis. We demonstrate that our pattern similarity results do not track with the temporal distance between study and test, rendering an account based purely on lag improbable. While these factors may certainly contribute to our pattern similarity measures, they cannot fully explain our findings. Thus, we believe that the pattern similarity data reported here reflect the true engagement of a common neural state during BC study and AC test.

Interestingly, the memory integration signatures observed in CA<sub>1</sub> in the present study converge with studies of rodent hippocampus. First, recent work has implicated CA<sub>1</sub> processes in both detecting novelty in the environment as well as enabling memory updating via increases in plasticity (Larkin et al., 2014). This finding is consistent with an interpretation of the present results as reflecting CA<sub>1</sub>-mediated integration of new content into existing memories during study. Additional work has highlighted the importance of CA<sub>1</sub> for nodal coding, in which shared content is represented by the same population of neurons across distinct events (Wood et al., 2000; Singer et al., 2010; McKenzie et al., 2013). For instance, one study demonstrated that while the firing patterns of some CA<sub>1</sub> cells reflected individual episode representations, others fired similarly across different types of episodes that shared content (i.e., a common spatial location) (Wood et al., 2000). Another study showed that in environments with overlapping elements (related locations), a subset of



CA<sub>1</sub> and CA<sub>3</sub> neurons responded similarly to the related locations both within and across environments (Singer et al., 2010). Together, these studies suggest that hippocampal neurons in the CA fields can develop generalized firing patterns that encode similarities across episodes. Consistent with this idea, recent work demonstrated that some CA<sub>1</sub> neurons respond similarly to multiple spatial sequence locations learned in a single environment, further evidence that this region may code for “nodes,” or commonalities across experiences (McKenzie et al., 2013). Our results are consistent with such a nodal coding scheme, in which integrated memories representing shared B elements across AB and BC learning episodes are formed in CA<sub>1</sub> during learning. Moreover, the data presented here build upon the existing animal literature to demonstrate the potential behavioral significance of nodal coding—specifically, that integrated memories may support novel judgments, allowing for rapid and appropriate action in the absence of direct experience.

## Conclusions

The present results suggest that area CA<sub>1</sub> plays a specialized role during the encoding of overlapping information. Utilizing new methodologies—high-resolution fMRI in combination with neural pattern similarity analysis—we provide a direct comparison of the neural states engaged during study of overlapping events and subsequent inference test. These methods allow us to take advantage of distributed patterns of hippocampal activation reflecting important content and process, thereby suggesting a mechanistic account of the hippocampal subfield contributions to overlapping event encoding. More broadly, our data are consistent with the notion from the animal literature that CA<sub>1</sub> signals deviations of current events from memory-based expectations, allowing for the construction of memory representations that code the relationships among multiple experiences. The formation of such integrated memory representations supports flexible judgments like novel inference, reflecting the extension of memory beyond direct observation. Moreover, the present results reinforce the idea that encoding and retrieval do not occur in isolation, but rather are highly interactive processes.

## Acknowledgments

We thank Nicolaus Schmandt for help with data collection and analysis.

### Funding

Grant sponsor: National Institute of Mental Health of the National Institutes of Health; Grant number: R01MH100121 (ARP)

Grant sponsor: National Science Foundation CAREER; Grant number: 1056019 (ARP)

Grant sponsor: Army Research Office; Grant number: 55830-LS-YIP (ARP)

Grant sponsor: National Institute of Mental Health of the National Institutes of Health;

Grant number: NRSA F32MH094085 (DZ)

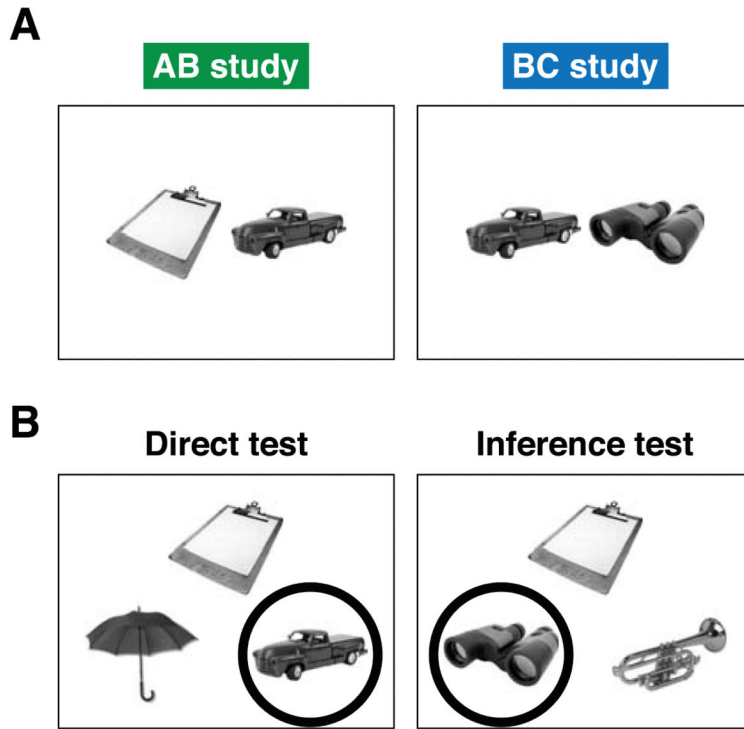
Grant sponsor: Department of Defense NDSEG (MLS)

## References

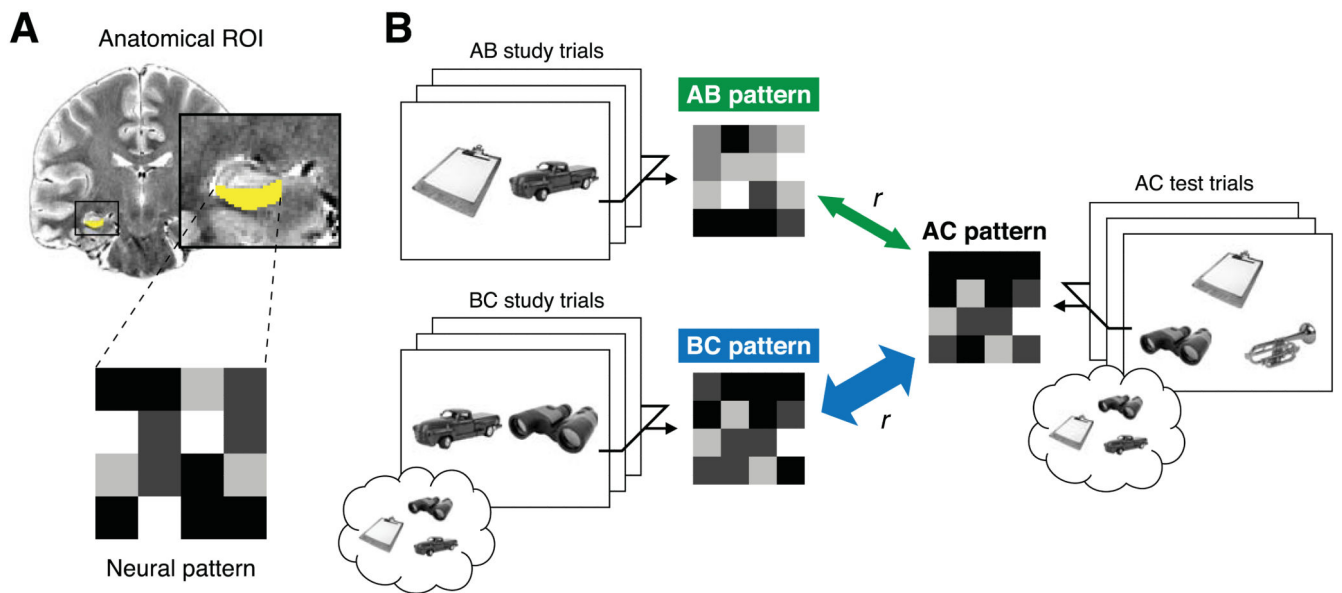
- Acuna BD, Eliassen JC, Donoghue JP, Sanes JN. Frontal and parietal lobe activation during transitive inference in humans. *Cereb Cortex*. 2002; 12:1312–1321. [PubMed: 12427681]
- Addis DR, Cheng T, Roberts RP, Schacter DL. Hippocampal contributions to the episodic simulation of specific and general future events. *Hippocampus*. 2011; 21:1045–52. [PubMed: 20882550]
- Addis DR, Schacter DL. The hippocampus and imagining the future: where do we stand? *Front Hum Neurosci*. 2012; 5:1–15.
- Addis DR, Wong AT, Schacter DL. Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*. 2007; 45:1363–77. [PubMed: 17126370]
- Alvarez P, Squire LR. Memory consolidation and the medial temporal lobe: a simple network model. *Proc Natl Acad Sci U S A*. 1994; 91:7041–5. [PubMed: 8041742]
- Amaral, DG.; Insausti, R. Hippocampal formation. In: Paxinos, G., editor. *The Human Nervous System*. Academic Press; San Diego: 1990. p. 711-755.
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011; 54:2033–44. [PubMed: 20851191]
- Bartlett, F. *Remembering: A Study in Experimental and Social Psychology*. : Cambridge University Press; Cambridge: 1932.
- Buckner RL. The role of the hippocampus in prediction and imagination. *Annu Rev Psychol*. 2010; 61:27–48. [PubMed: 19958178]
- Chen J, Olsen RK, Preston AR, Glover GH, Wagner AD. Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. *Learn Mem*. 2011; 18:523–528. [PubMed: 21775513]
- Cohen, N.; Eichenbaum, H. *Memory, amnesia, and the hippocampal system*. The MIT Press; Cambridge, MA: 1995.
- Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996; 29:162–73. [PubMed: 8812068]
- Dale AM. Optimal experimental design for event-related fMRI. *Hum Brain Mapp*. 1999; 8:109–114. [PubMed: 10524601]
- Duncan K, Ketz N, Inati SJ, Davachi L. Evidence for area CA1 as a match/mismatch detector: a high-resolution fMRI study of the human hippocampus. *Hippocampus*. 2012; 22:389–98. [PubMed: 21484934]
- Eichenbaum H. The hippocampus and mechanisms of declarative memory. *Behav Brain Res*. 1999; 103:123–33. [PubMed: 10513581]
- Eichenbaum H. A cortical-hippocampal system for declarative memory. *Nat Rev Neurosci*. 2000; 1:41–50. [PubMed: 11252767]
- Glover GH, Law CS. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magn Reson Med*. 2001; 46:515–22. [PubMed: 11550244]
- Hall G. Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. *Anim Learn Behav*. 1996; 24:233–255.
- Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D. Hippocampal activation during transitive inference in humans. *Hippocampus*. 2004; 14:153–62. [PubMed: 15098721]
- Holland PC. Acquisition of representation-mediated conditioned food aversions. *Learn Motiv*. 1981; 12:1–18.
- Hotelling H. The Selection of Variates for Use in Prediction with Some Comments on the General Problem of Nuisance Parameters. *Ann Math Stat*. 1940; 11:271–283.
- Insausti R, Juottonen K, Soinen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkänen A. MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *Am J Neuroradiol*. 1998; 19:659–71. [PubMed: 9576651]
- Jordanova MD, Good M, Honey RC. Retrieval-mediated learning involving episodes requires synaptic plasticity in the hippocampus. *J Neurosci*. 2011; 31:7156–62. [PubMed: 21562278]

- Van Kesteren MTR, Ruiters DJ, Fernández G, Henson RN. How schema and novelty augment memory formation. *Trends Neurosci.* 2012; 35:211–219. [PubMed: 22398180]
- Klein SB, Loftus J, Kihlstrom JF. Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Soc Cogn.* 2002; 20:353–379.
- Kriegeskorte N, Mur M, Bandettini P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci.* 2008; 2:1–28. [PubMed: 18958245]
- Larkin M, Lykken C, Tye L, Wickelgren JG, Frank LM. Hippocampal output area CA1 broadcasts a generalized novelty signal during an object-place recognition task. *Hippocampus.* 2014
- Lisman JE, Grace AA. The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron.* 2005; 46:703–13. [PubMed: 15924857]
- Manns J, Eichenbaum H. Evolution of declarative memory. *Hippocampus.* 2006; 16:795–808. [PubMed: 16881079]
- Marr D. Simple Memory: A Theory for Archicortex. *Philos Trans R Soc London.* 1971; 262:23–81. [PubMed: 4399412]
- Martin VC, Schacter DL, Corballis MC, Addis DR. A role for the hippocampus in encoding simulations of future events. *Proc Natl Acad Sci U S A.* 2011; 108:13858–63. [PubMed: 21810986]
- McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev.* 1995; 102:419–457. [PubMed: 7624455]
- McKenzie S, Robinson NTM, Herrera L, Churchill JC, Eichenbaum H. Learning causes reorganization of neuronal firing patterns to represent related experiences within a hippocampal schema. *J Neurosci.* 2013; 33:10243–10256. [PubMed: 23785140]
- McNaughton BL, Morris RGM. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* 1987; 10:408–415.
- Moscovitch M, Rosenbaum RS, Gilboa A, Addis DR, Westmacott R, Grady C, McAndrews MP, Levine B, Black S, Winocur G, Nadel L. Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *J Anat.* 2005; 207:35–66. [PubMed: 16011544]
- Norman KA, O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol Rev.* 2003; 110:611–46. [PubMed: 14599236]
- O'Keefe, J.; Nadel, L. *The Hippocampus as a Cognitive Map.* Clarendon; London: 1978.
- O'Reilly RC, Rudy JW. Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychol Rev.* 2001; 108:311–345. [PubMed: 11381832]
- Pfeuffer J, Van de Moortele P-F, Ugurbil K, Hu X, Glover GH. Correction of physiologically induced global off - resonance effects in dynamic echo - planar and spiral functional imaging. *Magn Reson Med.* 2002; 47:344–353. [PubMed: 11810679]
- Polyn SM, Natu VS, Cohen JD, Norman KA. Category-specific cortical activity precedes retrieval during memory search. *Science.* 2005; 310:1963–6. [PubMed: 16373577]
- Preston AR, Bornstein AM, Hutchinson JB, Gaare ME, Glover GH, Wagner AD. High-resolution fMRI of content-sensitive subsequent memory responses in human medial temporal lobe. *J Cogn Neurosci.* 2010; 22:156–173. [PubMed: 19199423]
- Preston AR, Shrager Y, Dudukovic N, Gabrieli JDE. Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus.* 2004; 14:148–152. [PubMed: 15098720]
- Pruessner JC, Köhler S, Crane J, Pruessner M, Lord C, Byrne A, Kabani N, Collins DL, Evans AC. Volumetry of entorhinal and parahippocampal cortex from high-resolution MR images: Considering the variability of the collateral sulcus. *Cereb Cortex.* 2002; 12:1342–1353. [PubMed: 12427684]
- Pruessner JC, Li LM, Serles W, Pruessner M, Collins DL, Kabani N, Lupien S, Evans AC. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software:

- Minimizing the discrepancies between laboratories. *Cereb Cortex*. 2000; 10:433–42. [PubMed: 10769253]
- Shohamy D, Wagner AD. Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron*. 2008; 60:378–89. [PubMed: 18957228]
- Singer AC, Karlsson MP, Nathe AR, Carr MF, Frank LM. Experience-dependent development of coordinated hippocampal spatial activity representing the similarity of related locations. *J Neurosci*. 2010; 30:11586–604. [PubMed: 20810880]
- Stark CEL, Squire LR. When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proc Natl Acad Sci U S A*. 2001; 98:12760–6. [PubMed: 11592989]
- Tolman EC. Cognitive maps in rats and men. *Psychol Rev*. 1948; 55:189–208. [PubMed: 18870876]
- Tulving E, Thomson D. Encoding specificity and retrieval processes in episodic memory. *Psychol Rev*. 1973; 80:352–373.
- Wang S-H, Morris RGM. Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annu Rev Psychol*. 2010; 61:49–79. [PubMed: 19575620]
- Weiler JA, Suchan B, Daum I. When the future becomes the past: differences in brain activation patterns for episodic memory and episodic future thinking. *Behav Brain Res*. 2010a; 212:196–203. [PubMed: 20399235]
- Weiler JA, Suchan B, Daum I. Foreseeing the future: occurrence probability of imagined future events modulates hippocampal activation. *Hippocampus*. 2010b; 20:685–90. [PubMed: 19693779]
- Wheeler M, Petersen SE, Buckner RL. Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci U S A*. 2000; 97:11125–29. [PubMed: 11005879]
- Williams EJ. The Comparison of Regression Variables. *J R Stat Soc Ser B*. 1959; 21:396–399.
- Wood ER, Dudchenko PA, Robitsek RJ, Eichenbaum H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*. 2000; 27:623–33. [PubMed: 11055443]
- Zalesak M, Heckers S. The role of the hippocampus in transitive inference. *Psychiatry Res Neuroimaging*. 2009; 172:24–30.
- Zeineh MM, Engel SA, Bookheimer SY. Application of cortical unfolding techniques to functional MRI of the human hippocampal region. *Neuroimage*. 2000; 11:668–83. [PubMed: 10860795]
- Zeineh MM, Engel SA, Thompson PM, Bookheimer SY. Dynamics of the hippocampus during encoding and retrieval of face-name pairs. *Science*. 2003; 299:577–580. [PubMed: 12543980]
- Zeithamova D, Dominick AL, Preston AR. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*. 2012a; 75:168–79. [PubMed: 22794270]
- Zeithamova D, Preston AR. Flexible memories: Differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *J Neurosci*. 2010; 30:14676–14684. [PubMed: 21048124]
- Zeithamova D, Schlichting ML, Preston AR. The hippocampus and inferential reasoning: Building memories to navigate future decisions. *Front Hum Neurosci*. 2012b; 6:1–14. [PubMed: 22279433]

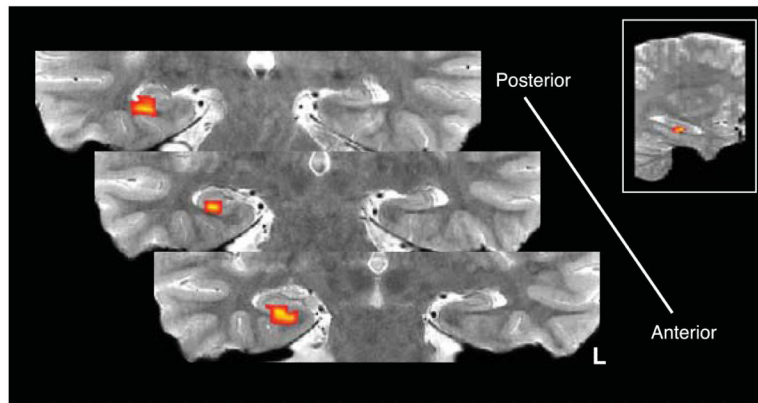


**Figure 1.** Associative inference task. (a) Participants learned overlapping pairs of objects during the study phases. AB (e.g., clipboard-truck) pairs were presented first. BC (e.g., truck-binoculars) pairs were learned later and included familiar items from the AB pairs (i.e., the truck in this example). (b) During test phases, participants were presented with three objects. The top item served as the cue; the bottom items were the two choices. A direct test trial is shown on the left, in which the participant is required to select truck when cued with clipboard. In the inference example (right), the participant should choose the binoculars, as both the clipboard and binoculars were paired with the truck during learning. For both direct and inference test trials, familiar items that were members of a different triad from the same study scan served as foils. Correct choices are circled for illustrative purposes only (not shown to participants).



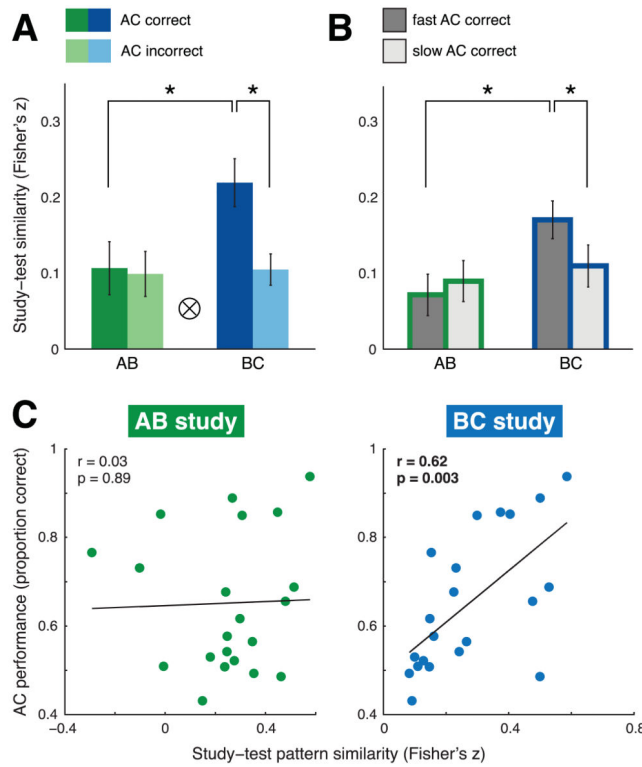
**Figure 2.**

Schematic depiction and rationale of neural pattern similarity analysis (NPSA). (a) Average patterns of activation associated with specific trial types were extracted for each anatomical ROI. Here we depict the cross-participant analysis (see *Neural pattern similarity analysis* section of *Materials and Methods*), which was estimated irrespective of memory performance. Trials were modeled according to event type (AB, BC for study phase; AC for test phase) using a GLM. Parameter estimates associated with conditions of interest were then extracted for each voxel within the ROI (example ROI shown in yellow). The intensity in each cell in the grayscale matrix schematic represents the parameter estimate for a single voxel in the brain. The similarity of two patterns was then determined by computing a Pearson correlation between the two matrices. (b) Predictions for NPSA when existing memories are retrieved and linked to current experience during overlapping event encoding. Example AB study, BC study, and inference test screens are shown; simplified hypothetical mean patterns of activation associated with each trial type are depicted next to the corresponding condition. As BC study trials provide a unique opportunity to link prior memories with current experience, we predicted greater reinstatement of BC than AB study patterns during AC inference judgments. Reinstatement of study patterns evoked during overlapping event encoding would be reflected by a higher correlation between BC-AC study-test (thick blue arrow) than between AB-AC study-test (thin green arrow).



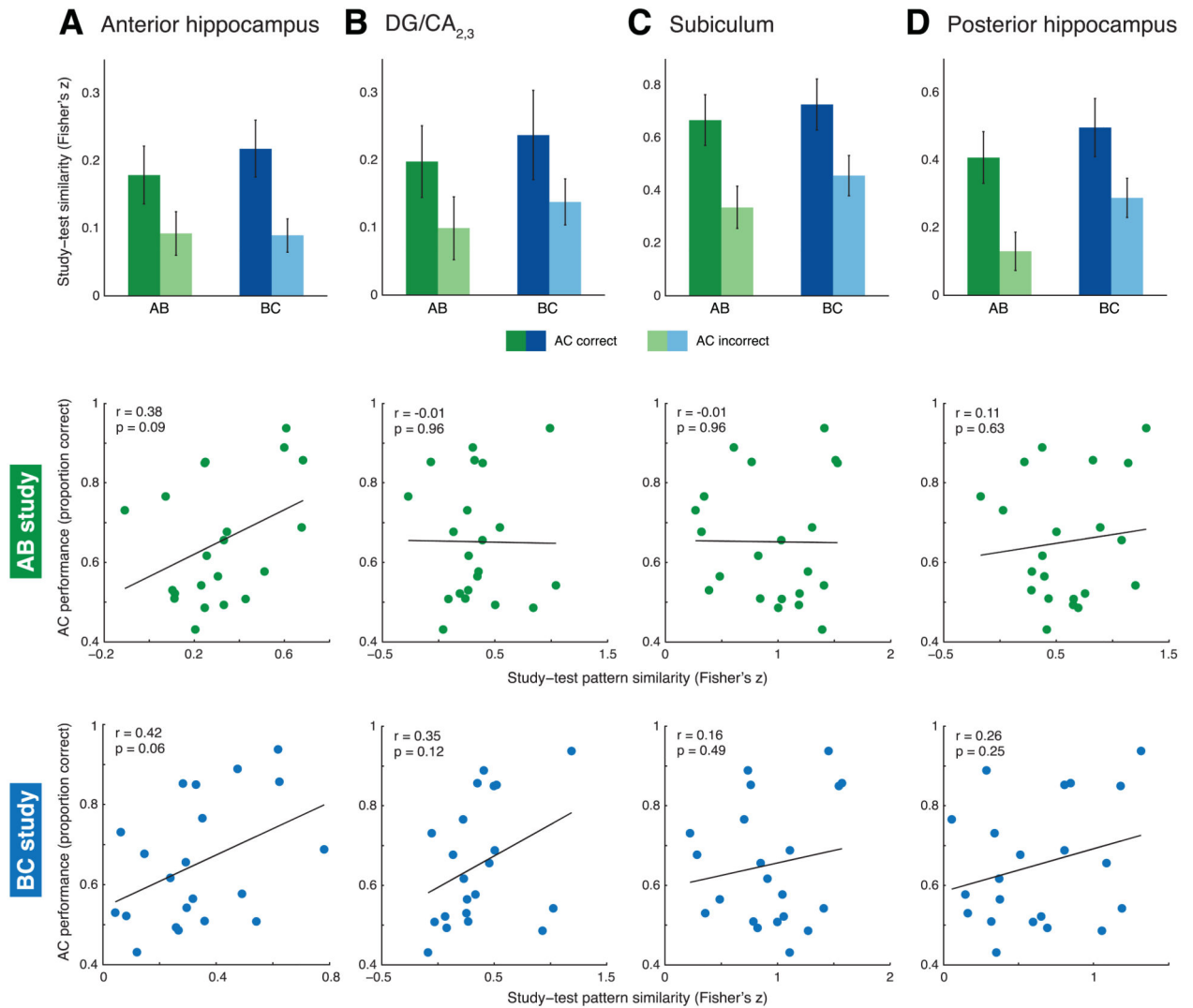
**Figure 3.**

Results from univariate analysis demonstrating relationship between CA<sub>1</sub> processes during overlapping event encoding and subsequent inference success. Right CA<sub>1</sub> was the only hippocampal region to demonstrate a signature consistent with study-phase retrieval of prior memories in service of later inference. This region showed greater subsequent inference effects for BC relative to AB study trials as a function of AC performance. Inset, sagittal view showing location of cluster along anterior-posterior extent of hippocampus. Significant activation was restricted to the hippocampal body. Activation map has been transformed to the space of a single participant's T2 coronal image for visualization purposes. Cluster is significant after correction for multiple comparisons (voxel threshold:  $p < 0.025$ , uncorrected; cluster size threshold:  $p < 0.05$ ).

**Figure 4.**

Results from multivariate neural pattern similarity analysis demonstrating an encoding signature specific to overlapping events in bilateral CA<sub>1</sub>. **(a)** Performance-based analysis. Study-test pattern similarity limited to triads for which both direct memory judgments (AB, BC) were correct. Trials were split based on whether the corresponding inference (AC) judgment was later correct (dark bars) or incorrect (light bars). AB-AC study-test pattern similarity is shown in green; BC-AC study-test pattern similarity is shown in blue. **(b)** RT-based analysis. Study-test pattern similarity for direct correct and AC inference correct triads. Trials were further split based on median reaction time into fast (dark gray bars) and slow (light gray bars) inference judgments. Bars depicting AB-AC study-test pattern similarity are outlined in green; bars for BC-AC pattern similarity are outlined in blue. For both (a) and (b), asterisks (\*) denote significant follow-up paired t-tests ( $p < 0.05$ ); tensor product symbol ( $\otimes$ ) denotes significant interaction ( $p < 0.01$ ). Error bars denote across-participant SEM. **(c)** Cross-participant analysis. Scatterplots depict continuous relationship between study-test pattern similarity (AB-AC study-test, green; BC-AC study-test, blue) and inference performance. Best-fit lines and statistics on plots were calculated using Pearson correlation. Significant correlation at  $p < 0.005$  denoted with bold type.





**Figure 5.**

Neural pattern similarity results in anterior hippocampus (a), DG/CA<sub>2,3</sub> (b), subiculum (c), and posterior hippocampus (d). None of these regions showed evidence of a specialized BC encoding mechanism. Data are presented as in Figure 4a and c. Top charts, performance based analysis. Significant main effects of inference performance were observed in anterior hippocampus, subiculum, and posterior hippocampus (all  $p < 0.01$ ; not marked on charts). Bottom scatterplots, cross-participant analysis depicting continuous relationship between study-test pattern similarity and inference performance. All correlations were non-significant at the critical  $p$  threshold of 0.005.