PNAS
www.pnas.org

Supplementary Information for

Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus

Lynn J. Lohnas, Katherine Duncan, Werner K. Doyle, Thomas Thesen, Orrin Devinsky, Lila Davachi

Lila Davachi
Email: ld24@columbia.edu

This PDF file includes:

Supplementary text
Figs. S1 to S6
Tables S1 to S4
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1717088115

Supplementary Information Text

Materials and Methods

Participants. Relevant clinical and demographic information is provided in Table S2.

Electrophysiology. *Recording.* ECoG activity was recorded continuously during both blocks. Each participant had grid, strip and depth electrodes, and electrode placements were determined based on clinical criteria. ECoG activity was recorded with a custom built neural recording system (NSpike) at 10kHz. Sync pulses sent at the onset of each stimulus presentation and each participant response allowed for alignment of the ECoG data with trial onsets as well as behavioral responses.

Electrode localization. Hippocampal electrodes were manually identified with individual patient's post-implantation magnetic resonance (MR) images using visual inspection of synchronized axial, coronal and sagittal slices (according to (1); see e.g. Fig. 1C,D). We first defined the posterior border by the first slice where gray matter appeared inferior and medial to the lateral ventricle. Then, moving anteriorly, wherever possible we used the landmarks of the lateral ventricle, white matter, and uncal recess to inform the borders. Across participants, electrodes were identified in the hippocampal head, body, and tail, yet there were insufficient electrodes to consider these subregions separately.

In addition to the hippocampus, we analyzed data from two additional regions: dorsolateral prefrontal cortex (DLPFC) and occipitotemporal cortex (OTC). DLPFC electrodes were identified as any electrodes in the middle frontal gyrus (anterior to premotor cortex, i.e. Brodmann areas 9 and 46; (2)) based on visual inspection of each participant's reconstructed 3D cortical surface using pre- and post-implantation MR scans (3). OTC electrodes were identified using a combination of the MR images and 3D brain reconstructions. This ROI was bounded superiorly by the inferior temporal sulcus and posteriorly by the occipital lobe, guided by the parieto-occipital fissure and the temporo-occipital incisures (2), anteriorly by the hippocampal tail (1), and ventrally by the occipitotemporal sulcus (4). The number of electrodes per region of interest is provided in Table S3.

Preprocessing. Data were downsampled to 300 Hz and each electrode was referenced to the mean activity across all of the patient's electrodes, with the mean weighted such that each grid, strip or depth contributed equally (5). To remove electrical line noise, data were filtered at 60 Hz with a fourth order 2 Hz stopband Butterworth notch filter.

Analysis. *High frequency activity (HFA) univariate power.* We calculated the mean HFA power over time for four non-overlapping 500ms time bins. A Wilcoxon rank sum test was then used to compare pairs of conditions separately for each participant, electrode, frequency band and time bin. To determine significance between pairs of conditions, we used the summed Z method, an approach meant to assess significance with many observations per participants but few participants (6–8). With this approach, an empirical Z value is obtained from the experimental data (reported as the actual Z in the Results) and compared to a null distribution obtained using a permutation procedure (reported as the null mean Z in the Results, taken from 1000 random shuffles of the labels for each condition, with the same shuffled labels across all time bins for a given region of interest (ROI) and subject). Briefly, after calculating the statistic for each participant electrode, Z values and null distributions are then determined within each participant, across all electrodes in a given ROI. In this way, the null distributions are defined based on each region. Across participants, the point at which the empirical Z score fell in the region-specific null distribution determined the p value between conditions. Unless noted otherwise in the text, reported p values are Bonferroni corrected for the number of time windows.

Spatiotemporal pattern similarity (STPS) across conditions. In a similar way to the approach of comparing STPS for a particular condition to baseline (e.g. correct old/correct new), when comparing across conditions, we calculated the difference between the values of the matched pairs (e.g. between two conditions or the difference between the differences of conditions). We compared this to the difference between the values for the unmatched pairs in the null distribution, such that the point at which the actual matched difference fell on the null difference distribution determined the p values. In the Results, we report the mean STPS values from the empirical and null distributions.

Correlation between univariate HFA and HFA STPS. For each participant, we then examined whether univariate HFA was related to STPS. Specifically, hippocampal HFA STPS was calculated for matched old-new pairs in each of the 500ms time bins. We then took the trial-by-trial correlation of the HFA pattern similarity difference values with HFA univariate activity during the same time bin as the STPS (i.e. for each of the 500ms bins). Next, for each participant we calculated an expected null distribution of correlation values by randomly shuffling the trial labels of HFA pattern similarity values, and taking the correlation of these shuffled pairs, for 200 shuffles of the data. The p value was determined by where the actual mean correlation fell on the mean shuffled distribution, and was Bonferroni corrected. In the Results, we report the mean STPS values from the empirical and null distributions.

Results

Response times (RTs) by task and condition. We included behavioral performance and RT by condition as a useful reference to the related ECoG measures (Fig. S1). As noted in the main text, RTs were faster in the fine-grain task than in the coarse-grain task.

Encoding across tasks. Our primary comparisons between the fine-grain and coarse-grain task always included items being presented for the second time, either as exact repeats from their first presentations at encoding (old) or highly similar but not identical to their first presentations (similar). However, it is important to ensure that the differences from these second presentations are not simply an artifact of differences between their first presentations during encoding. To examine potential differences during encoding between tasks, for each region we compared HFA of correctly classified new items (i.e., during initial item encoding) between task types (fine-grain vs. coarse-grain). We only include correctly classified new items because neural activity of incorrect new items was never considered in our analyses, nor did all participants have incorrect new trials in both tasks. As shown in Fig. S2, there were no significant differences between tasks in any time window or region (p's>.5). This suggests that the differences (or lack thereof) across tasks during retrieval cannot be explained by differences during encoding.

STPS in DLPFC. We calculated STPS in DLPFC for the same five conditions as reported in the main text for hippocampus and OTC (Fig. S3). However, STPS in this region was not significantly different than the shuffled baseline in any condition for any time bin (all p's>.06).

Contributions of univariate HFA to STPS. It is important to control for univariate differences between conditions when calculating multivariate patterns of activity across conditions. To this end, for each pattern of univariate activity considered in the STPS analyses, we subtracted the mean univariate HFA for that pattern's condition (e.g., fine-grain old item). However, prior work suggests that HFA may still bias STPS results (9). To determine whether this was an issue in our STPS results, we created a set of regression models, one for each condition and time bin where we found a significant difference between HFA of first presentations of items (as new) and HFA of second presentations (as old or similar). For each of these models, our goal was to assess whether univariate HFA contributed significantly to STPS. This may mean that a significant STPS effect may be driven by, or obfuscated by, a significant univariate difference. Thus, we considered these models irrespective of whether STPS was significant during the time bin of interest. In these models, STPS was the dependent measure, with one observation for each item's matched presentations. We considered as random effects: univariate HFA from the first presentation, univariate HFA from the second presentation, an interaction term between the two univariate HFA measures, and participant. To assess the significant contributions of univariate HFA to each model, we compared each model to a null model where only participant was a random effect. (Note that it is not appropriate to use a stepwise model here, as the univariate HFA measures for the same item are not independent.) Table S1 includes the p value between each null model and the univariate model. None of the univariate models were significantly different from their corresponding null models, thus mitigating concerns that significant differences in univariate HFA may bias the STPS results.

Effect of study-retrieval lag on behavior and neural activity. In our experimental tasks, participants were presented with (a version of) every item twice, with its first presentation as a new item and its second presentation as an exact repeat (old) or a non-identical but highly similar item (similar). The lag between first and second presentations varied between 1-8 items. In all analyses reported in the Results, we collapsed analyses across all items and thus across all lag values. We performed several analyses to ensure that the reported differences were not confounded by lag.

Behavior. To examine whether the intervening lag between first and second presentations impacted recognition accuracy, we conducted a repeated measures ANOVA (rmANOVA) with accuracy as the dependent variable, with stimulus type (old or similar) and lag (1,2,…,8) as factors. (One participant who did not complete the coarse-grain task did not have any items with a lag=7, and thus this lag was not considered in the rmANOVA for the coarse-grain task.) We found no main effect of lag (p’s>.1), nor an interaction between lag and stimulus type on memory (p’s>.09; Fig. S4A). We also considered memory accuracy when binned into shorter and longer lags (lag=1,2 vs. lag=3,4,5,6,7,8), paralleling the neural analyses described below. Memory accuracy did not differ by shorter or longer lags in either task, for either old or similar items (all p’s>.2).

Univariate HFA. To examine whether HFA and STPS varied by lag, we wanted to ensure that for each participant and condition there were a sufficient number of observations at each lag value. Unlike the behavioral analyses above, which included all presented stimuli, all of the neural analyses included items only if they were correctly classified as new during their first presentations. Further, most analyses only included items that were correctly classified during their second presentations. With this more limited data, there was not sufficient data to assess neural activity at every lag. Instead, for items correctly classified during their second presentations, we aggregated observations into bins of shorter and longer lag values. Because more items were presented at shorter lags than longer lags, to have a sufficient number of observations per lag bin, we defined shorter lags as lag=1,2 and longer lags as lag=3,4,5,6,7,8. For incorrect items (namely, similar items classified as ‘old’ in the fine-grain task), participants made sufficiently few errors that we could not divide the data further into different lag values, and thus there were not enough observations to perform these analyses. Nonetheless, given that memory performance did not vary by lag, we are less concerned about contributions of lag to these error trials. We compared HFA between shorter vs. longer lags, for all of the correct stimulus/response conditions reported in the Results, across regions and time bins. We found no significant differences between conditions based on lag.

STPS. We also examined whether the lag between first and second presentations of items influenced the STPS measures. Paralleling the analyses in HFA, we compared STPS between shorter lags (lag=1,2) and longer lags (lag=3,4,5,6,7,8), for all reported correct stimulus/response conditions, across tasks, regions of interest and time bins. We found one significant difference between conditions based on lag (Fig. S4B): STPS in OTC during 1.5-2s of the coarse-grain task was *significantly greater for similar items with longer lags than shorter lags* (p<.001, actual mean=.0785, null mean=-.0068). A post-hoc test of STPS during this time bin revealed significantly greater STPS than baseline for items with longer lags (p=.035, actual mean=.0561, null mean=-.001), but there was no significant difference in STPS from baseline for items with shorter lags (p>.5). This is somewhat surprising, given that more recently presented information is better remembered, and so one may expect items with *shorter* lags to evoke stronger reinstatement and thus greater STPS. Instead, it is tempting to speculate that greater STPS with longer lags might reflect more effortful retrieval for more distant memories.

Collapsing across all lags, as in the main text (Fig. 5B), we did not find a significant difference in STPS from baseline in this region, time bin and condition. This suggests that STPS does interact with lag, and collapsing across lag occluded the fact that this effect is significant at longer lags but not shorter lags. Nonetheless, our findings of significant reinstatement in STPS for old and similar items across tasks, as well as similar items classified as old in the fine-grain task, suggests that the OTC reinstatement effects are not modulated by lag.

Significance of effects by participant. It is important to consider the contributions of individual participants to significant effects, especially with a small number of participants. Therefore, for each significant effect reported in the Results, we examined the significance of each effect at the participant level for HFA (Fig. S5, Table S4) and STPS (Fig. S6, Table S4). Recall that for HFA and STPS, we calculated a null distribution for each participant and ROI. Thus, within each ROI we could determine where each participant's observed value fell on his/her null distribution to calculate a p value. For plotting purposes, we convert the p value to Z values. Critically, for all but one HFA effect and two STPS effects, there was at least one participant who exhibited a significant effect on an individual level. For those effects where no one participant exhibited a significant effect, participants had values clustered near a trending value, and thus this tighter distribution yielded an across-participant significant difference. In addition, none of the STPS or HFA effects are driven by a single outlier participant. Further, different participants are more significant across different conditions, regions and analysis type. Thus, although not every participant exhibits significant effects for every analysis, there is minimal concern that the mean-level differences do not reflect the distribution of values across participants.

Mnemonic reinstatement in OTC. When measuring reinstatement in a task where a similar version of an item presented during encoding is also presented during retrieval, this raises the question of whether reinstatement might just reflect the overlap in perceptual processing between similar stimuli. This is of particular concern in OTC, as STPS in this region was significantly above baseline irrespective of whether similar items were classified as similar or old, and STPS was above baseline for correct old and similar items in both tasks. If STPS in OTC reflects the perceptual overlap between stimuli, rather than mnemonic operations, we would not expect OTC STPS to differ based on participants' memory decision. To this end, we contrasted STPS for all old and similar items based on their classification as (a) indicating an awareness that the item was being presented for the second time, i.e. classified as 'old' or 'similar' in the fine-grain task, or as 'old' in the coarse grain task; (b) not indicating an awareness that the item was being presented for the second time, i.e. classified as new. Further, all items had to be classified as new during their first presentation. We collapsed across task and trial types in order to acquire as many incorrect trials as possible, as accuracy was relatively high (see Fig. 1A,B; one participant did not have enough incorrect trials to be included in this analysis). This approach is reasonable given that OTC STPS is not significantly different between tasks, or between old and similar items (see Results).

If STPS in OTC only reflected a perceptual similarity between the two presentations, we would not expect STPS to differ between conditions a and b above. However, there was significant reinstatement from 0-0.5s and 0.5-1s when participants responded 'old' or 'similar' (0-.5s: p<.001, actual mean=.0545, null mean=-.001; 0.5-1s: p<.001, actual mean=.0425, null mean=.0003), but no reinstatement for items to which they responded 'new' (p's>.5). Critically, post-hoc comparisons on the two time windows revealed that STPS was significantly less for items classified as new, during 0-0.5s (p=.040, actual mean=.1071, null mean=.0061) but not 0.5-1s (p>.5). Thus, there is greater STPS for second presentations of old or similar items not classified as new, which is more consistent with a memory-related effect of STPS in OTC. Moreover, this significant effect occurs during 0-0.5s, the same bin as when OTC STPS is significantly greater than baseline for correctly classified old or similar items in both tasks, as well as similar items classified as ‘old’ in the fine-grain task. Taken together these results suggest that OTC STPS reflects mnemonic processes not explained by the perceptual similarity between an item’s first and second presentations.

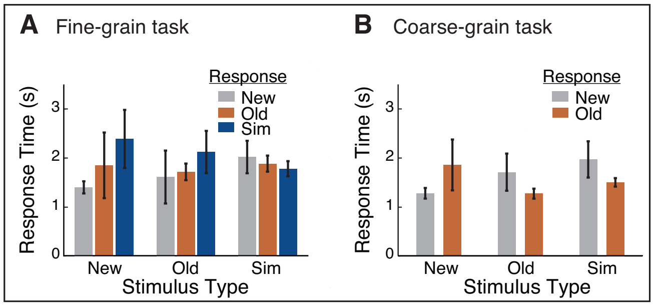

Fig. S1. Response times by stimulus type and response type. (A) Response times in the fine-grain task. N=5 except for the following Stimulus Type-Response conditions, as not all participants made errors, and thus not all Stimulus Type-Response conditions were realized in all participants: New-“Sim”, N=4; Sim-“New”, N=4; Old-“New”, N=3; New-“Old”, N=2. (B) Response times in the coarse-grain task. N=5 except for the following Stimulus Type-Response conditions, as not all participants made errors, and thus not all Stimulus Type-Response conditions were realized in all participants: New-“Old”, N=4; Old-“New”, N=4. Error bars indicate mean ±SEM across participants.

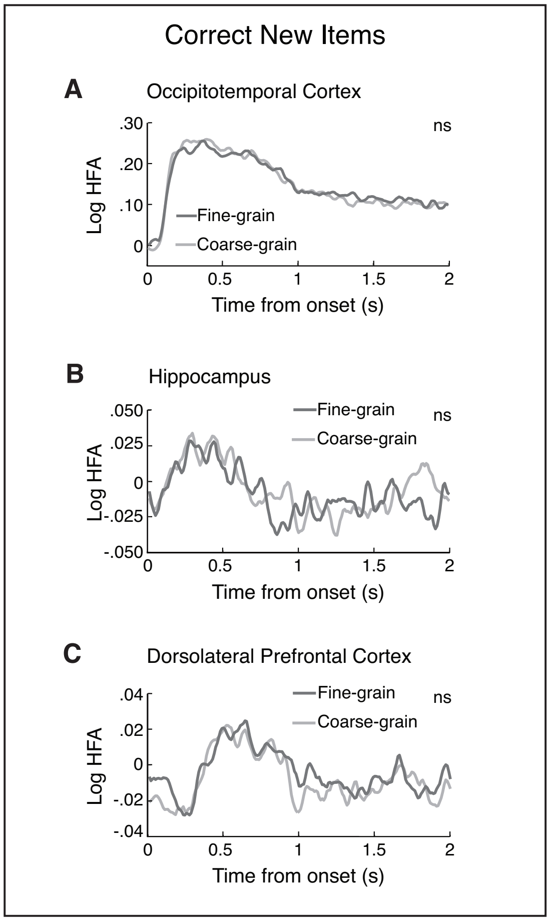

Fig. S2. In the fine-grain task, high-frequency activity (HFA) did not differ for correct new items between tasks. Significance was assessed in the 2s following post-stimulus onset divided into four 500ms time bins. For illustrative purposes only, HFA is plotted as the mean across every 50ms with a 10ms sliding time window. (A) HFA in occipitotemporal cortex (OTC). (B) HFA in hippocampus. (C) HFA in dorsolateral prefrontal cortex (DLPFC). ns=not significant. N=5.

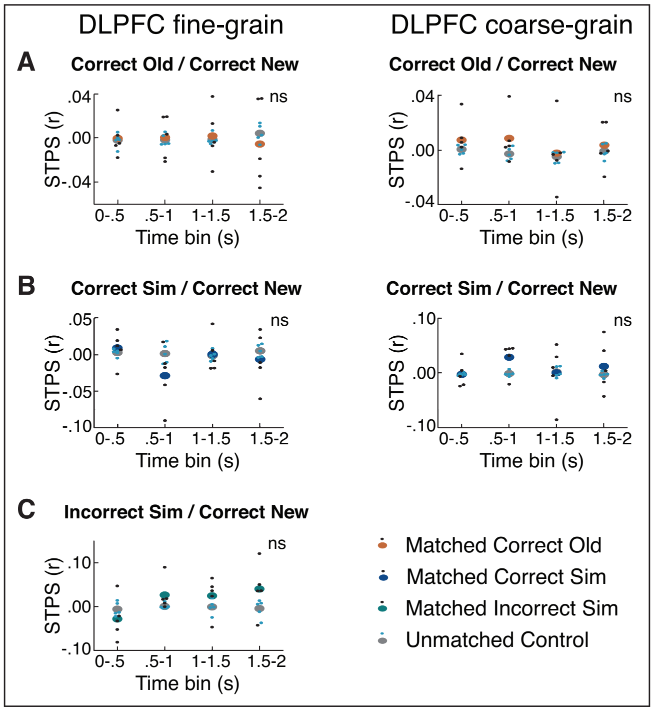

Fig. S3. Spatiotemporal pattern similarity (STPS) in dorsolateral prefrontal cortex (DLPFC) across tasks and condition types. No STPS values were significantly above baseline in any condition or time bin. ns=not significant. N=5.

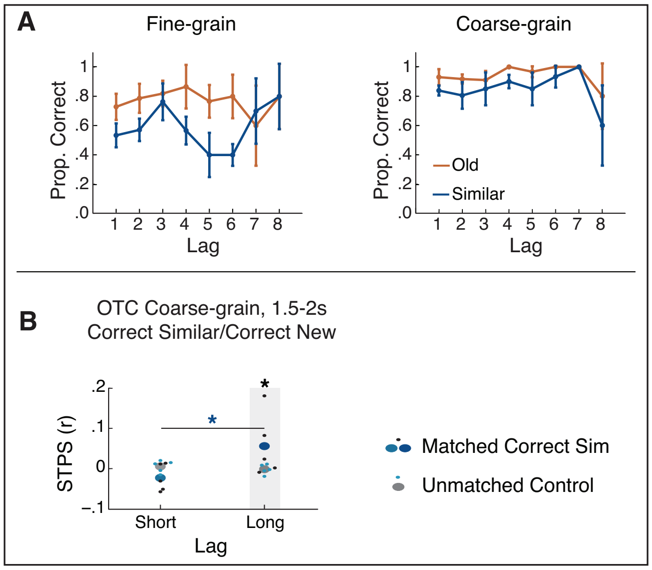

Fig. S4. Contributions of lag to behavioral performance and neural activity. (A) Proportion of correct responses (Prop. Correct) for old and similar items as a function of the lag intervening between first and second presentations. Error bars indicate mean ±SEM across participants. (B) In the coarse-grain task, spatiotemporal pattern similarity (STPS) in occipitotemporal cortex (OTC) was significantly greater for similar items with longer lags (3-8) than shorter lags (1-2) during 1.5-2s. *p<.05. N=5.

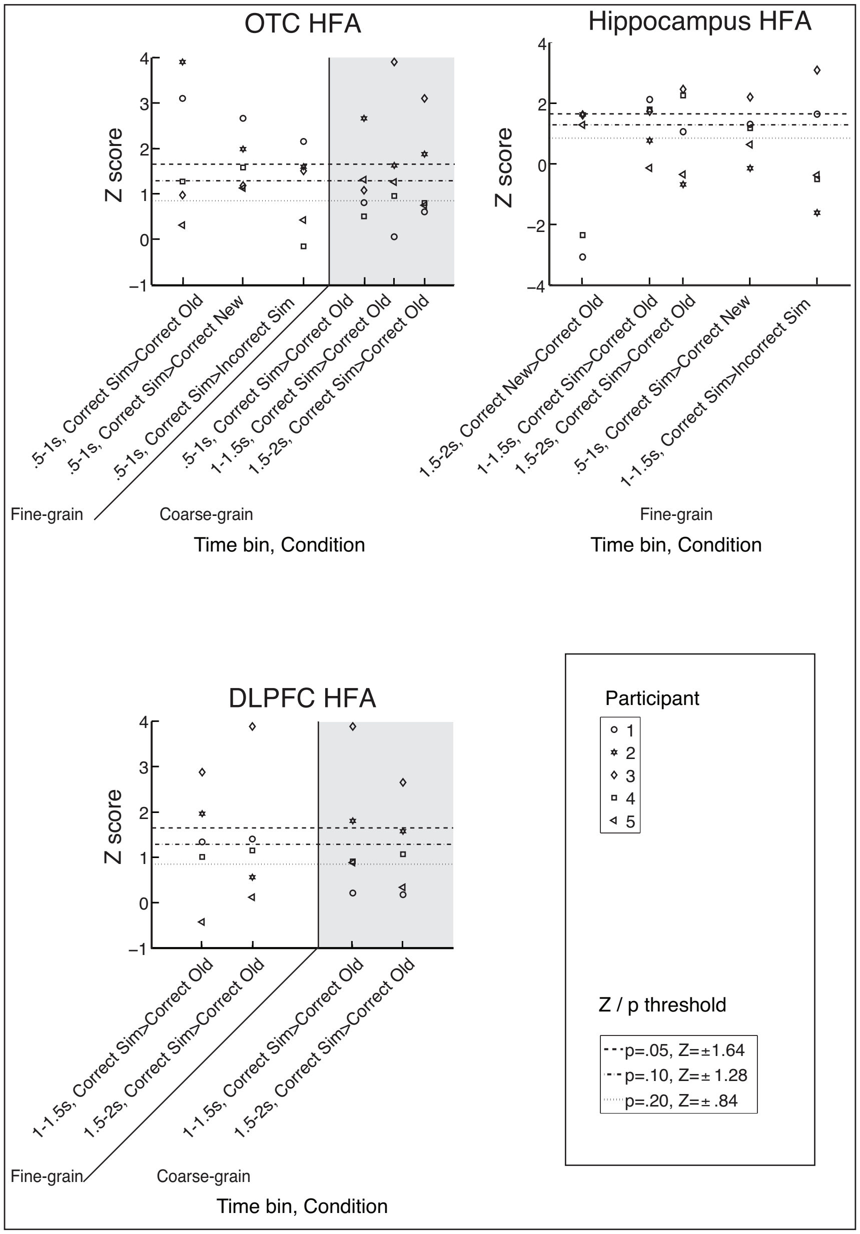

Fig. S5. Participant-level statistical significance for all significant high-frequency activity (HFA) comparisons reported in the Results section. OTC = Occipitotemporal cortex; DLPFC = Dorsolateral prefrontal cortex. N=5.

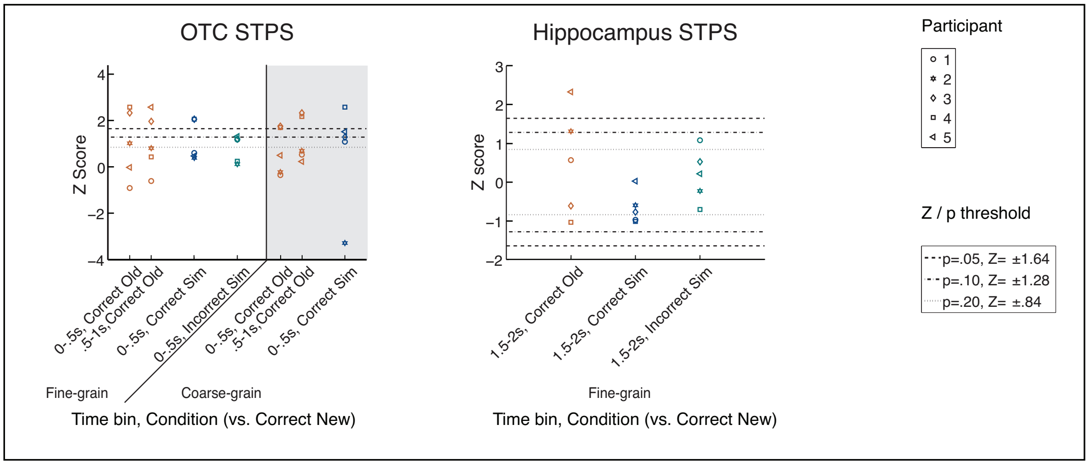

Fig. S6. Participant-level statistical significance for all significant spatiotemporal pattern similarity (STPS) results reported in the Results section. OTC = Occipitotemporal cortex. For completeness with Fig. 3, hippocampal STPS of incorrect similar items in the fine-grain task from 1.5-2s is also included, even thought STPS in this time bin was not significant. N=5.

Table S1. Contributions of univariate high frequency activity to spatiotemporal pattern similarity. Each row includes the task-specific conditions and time bin where univariate high frequency activity was significant. The *p* value indicates the significance of an analysis of variance (ANOVA) between two regression models with STPS as the dependent measure and: (a) a null model with subject as a predictor; (b) a univariate model with three additional predictors: univariate activity from each of the two conditions reported in the table, as well as an interaction term between the two univariate predictors. OTC = Occipitotemporal cortex; DLPFC = Dorsolateral prefrontal cortex.

Region	**Task**	**Conditions**	**Time Bin**	***p* value**
Hippocampus	Fine-grain	Correct Old, Correct New	1.5-2s	0.88
Hippocampus	Fine-grain	Correct Similar, Correct New	.5-1s	0.92
OTC	Fine-grain	Correct Similar, Correct New	.5-1s	0.997
OTC	Fine-grain	Incorrect Similar, Correct New	1-1.5s	0.84
DLPFC	Fine-grain	Correct Old, Correct New	.5-1s	0.999
DLPFC	Fine-grain	Correct Similar, Correct New	.5-1s	0.69
DLPFC	Fine-grain	Correct Similar, Correct New	1-1.5s	0.91
DLPFC	Fine-grain	Correct Similar, Correct New	1.5-2s	0.32
DLPFC	Fine-grain	Incorrect Similar, Correct New	.5-1s	0.999
DLPFC	Fine-grain	Incorrect Similar, Correct New	1-1.5s	0.52
DLPFC	Fine-grain	Incorrect Similar, Correct New	1.5-2s	0.55
DLPFC	Coarse-grain	Correct Old, Correct New	.5 -1 s	0.97
DLPFC	Coarse-grain	Correct Similar, Correct New	.5 -1 s	0.999
DLPFC	Coarse-grain	Correct Similar, Correct New	1-1.5s	0.06
DLPFC	Coarse-grain	Correct Similar, Correct New	1.5-2s	0.59

Table S2. Participant demographic information. Participant information includes: age (years); gender (F = Female, M = Male); Language lateralization as determined by Wada procedure (L = Left); Wechsler Adult Intelligence Scale III indices: VCI = Verbal Comprehension Index; POI = Perceptual Organization Index; WMI =Working Memory Index; PSI = Processing Speed Index. NA = Not available.

Participant	Age	Gender	WADA	VCI	POI	WMI	PSI
1	42	F	L	126	111	99	111
2	24	M	L	122	121	124	111
3	39	F	L	82	97	91	88
4	25	F	L	134	84	86	102
5	19	F	L	NA	NA	NA	NA

Table S3. Number of electrodes for each participant in each considered region of interest. Regions of interest: Hipp = hippocampus, OTC = occipitotemporal cortex, DLPFC = dorsolateral prefrontal cortex; Implant: hemisphere of implantation (L = Left, R = Right, B = Both); ROIs: hemisphere from which the electrodes were recorded for the regions of interest.

Participant	Hipp	OTC	DLPFC	Implant	ROIs
1	4	2	10	R	R
2	1	3	7	L	L
3	5	1	6	L	L
4	3	2	8	R	R
5	6	3	9	B	L

Table S4. Participant-level significant effects. N refers to the number of participants exhibiting a p value less than the value indicated. Region: Hipp = Hippocampus; OTC = Occipitotemporal cortex; DLPFC = Dorsolateral prefrontal cortex. Measure: HFA = high frequency activity. STPS = spatiotemporal pattern similarity. Task: Fine = fine-grain task; Coarse = coarse-grain task.

Region	Measure	Task	Conditions	Time Bin	N, p<.05	N, p<.1	N, p<.2
Hipp	HFA	Fine	Correct New, Correct Old	1.5-2s	0	2	3
Hipp	HFA	Fine	Correct Similar, Correct Old	1-1.5s	3	3	3
Hipp	HFA	Fine	Correct Similar, Correct Old	1.5-2s	2	2	3
Hipp	HFA	Fine	Correct Similar, Correct New	0.5-1s	1	2	3
Hipp	HFA	Fine	Correct Similar, Incorrect Similar	1-1.5s	1	2	2
OTC	HFA	Fine	Correct Similar, Correct Old	0.5-1s	2	2	4
OTC	HFA	Fine	Correct Similar, Correct New	0.5-1s	2	3	5
OTC	HFA	Fine	Correct Similar, Incorrect Similar	0.5-1s	1	3	3
OTC	HFA	Coarse	Correct Similar, Correct Old	0.5-1s	1	2	3
OTC	HFA	Coarse	Correct Similar, Correct Old	1-1.5s	1	2	4
OTC	HFA	Coarse	Correct Similar, Correct Old	1.5-2s	2	2	2
DLPFC	HFA	Fine	Correct Similar, Correct Old	1-1.5s	2	3	4
DLPFC	HFA	Fine	Correct Similar, Correct Old	1.5-2s	1	2	3
DLPFC	HFA	Coarse	Correct Similar, Correct Old	1-1.5s	2	2	4
DLPFC	HFA	Coarse	Correct Similar, Correct Old	1.5-2s	1	2	3
Hipp	STPS	Fine	Correct Old, Correct New	1.5-2s	1	2	2
Hipp	STPS	Fine	Correct Similar, Correct New	1.5-2s	0	0	2
OTC	STPS	Fine	Correct Old, Correct New	0-0.5s	2	2	3
OTC	STPS	Fine	Correct Old, Correct New	0.5-1s	2	2	2
OTC	STPS	Fine	Correct Similar, Correct New	0-0.5s	2	2	2
OTC	STPS	Fine	Incorrect Similar, Correct New	0-0.5s	0	1	3
OTC	STPS	Coarse	Correct Old, Correct New	0-0.5s	2	2	2
OTC	STPS	Coarse	Correct Old, Correct New	0.5-1s	2	2	2
OTC	STPS	Coarse	Correct Similar, Correct New	0-0.5s	1	2	4

References

1. Pruessner JC, et al. (2000) Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. *Cereb Cortex* 10(4):433–442.
2. Duvernoy HM (1999) *The human brain: Surface, blood supply, and three-dimensional sectional anatomy* (Springer-Verlag Wien, New York, NY). 2nd Ed.
3. Yang AI, et al. (2012) Localization of dense intracranial electrode arrays using magnetic resonance imaging. *Neuroimage* 63(1):157–65.
4. Grill-Spector K, Weiner KS (2013) The functional architecture of the ventral temporal cortex and its role in categorization. 18(9):1199–1216.
5. Serruya MD, Sederberg PB, Kahana MJ (2014) Power shifts track serial position and modulate encoding in human episodic memory. *Cereb Cortex* 24(2):403–13.
6. Sederberg PB, et al. (2007) Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cereb Cortex* (1190–1196).
7. Sederberg PB, et al. (2006) Oscillatory correlates of the primacy effect in episodic memory. *Neuroimage* 32(3):1422–31.
8. Sederberg PB, Kahana MJ, Howard MW, Donner EJ, Madsen JR (2003) Theta and gamma oscillations during encoding predict subsequent recall. *J Neurosci* 23(34):10809–14.
9. LaRocque KF, et al. (2013) Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J Neurosci* 33(13):5466–5474.