Supporting Information for

Hippocampal activity predicts contextual misattribution in false memories

Noa Herz, Bernard R. Bukala, James E. Kragel and Michael J. Kahana

Noa Herz.
E-mail: herz.noa@gmail.com

This PDF file includes:

Supporting text
Figs. S1 to S6
Tables S1 to S3
SI References

Supporting Information Text

Results

A. Output Position Models. To control for output position, we predicted hippocampal power (either HFA, LFA or low-theta) as a function of retrieval type while adding output position into the model. In the uncategorized experiment, we found a main effect of retrieval type on HFA ($\chi^2_{(2)} = 318.469, p < .001$), with decreased HFA for intrusions relative to correct recalls ($z = 6.528, p < .001$) but increased relative to deliberations ($z = 8.561, p < .001$). In addition, we found a main effect of retrieval type on LFA ($\chi^2_{(2)} = 58.457$, $p < .001$), with increased LFA for intrusions relative to correct recalls ($z = 2.640$, $p = .010$) but decreased relative to deliberations ($z = -3.789$, $p < .001$). In the low-theta range, we found decreased low-theta power for intrusions relative to correct recalls ($z = -2.495$,$p = .015$) and deliberations ($z = -6.410$, $p < .001$) ($\chi^2_{(2)} = 45.595$, $p < .001$).

In the categorized experiment, we found a main effect of retrieval type on HFA ($\chi^2_{(2)} = 140.805$, $p < .001$), with decreased HFA for intrusions relative to correct recalls ($z = -5.786$, $p < .001$) but increased relative to deliberations ($z = 3.198$, $p = 0.001$). In addition, we found a main effect of retrieval type on LFA ($\chi^2_{(2)} = 31.7128$, $p < .001$), with decreased LFA for intrusions ($z = -3.139$, $p = .002$) and correct recalls ($z = -5.434$, $p < .001$) relative to deliberations. Intrusions were not significantly different from correct recalls ($z = 1.049$, $p = .294$). In the low-theta range, a similar main effect of retrieval type to the one found for the LFA emerged, with decreased low-theta for correct recalls ($z = -2.427$, $p = 0.015$) and intrusions ($z = -3.590$, $p < .001$) relative to deliberations ($\chi^2_{(2)} = 14.740$, $p < .001$). Table S1 shows the beta coefficients of the linear mixed effect model before (left column) and after (right column) adding output position to the model. Similar beta coefficients before and after including output position suggest that the observed effects are not explained by the different output position of correct recalls and intrusions. To obtain an estimation of the total amount of variance explained by output position we followed the formula shown in (1):

$$PRV = (Var_{recall} - Var_{recall+output})/Var_{recall} \quad [1]$$

where PRV is the proportion reduction in variance, Var_{recall} is the residuals variance estimate from the recall type model and $Var_{recall+output}$ is the residuals variance estimate from the full model, containing both recall type and output position. The percent reduction in unexplained variance when adding output position to the model was only .009% in the categorized FR and 0.1% in the uncategorized FR.

B. Electrode Location Across the Anterior - Posterior Axis. To determine whether LFA reduction changes as a function of contact location along the hippocampal anterior-posterior axis, we measured each contact distance from the uncal apex ($y = -20$ mm in Talairach space), a standard landmark segmenting the anterior and posterior portions of the hippocampus (2). Using this continuous measure, we tested whether LFA can be predicted by an electrode location x event type interaction using a linear mixed-effects model. We found a significant interaction effect ($\chi^2_{(3)} = 17.3297, p = 0.0006$), with intrusions that shared only one type of contextual similarity with the target context (nS-PLI and S-ELI) exhibiting increased LFA relative to correct recalls specifically in the posterior portion of the hippocampus (z=-2.660, p=0.008 and z=-3.25, p= 0.001, respectively). Contextually similar intrusions (S-PLI) were not different from correct recalls in either the anterior or posterior section of the hippocampus (z=-0.326, p= 0.744) (Figure S5.B). In the uncategorzied experiment we did not find any interaction between electrode location along the hippocampus and event type ($\chi^2_{(3)} = 1.5410, p = 0.672$) (Figure S5.A).

1. Materials and Methods

A. Word Pools. We used the MRC Psycholinguistic Database to retrieve the characteristics of words included in the uncategorized and categorized free recall experiments (for the full word pool, see 'RAM Free Recall' in: $https://memory.psych.upenn.edu/WordPools$). Table S3 shows the averaged word concreteness (ranging from abstract to concrete, on a 100 to 700 scale), word length (the number of letters in each word), number of syllables and number of phonemes of words in the two experiments.

B. Linear Mixed Effect Models. To investigate whether HFA, LFA or low-theta change as a function of retrieval type, we used the linear mixed effects model:

$$power \sim recall\_type + (1|participant/session) \quad [2]$$

where *power* is the averaged power in the frequency range of interest, *recall_type* is recall type (either correct recall/intrusions/deliberations for the recall veridicality model, or intrusion types for the contextual similarity models) and $1|participant/session$ are random intercepts for session nested in participant.

To test whether the observed differences between recall types resulted from their different output positions, we used the model:

$$power \sim recall\_type + output\_position + (1|participant/session) \quad [3]$$

where *output_position* is output position of each recalled event and the other factors are the same as in Eq. (2).

In the Irregular-Resampling Auto-Spectral Analysis (IRASA) we used the model:

$$IRASA\_param \sim recall\_type * exp + (1|participant/session) \quad [4]$$

where $IRASA\_param$ is either the broadband parameters (intercept and slope) or oscillatory components (low-theta, LFA, HFA) extracted from the IRASA analysis. $recall\_type$ is the same as in Eq. (2) and exp is a binary variable indicating the uncategorized or categorized free-recall experiment.

In the temporal specificity analysis we used the formula:

$$delta \sim freq * time + (1|participant/session) \quad [5]$$

where $freq$ is a categorical variable indicating the frequencies of interest (low-theta, LFA, HFA), $time$ is a binary variable of pre/post retrieval, and $delta$ is the difference between correct recalls and intrusions.

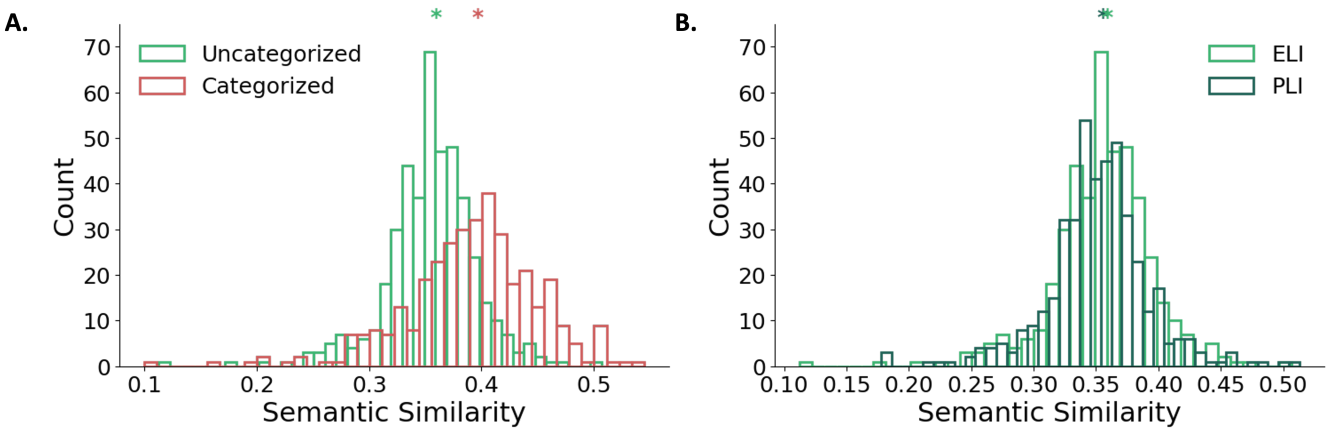

Fig. S1. Distributions of semantic similarity values, computed using word2vec. **A.** Extra-list intrusions (ELIs) in the categorized experiment (pink) shared a higher semantic similarity with the encoded list relative to ELIs in the uncategorized experiment (green) ($z = 10.294, p < .001$). **B.** In the uncategorized experiment, ELIs (light green) did not share greater semantic similarity with the encoded list relative to prior-list intrusions (PLI; dark green) ($z = 1.266, p = .205$). Asterisks denote distributions' means.

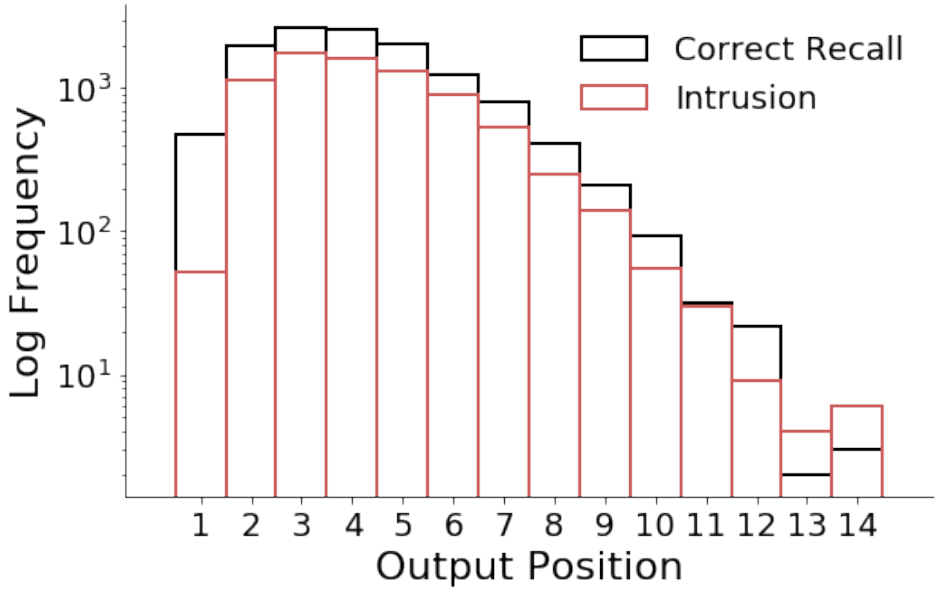

Fig. S2. Distribution of retrievals as a function of output position. Frequency (log transformed) of correct recalls (black) and intrusions (pink) for each output position. Intrusions tended to arrive at later output positions relative to correct recalls during the recall phase.

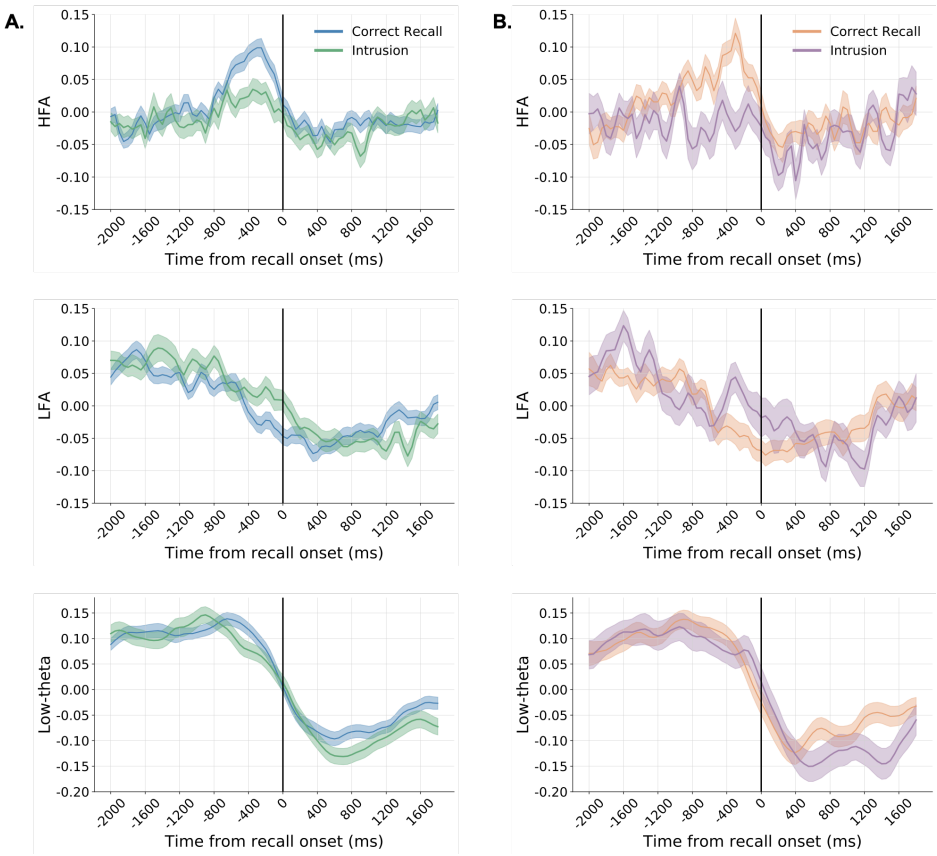

Fig. S3. Temporal specificity in the hippocampus for each experimental task. Mean high-frequency activity (HFA; top), low-frequency activity (LFA; middle) and low-theta (bottom) measured at each time point from two seconds prior to two seconds following vocalization for either the **A.** Uncategorized free recall, or **B.** Categorized free recall experiment. No differences in temporal specificity were found between the two experiments for either HFA, LFA or low-theta (all $p's > .334$). Shaded area represent ±1 standard error of the mean.

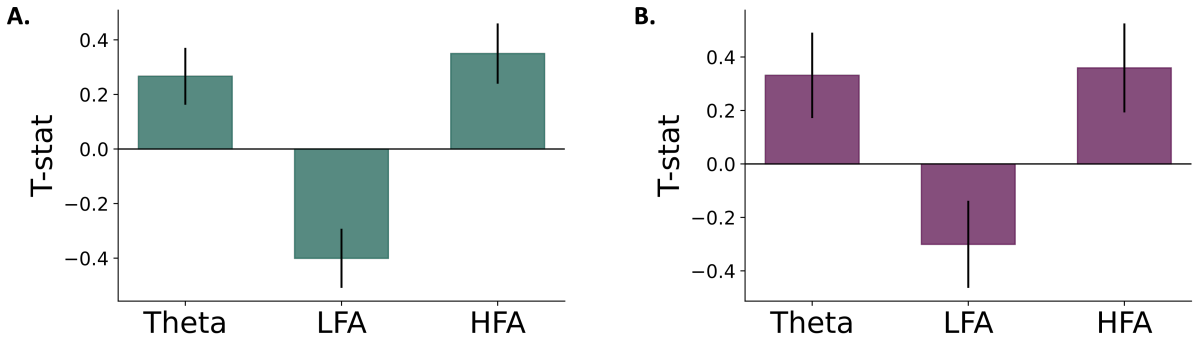

Fig. S4. Oscillatory activity, computed as the difference between the mixed and broadband power spectra, for the low-theta, low-frequency (LFA) and high-frequency activity (HFA). Bars represent the averaged within-subject t-statistic of the difference between correct recalls and intrusions. Correct recalls exhibit increased low-theta, decreased low-frequency and increased high-frequency oscillations relative to intrusions in both the **A.** Uncategorized free recall, and **B.** Categorized free recall experiment. Error bars represent ±1 standard error of the mean. No significant interaction between experiment and event type emerged for neither theta ($p's > .077$), LFA ($p's > .302$) or HFA ($p's > .425$).

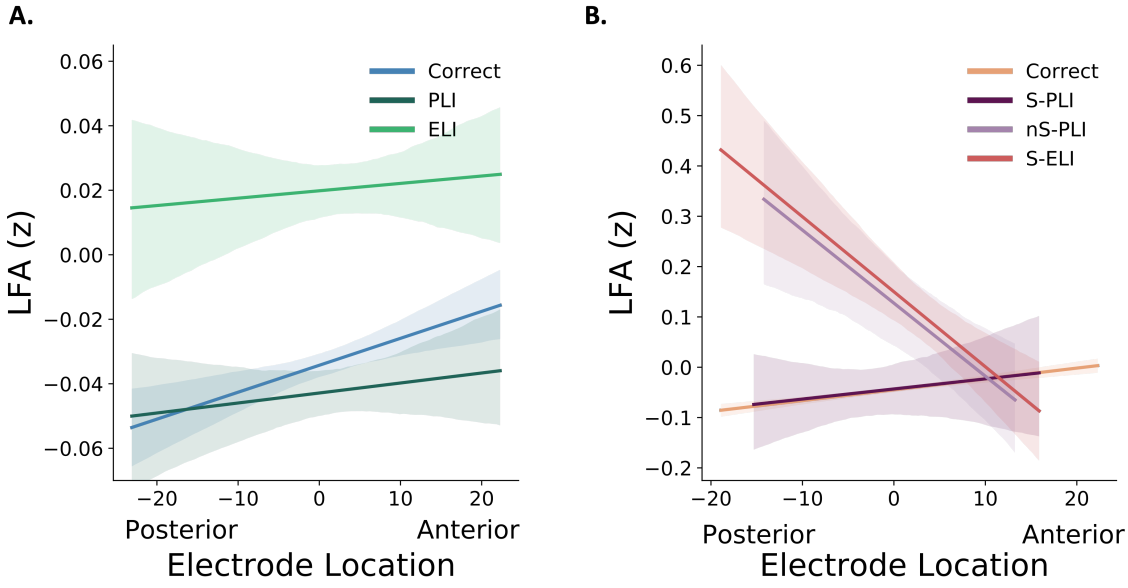

Fig. S5. Low-frequency activity of each retrieval type as a function of electrode location along the hippocampal anterior-posterior axis. **A.** No interaction between recall type and electrode location emerged in the uncategorized free recall experiment. **B.** Interaction between electrode location and low-frequency activity in the categorized free recall experiment. Contextually dissimilar intrusions (nS-PLI and S-ELI) exhibited increased LFA relative to correct recalls and contextualy similar intrusions (S-PLI) specifically in the posterior portion of the hippocampus. Error bars represent 68% confidence interval of the linear fit of the data (Correct: correct recall, PLI; prior-list intrusion, ELI; extra-list intrusion, S-PLI: semantic prior-list intrusion, nS-PLI: non-semantic prior-list intrusion, S-ELI: semantic extra-list intrusion).

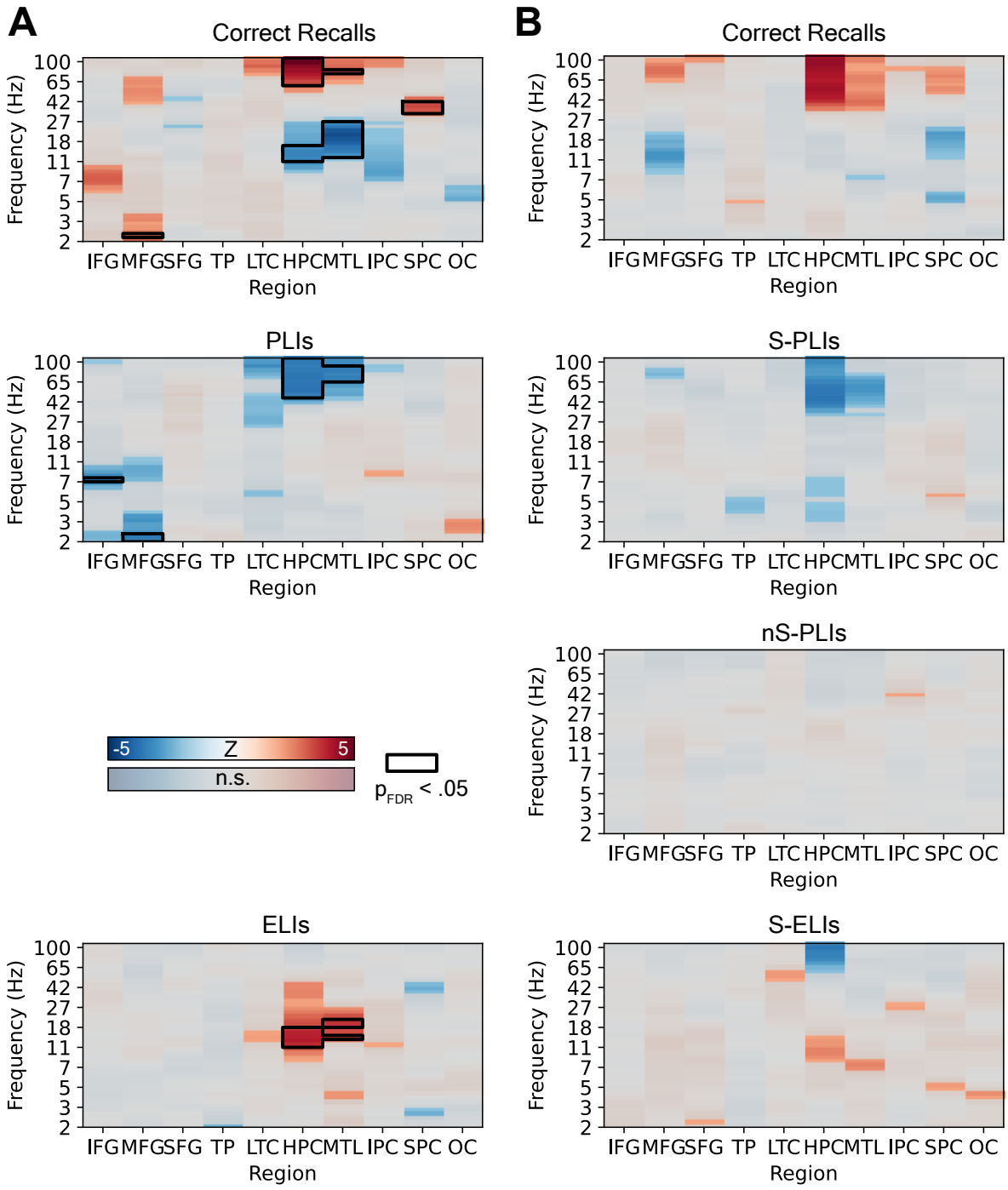

Fig. S6. Forward-model estimates of feature importance broken down for uncategorized (**A**) and categorized (**B**) lists. Significant ($p < .05$) increases and decreases in power important for classification are shown in red and blue, respectively. Features that survived FDR correction are outlined in black. Non-significant (n.s.) features are masked in grey. IFG, inferior frontal gyrus; MFG, middle frontal gyrus; SFG, superior frontal gyrus; TP; temporal pole; MTL, medial temporal lobe; HPC, hippocampus; LTC, lateral temporal cortex; IPC, inferior parietal cortex; SPC, superior parietal cortex; OC, occipital cortex.

Table S1. Beta coefficients in the linear mixed effects model of recall veridicality.

			Coefficient (ste) uncorrected	Coefficient (ste) with output position
Uncategorized FR	Low-theta	Correct vs. Delib	-0.041 (0.010)	-0.040 (0.010)
		Correct vs. Intrusions	0.029 (0.012)	0.030 (0.012)
		Intrusions vs. Delib	-0.070 (0.011)	-0.071 (0.011)
	LFA	Correct vs. Delib	-0.060 (0.008)	-0.060 (0.008)
		Correct vs. Intrusions	-0.027 (0.010)	-0.026 (0.010)
		Intrusions vs. Delib	-0.033 (0.009)	-0.034 (0.009)
	HFA	Correct vs. Delib	0.112 (0.006)	0.112 (0.006)
		Correct vs. Intrusions	0.051 (0.008)	0.051 (0.008)
		Intrusions vs. Delib	0.061 (0.007)	0.061 (0.007)
Categorized FR	Low-theta	Correct vs. Delib	-0.034 (0.014)	-0.034 (0.014)
		Correct vs. Intrusions	0.031 (0.019)	0.030 (0.019)
		Intrusions vs. Delib	-0.065 (0.018)	-0.065 (0.018)
	LFA	Correct vs. Delib	-0.060 (0.011)	-0.060 (0.011)
		Correct vs. Intrusions	-0.016 (0.015)	-0.016 (0.015)
		Intrusions vs. Delib	-0.045 (0.014)	-0.045 (0.014)
	HFA	Correct vs. Delib	0.108 (0.009)	0.108 (0.009)
		Correct vs. Intrusions	0.071 (0.012)	0.071 (0.012)
		Intrusions vs. Delib	0.037 (0.012)	0.037 (0.012)

Table S2. Overall number of intrusions for each intrusion type in the categorized free recall paradigm (ELIs; extra-list intrusions, PLIs; prior-list intrusions). ELIs are most-frequently semantically related to the encoded list, while ELIs that are not semantically related are very rare. This unbalanced distribution was not evident for PLIs ($p < .001$).

	ELIs	PLIs
Semantic intrusions	216	108
Non-semantic intrusions	24	136

Table S3. Word characteristics in the uncategorized and categorized free-recall experiments.

	Uncategorized FR Mean (std)	Categorized FR Mean (std)
Concreteness	588.7 (31.5)	596.2 (37.4)
Word length	4.1 (0.8)	5.6 (1.8)
Num syllables	1.0 (0.05)	1.7 (0.7)
Num phonemes	3.2 (0.7)	4.4 (1.6)

References

1. JL Peugh, A practical guide to multilevel modeling. *J. Sch. Psychol.* **48**, 85–112 (2010).
2. J Poppenk, HR Evensmoen, M Moscovitch, L Nadel, Long-axis specialization of the human hippocampus. *Trends Cogn. Sci.* **17**, 230–240 (2013).