PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES
NEUROSCIENCE

Hippocampal activity predicts contextual misattribution of false memories

Noa Herz[a,1], Bernard R. Bukala[a], James E. Kragel[b], and Michael J. Kahana[a]

Edited by Daniel Schacter, Harvard University, Cambridge, MA; received March 31, 2023; accepted August 2, 2023

Failure of contextual retrieval can lead to false recall, wherein people retrieve an item or experience that occurred in a different context or did not occur at all. Whereas the hippocampus is thought to play a crucial role in memory retrieval, we lack understanding of how the hippocampus supports retrieval of items related to a target context while disregarding related but irrelevant information. Using direct electrical recordings from the human hippocampus, we investigate the neural process underlying contextual misattribution of false memories. In two large datasets, we characterize key physiological differences between correct and false recalls that emerge immediately prior to vocalization. By differentiating between false recalls that share high or low contextual similarity with the target context, we show that low-frequency activity (6 to 18 Hz) in the hippocampus tracks similarity between the current and retrieved context. Applying multivariate decoding methods, we were able to reliably predict the contextual source of the to-be-recalled item. Our findings elucidate one of the hallmark features of episodic memory: our ability to distinguish between memories that were formed on different occasions.

false memory | context | hippocampus | free-recall

Significance

The hippocampus is known to be involved in the retrieval of items bound to a specific context. Failure of contextual retrieval can lead to false recalls, wherein people recall an item outside of its associated context. Using direct intracranial recordings in humans, we show that hippocampal activity at the moments leading up to vocalization can predict the veridicality of the to-be-recalled information. We further show that hippocampal low-frequency activity (LFA; 6 to 18 Hz) reflects the associative strength between items and their associated context, with greater LFA reduction signaling greater overlap between the retrieved items' context and the target context. The results shed light on the neural process underlying our ability to distinguish between memories that were formed on different occasions.

Humans possess the remarkable ability to mentally time travel and relive past events. Our ability to retrieve the temporal and situational context of our past experiences is a hallmark of episodic memory and relies on hippocampal activity (1–3). However, our memories are not infallible and can be susceptible to errors. For instance, when trying to recall the items we purchased yesterday at the supermarket we may, mistakenly, recall an item bought at a different store. Retrieval of items learned in an irrelevant context can occur due to source misattribution (where "source" refers to the episode, or context, in which the information was originally presented) (4). Such misattribution, can result from semantic relatedness or sensory resemblance of mnemonic details (5). These cases of "miscontextualized" memories allow us to examine the role of the hippocampus in retrieving the association of items with their encoded context. First, if the hippocampus stores the associations of items with their originally encoded contexts, hippocampal activity should exhibit distinct activity patterns for correct and false recalls. Second, if the hippocampus represents the associative strength between items and their encoded context, then items encoded in similar contexts should also elicit similar hippocampal activity. The hippocampus should thus gradually distinguish between correct and false recalls as a function of the similarity between the target context and the context associated with the erroneous memory.

Previous studies investigating false recall have focused on differentiating true and false memories without considering their degree of similarity. To induce false recalls in a controlled setting, researchers commonly employ the Deese–Roediger–McDermott (DRM) procedure (6), where participants are presented with a list words strongly associated with a critical nonpresented word (e.g., bed, awake, and night being associates of the word "sleep"). False recall of the critical item may arise because the related list items cause participants to think of the critical item during study, or because of the similarity of the cues at test, or a combination of both factors (7). Studies investigating the neural correlates of semantically associated false memories have implicated the prefrontal (8, 9) and anterior temporal (10, 11) cortex in false recollection, areas with extensive anatomical and functional connections with memory-related medial temporal lobe (MTL) regions. Recognition-based neuroimaging studies show differences between true and false memories in early sensory regions or in the posterior MTL, without any observed differences in the hippocampus (12–14, but see refs. 15 and 16). However, recordings from hippocampal depth electrodes may provide a more direct readout of hippocampal physiology. Indeed, analyzing hippocampal Intracranial electroencephalography (iEEG), Long et al. found elevated high-frequency activity (HFA)

Author affiliations: [a]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104; and [b]Department of Neurology, University of Chicago, Chicago, IL 60637

Author contributions: M.J.K. designed research; N.H. contributed new reagents/analytic tools; N.H., B.R.B., and J.E.K. analyzed data; and N.H., B.R.B., J.E.K., and M.J.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: herz.noa@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2305292120/-/DCSupplemental.

Published September 26, 2023.

(44 to 100 Hz) for correct relative to false recalls just prior to item vocalization (17). Increased HFA also emerged during encoding of subsequently remembered items (18, 19), possibly reflecting successful item–context binding that enables later retrieval of the encoded information (20).

Although comparing correct and false recalls allows examination of hippocampal engagement in contextually mediated retrieval, it does not provide insights into the neural representation of item–context associations. Furthermore, the majority of electrophysiological studies demonstrating hippocampal involvement in episodic recall contrasted neural activity during correct recall with matched periods of silent memory search (18, 21). These events do not uniquely isolate the correlates of contextual retrieval, as they also diverge in terms of the retrieval process, as well as by the motor activity associated with item vocalization. In these studies, decreased low-frequency activity (LFA) typically accompanied increased HFA (18, 22, 23), possibly reflecting a "tilt" in the broadband power spectrum. It has been proposed that this spectral tilt serves as a proxy for neuronal firing rate (24, 25). Alternatively, increased HFA and decreased LFA may reflect distinct processes, each serving a different memory function (26). HFA increase has been found across widespread brain regions and cognitive tasks (27) and was therefore proposed to subserve a nonspecific marker of brain activation (28, 29). In contrast, LFA desynchronization was proposed to enable rich memory representation (30, 31) via reduction in neural noise (32). In recognition memory studies, LFA desynchronize to a greater extent during retrieval of associative information (33) and retrieval of highly detailed memories (34), and was therefore suggested to reflect memory strength or mnemonic specificity. Whereas high-theta and alpha/beta oscillations decrease prior to successful recall, low-frequency theta oscillations (~2 to 5 Hz) sometimes increase (35, 36), an effect that can be masked when aggregating across the full 2 to 8 Hz theta band (37).

Here, we test the prediction that the hippocampus enables episodic retrieval by representing the associative strength between retrieved items and the current context. Specifically, we predicted that false memories sharing greater contextual similarity with the target context will show less discriminable hippocampal activity from correct recalls. This prediction accords with computational models defining a slowly drifting representation of context to which items become associated, and which later cues retrieval (3, 38). To test our prediction, we analyzed direct electrical recordings from the human hippocampus of neurosurgical epileptic patients as they studied and subsequently recalled lists of uncategorized or semantically categorized lists of items. Following findings from previous studies, we concentrated our analyses on three frequency bands of interest: HFA (44 to 100 Hz), high-theta/alpha/beta power (henceforth LFA; 6 to 18 Hz), and low-theta (2 to 5 Hz). We deployed univariate as well as multivariate classification methods to characterize the hippocampal activity distinguishing correct from false recalls varying in their contextual similarity to the target context. We show that hippocampal activity can reliably differentiate correct from false retrievals, and that this activity emerges specifically in the moments (< 1 s) preceding memory retrieval and fades rapidly afterward. We further show that hippocampal LFA tracks the degree of similarity between the falsely recalled item's context and the target context, with greater LFA reduction signaling greater overlap between the target context and the context of the retrieved item.

Results

We report six major sets of analyses across two large studies of human hippocampal electrophysiology. Analyzing data from free recall of unrelated word lists, we first ask whether human hippocampal activity at the moments preceding item vocalization could reliably differentiate correct from false recalls. We then test how these observed biomarkers of recall veridicality change as a function of the contextual similarity between the false recall and the target context. Using data from free recall of semantically organized word lists, we then test whether similar hippocampal biomarkers reflective of contextual similarity in unrelated word lists also reflect the semantic similarity between the false recall and the target context in semantically organized lists. We then investigate whether the observed spectral pattern differentiating correct from false recall reflects oscillatory activity or a broadband "tilt" of the power spectrum. Next, we ask whether the observed hippocampal biomarkers of recall veridicality play a part in the retrieval process by testing for their emergence specifically at the time prior to item vocalization. Finally, using multivariate classification methods, we predict the contextual similarity of the to-be-recalled information and test whether such prediction is possible not only at the group level, but also at a single-subject level.

Hippocampal Activity Distinguishes Veridical Recall from False Memories. We first investigated whether hippocampal activity at the moments preceding item vocalization reliably differentiates correct from false recalls. We also examined silent periods of memory search (henceforth, "deliberations") to enable comparison of the present findings to previous studies, e.g., refs. 17, 37, and 39. (The comparison between correct and false recalls, however, better controls for response vocalization and other cognitive processes that might be taking place during periods of silence). In line with prior studies, we found that correct recalls exhibited the twin findings of increased HFA and decreased LFA relative to deliberation periods. False recalls exhibited a reduction of these effects relative to correct recalls (Fig. 1*D*). To statistically evaluate these effects, we predicted hippocampal power (HFA, LFA, or low-theta, separately) as a function of retrieval type using a linear mixed-effects model. We found a main effect of retrieval type on HFA ($\chi^2_{(2)} = 317.32$, $P < 0.001$), with HFA for intrusions decreased relative to correct recalls ($z = -6.46$, $P < 0.001$) but increased relative to deliberations ($z = 8.59$, $P < 0.001$). In addition, we found a main effect of retrieval type on LFA ($\chi^2_{(2)} = 58.74$, $P < 0.001$), with increased LFA for intrusions relative to correct recalls ($z = 2.70$, $P = 0.01$) but decreased LFA relative to deliberations ($z = -3.74$, $P < 0.001$). A main effect in the low-theta range was also found ($\chi^2_{(2)} = 45.21$, $P < 0.001$), with decreased low-theta power for intrusions relative to correct recalls ($z = -2.42$, $P = 0.01$) and deliberations ($z = -6.36$, $P < 0.001$) . These results extend prior studies pointing to decreased LFA and increased HFA, as well as indications for a low-theta increase, as a marker of successful retrieval (35, 37) to false recalls, which exhibit an attenuation of this neural pattern.

The observed neural pattern of false recalls may reflect a lower confidence of patients when falsely recalling items, rather than reflecting miscontextualized recall. Indeed, false recalls tend to happen later during the retrieval phase relative to correct recalls (*SI Appendix*, Fig. S1), and prior studies have shown that recall confidence decreases with output position (i.e., the position of the recalled item in the sequence of recalls) (40). To investigate whether the differences between correct and false recalls resulted from their different output positions, we repeated the analyses while taking output position of each recalled item into account (*Materials and Methods*). After including output position in the model, all of the above effects remained significant (all $P's < 0.05$), suggesting that confidence in the recalled event does not

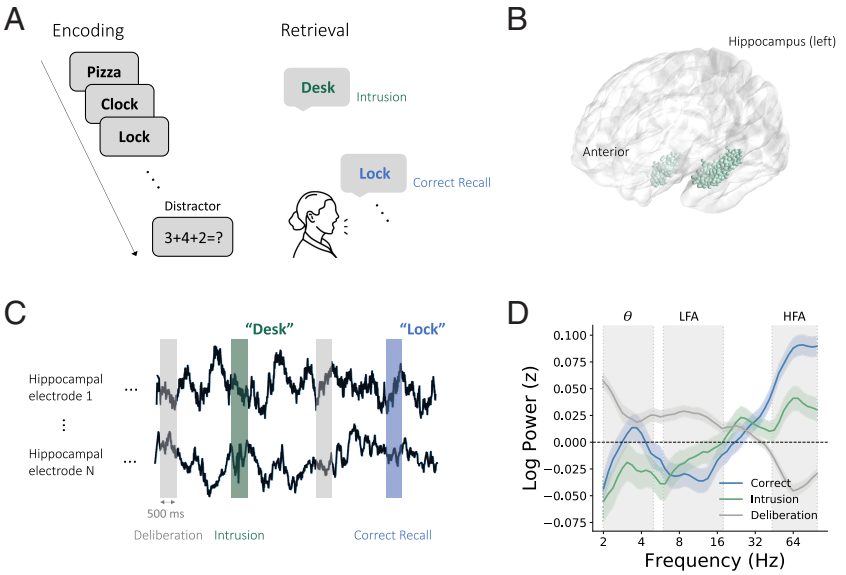

Fig. 1. Hippocampal biomarkers of false memory in uncategorized word lists. (*A*) During encoding, participants studied lists of semantically unrelated words. During retrieval, we asked participants to recall as many words as they could remember from the recent list, in any order. We classified responses as either correct recalls of one of the recently presented words, or false recalls of a word not recently presented (intrusion). (*B*) Regional distribution of hippocampal electrodes across patients performing uncategorized free recall. (*C*) We computed spectral power across hippocampal electrodes during the −2,500 to −100 ms preceding vocalization. Then, we extracted the mean power across the 500 ms preceding vocalization (either correct recalls or intrusions). Deliberation periods were matched 500 ms of silent memory search. (*D*) Correct recalls (blue) exhibit increased HFA, decreased LFA, and decreased low-theta relative to silent periods of memory search ("deliberations") in the 500 ms preceding recall in the hippocampus. Intrusions show an attenuation of the HFA and LFA effects (green). Gray regions mark the frequency ranges analyzed: low-theta, low-frequency activity (LFA), and high-frequency activity (HFA).

account for these hippocampal biomarkers (see *SI Appendix* for the full results).

Overall, these results demonstrate that increased HFA and decreased LFA, a biomarker of successful memory encoding and retrieval (19, 35, 41, 42), distinguish correct from false recalls. The results also mirror recent findings of a low-theta increase for correct recalls (35, 37), a signal that is attenuated for false recalls.

Spectral Correlates of False Recalls Reflect Their Contextual Similarity. We hypothesize that the degree of separation between correct and false recalls depends on the similarity between the context in which these items were encoded. The free-recall paradigm enables differentiation of intrusions based on the similarity of their associated context to the recently encoded list. Prior-list intrusions (PLIs) reflect cases where participants incorrectly recall an item that was not presented on the target list but was nonetheless presented in one of the prior lists of the experiment. Extralist intrusions (ELIs), on the other hand, are incorrect recalls of items never presented in the experiment. Since patients encoded PLIs in a prior phase of the experiment, these intrusions share greater source similarity with the current list's context relative to ELIs. This greater similarity results both from temporal factors (i.e., PLIs being encoded more recently in time) as well as other potential factors (e.g., words are encoded as part of an experiment, while sitting in the same room, etc). (In contrast to these factors, ELIs and PLIs did not differ in their semantic similarity to the recent list in this experiment; see *SI Appendix*, Fig. S1*B*). If the hippocampus stores information about the context in which items were encoded, we should expect PLIs to share a greater neural similarity with correct recalls relative to ELIs. This prediction is in line with findings showing that events encoded in greater temporal proximity have higher chances of becoming linked into an integrated representation (43), as well as theories suggesting the role of the hippocampus in the association of items with an intrinsic and gradually drifting representation of time (44–46). This prediction stands in contrast, however, to pattern separation accounts, suggesting that episodic retrieval requires separation of memories with overlapping features to help disambiguate between them (47, 48).

To test these competing hypotheses, we used a linear mixed-effects model predicting hippocampal power (HFA, LFA, or low-theta separately) as a function of retrieval type, while differentiating between intrusion types (PLIs/ELIs/correct recalls). When predicting HFA, we found a general HFA reduction for both intrusion types relative to correct recalls (PLIs vs. correct recalls: $z = -6.33$, $P < 0.001$; ELIs vs. correct recalls: $z = -4.13$, $P < 0.001$), without any difference between PLIs and ELIs ($z = -1.65$, $P = n.s.$).

For LFA, however, both correct recalls and contextually similar intrusions (i.e., PLIs) exhibited reduced LFA relative to contextually dissimilar intrusions (ELIs) (ELIs vs. correct recalls: $z = 3.48$, $P < 0.001$; ELIs vs. PLIs: $z = 2.18$, $P = 0.02$). LFA did not distinguish between correct recalls and contextually similar intrusions (PLIs vs. correct recalls: $z = 0.84$, $P = n.s.$) (Fig. 2*B*). In the low-theta range, contextually similar and dissimilar intrusions exhibited similar low-theta reduction (PLIs vs. ELIs: $z = 1.20$, $P = n.s.$).

These results demonstrate that hippocampal LFA is reduced as a function of intrusions' source similarity to the studied list, with PLIs showing greater LFA reduction relative to ELIs (Fig. 2). This pattern emerged both in the posterior and anterior sections of the hippocampus (*SI Appendix*, Fig. S5*A*). These findings suggest that hippocampal LFA may reflect the association of items with the context in which they were encountered. Greater LFA reduction marks a stronger match between the retrieved and the target context. The attenuated LFA reduction reflective of ELIs may therefore underlie our ability to distinguish between items that were encoded under different contexts.

Spectral Correlates of False Recalls Reflect Their Semantic Similarity. To the extent that hippocampal decreases in LFA mark the successful reinstatement of the retrieved item's context, one might expect to find a similar LFA decrease for the reinstatement of semantic context. This prediction aligns with a recent study indicating that the hippocampus codes for semantic distances between words during retrieval (49). We used an independent dataset in which participants studied a list of items categorized into three semantic categories to test this prediction (see Fig. 3*A* and *Intracranial Recordings* for more details on the experimental design). By adding semantic structure to the list, we were able to manipulate intrusions' semantic similarity to the target context and test our prediction that greater contextual overlap leads to a greater LFA reduction (*SI Appendix*, Fig. S1*A*).

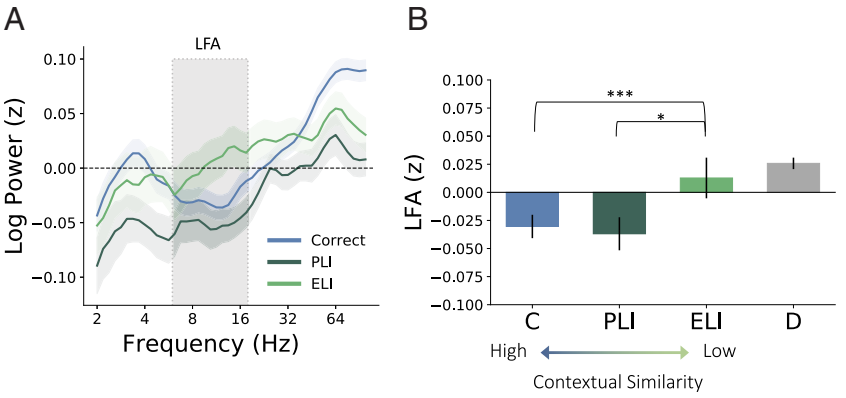

Fig. 2. Hippocampal low-frequency activity (LFA) decreases as a function of intrusions' contextual similarity to the recently encoded list. (*A*) Spectral power of correct recalls, contextually similar (PLIs) and dissimilar (ELIs) intrusions. PLIs exhibit similar LFA reduction to the one characterizing correct recalls. ELIs, on the other hand, does not show similar LFA reduction. The gray region marks the LFA range used in the analysis. (*B*) Mean hippocampal LFA for each retrieval type. LFA decreases as a function of intrusions' contextual similarity to the recently encoded list (C, correct recalls; PLI, prior-list intrusion; ELI, extralist intrusion; D, deliberations). Error bars represent ±1 SE of the mean. *$P < 0.05$, ***$P < 0.001$ in a linear mixed-effects model, FDR corrected.

We first sought to test whether the same hippocampal biomarkers differentiating correct from false recall emerge in this dataset of semantically organized lists. Similar to the results obtained in the uncategorized experiment (Fig. 1*D*), we again found that correct recalls exhibited increased HFA and decreased LFA relative to deliberation periods. Relative to correct recalls, false recalls exhibited reduction in these effects (Fig. 3*C*). When predicting hippocampal power as a function of retrieval type in the categorized free-recall experiment, we found a main effect of retrieval type on HFA ($\chi^2_{(2)} = 140.72$, $P < 0.001$), with HFA for intrusions decreased relative to correct recalls ($z = -5.79$, $P < 0.001$) but increased relative to deliberations ($z = 3.18$, $P = 0.001$). In addition, we found a main effect of retrieval type on LFA ($\chi^2_{(2)} = 31.70$, $P < 0.001$), with decreased LFA for intrusions ($z = -3.14$, $P = 0.001$) and correct recalls ($z = -5.43$, $P < 0.001$) relative to deliberations. LFA for intrusions was not significantly different from correct recalls ($z = 1.04$, $P = n.s.$), possibly due to the higher proportion of semantically associated false recalls in this dataset (*SI Appendix*, Fig. S1*A*). 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 ($z = -2.42$, $P = 0.01$) and intrusions ($z = -3.59$, $P < 0.001$) relative to deliberations ($\chi^2_{(2)} = 14.75$, $P < 0.001$). All of these three main effects remained significant after including output position in the model (all $P's < 0.05$; see *SI Appendix* for the full results).

These results demonstrate that similar hippocampal biomarkers of recall veridicality emerge in an independent dataset of semantically organized information. Correct recalls exhibit increased HFA and decreased LFA relative to deliberations. False recalls exhibit attenuation of these effects, although the LFA difference between correct recalls and intrusions did not reach significance.

In the uncategorized experiment, we found that hippocampal LFA decreases as a function of the source similarity between the intrusion and the target list. Here we asked - does hippocampal LFA also decrease as a function of the semantic similarity between the intrusion and the target context? We tested this question by contrasting two subclasses of intrusions observed in the categorized free recall experiment. Owing to the categorical nature of the study lists in this experiment (e.g., flowers, insects, and fruits categories), participants would often incorrectly recall nonstudied items belonging to one of the studied categories (e.g., a flower that was not presented on the target list). This allowed us to compare intrusions that are semantically related to the encoded information to those that lack such semantic relatedness. We hypothesized that the degree to which intrusions reflect retrieval of semantic context will determine the magnitude of the observed hippocampal LFA decrease. We therefore categorized each intrusion committed by participants as either S-I (semantic intrusion; an intrusion that was semantically related to at least one of the three semantic categories presented during encoding) or nS-I (nonsemantic intrusion; an intrusion that was not related to any of the three encoded categories) (see *Intrusions Semantic Categorization Procedure* under the *Materials and Methods* for more details on the categorization procedure).

We first assessed the frequency of intrusions sharing source (PLIs/ELIs) or semantic similarity (S-I/nS-I) with the recently encoded list. We found that ELIs were almost always semantically related to the recently encoded list, whereas PLIs had a more even distribution between semantically related and nonrelated intrusions (chi-square test of independence: $\chi^2_{(1)} = 112.31$, $P < 0.001$) (see *SI Appendix*, Table S1 for exact intrusions counts). This suggests that intrusions tend to share at least one type of contextual similarity with the encoded information (either source or semantic). Due to the rarity of intrusions that are neither semantically nor source related to the encoded information (i.e., nonsemantic ELIs), we investigated whether HFA, LFA, or low-theta change as a function of three intrusion types; contextually similar intrusions (semantic PLIs; S-PLI) and contextually dissimilar intrusions (including nonsemantic PLIs; nS-PLI and semantic ELIs; S-ELI).

As hypothesized, we found that hippocampal LFA changed as a function of the contextual similarity between the intrusion and the recently encoded list ($\chi^2_{(2)} = 14.38$, $P < 0.001$). Specifically, contextually similar intrusions (S-PLIs) showed the strongest LFA reduction, which was greater in comparison to nS-PLIs ($z = -2.51$, $P = 0.03$) and in comparison to S-ELIs ($z = -3.71$, $P < 0.001$). These effects remained significant after including output position in the model ($\chi^2_{(2)} = 15.05$, $P < 0.001$, S-PLI vs. nS-PLI: $z = -2.68$, $P = 0.02$, S-PLI vs. S-ELI:$z = -3.76$, $P < 0.001$). Fig. 3 shows the mean LFA for each retrieval type as a function of either the source (Fig. 3 *D*, *Top-Left*), semantic (Fig. 3 *D*, *Bottom-Left*), or combined semantic and source similarity of the intrusion to the recently encoded list (Fig. 3 *D*, *Right*). In contrast to LFA, neither HFA ($\chi^2_{(2)} = 1.07$, $P = n.s.$) nor low-theta ($\chi^2_{(2)} = 2.01$, $P = n.s.$) changed as a function of intrusions' source and semantic similarity to the recently encoded list. These findings show that hippocampal LFA covaries with both the semantic and the source similarity

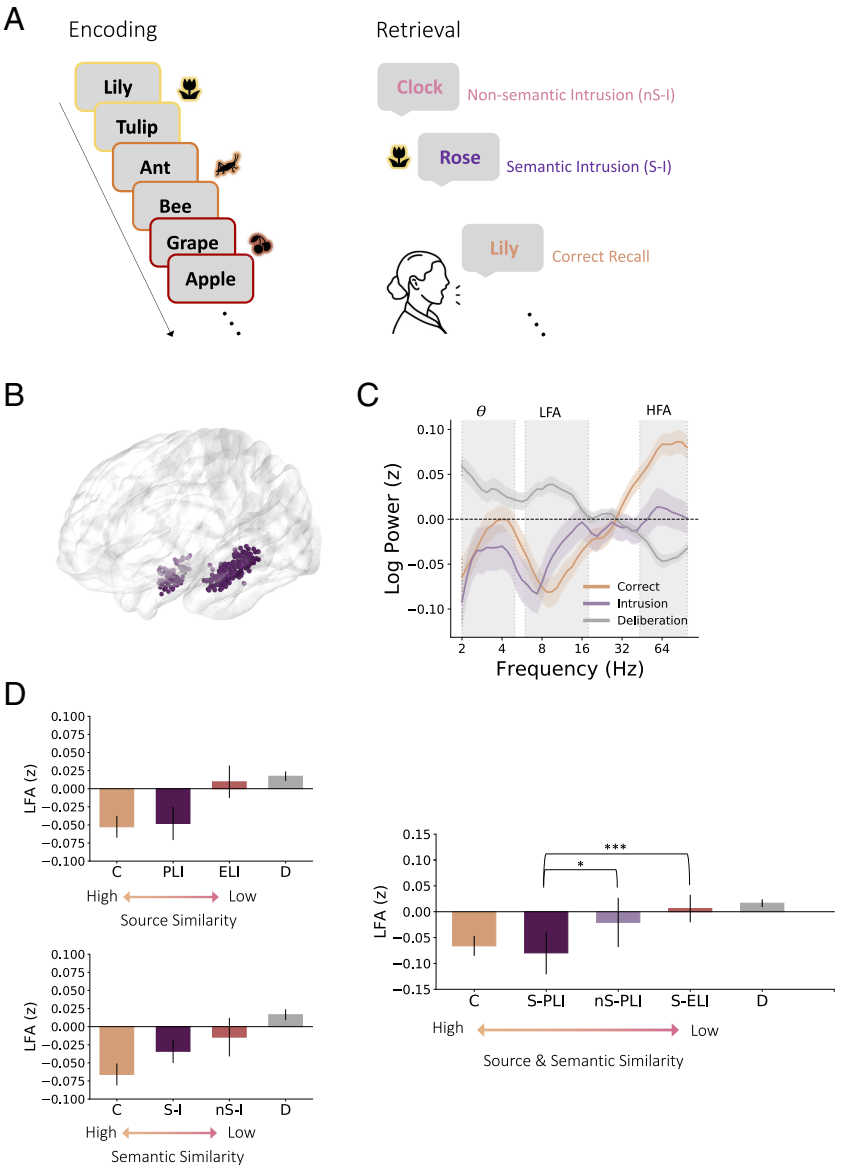

Fig. 3. Hippocampal biomarkers of false recall in categorized word lists. (*A*) Participants studied lists of 12 words drawn from three semantic categories (e.g., flowers, insects, and fruits) grouped in same-category pairs. During retrieval, we classify responses as either correct recalls (e.g., "Lily"), semantically related intrusions (e.g. "Rose"), or nonsemantically related intrusions (e.g., "Clock"). (*B*) Regional distribution of hippocampal electrodes across patients performing categorized recall. (*C*) Correct recalls (orange) exhibit increased high-frequency activity (HFA), decreased low-frequency activity (LFA) and decreased low-theta relative to silent periods of memory search ("deliberations") in the 500 ms preceding item vocalization in the hippocampus. Intrusions show an attenuation of the HFA and LFA effects (purple). Gray regions mark the frequency ranges analyzed: low-theta, low-frequency activity (LFA), and high-frequency activity (HFA). (*D*) Hippocampal LFA for each retrieval type as a function of intrusions' source (*Top Left*) or semantic (*Bottom Left*) similarity to the recently encoded list; we illustrate both source and semantic similarity in the *Right* panel. Different patients contributed to these analyses based on their meeting our minimum recall rates for each retrieval type (*Materials and Methods*). (C, correct recall; PLI, prior-list intrusion; ELI, extralist intrusion; S-I, semantic intrusion; nS-I, nonsemantic intrusion; S-PLI, semantic prior-list intrusion; nS-PLI, nonsemantic prior-list intrusion; S-ELI, semantic extralist intrusion; D, deliberation). Error bars represent ±1 SE of the mean. $*P < 0.05$, $***P < 0.001$ in a linear mixed-effects model, FDR corrected.

of the committed intrusion to the correctly recalled context. Increased LFA for contextually dissimilar relative to contextually similar intrusions was specific to the posterior portion of the hippocampus (*SI Appendix*, Fig. S5*B*). These results resonate with the proposal that the posterior hippocampus is biased toward pattern separation, while the anterior portion is biased toward pattern completion (50). The findings further support our conclusion that hippocampal LFA underlies retrieval of items bound to a given context, and are in line with the view that items associated with a similar context also share greater neural similarity. While LFA does not differentiate correct from false recalls sharing both source and semantic similarities to the target context (S-PLIs), it does differentiate between false recalls that differ by a single type of contextual similarity (either source or semantic), with LFA decrease for S-PLIs relative to nS-PLI and S-ELIs. This context-dependent LFA decrease maps with the fact that false recollection tends to increase for memories with similar attributes (51). Notably, the translation of this hippocampal LFA marker into a behavioral output (i.e., retrieval) is likely the result of a coordinated activity with neocortical areas involved in the source monitoring process (52).

Oscillatory, Rather Than Broadband, Activity Underlies Contextual Similarity. Standard measures of power extraction conflate oscillatory activity with a dominant $1/f^{\alpha}$ spectral pattern characteristic of neural signals (53), whereby power decreases with increasing frequency (54). This so-called broadband "tilt" manifests itself as a descending straight line on a log–log plot of the neural power and stands in contrast to oscillatory activity which spans narrow frequency bands (55, 56). If LFA reduction reflects a separate memory process from HFA, it should constitute an oscillatory desynchronization rather than emerge due to broadband tilt. To disentangle between oscillatory and broadband processes, we used the irregular-resampling auto-spectral analysis (IRASA) (57). IRASA takes advantage of mathematical properties of the $1/f^{\alpha}$ activity to separate it from the neural signal. By subtracting this $1/f^{\alpha}$ component from the overall power spectrum, a pure oscillatory measure can be obtained. We therefore used this method to isolate the oscillatory activity in each of the three frequency bands of interest, as well as extract the two parameters describing the broadband activity: the slope and intercept of the linear fit to the data.

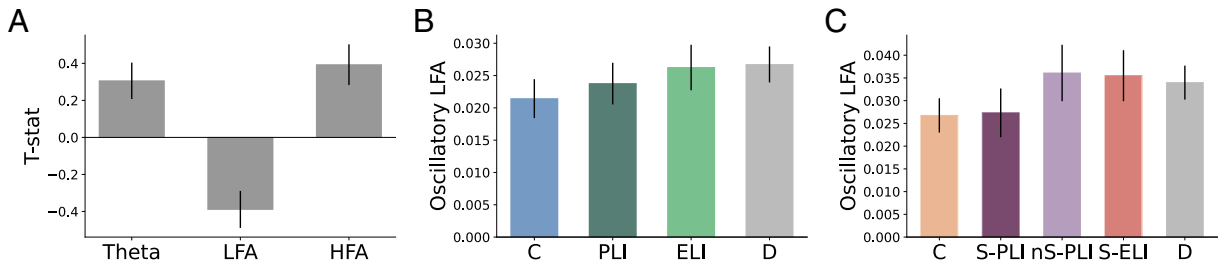

Fig. 4. Oscillatory effects of false recall in the hippocampus. (*A*) Averaged within-subject *t*-statistic of the difference between the oscillatory power spectrum of correct recalls and intrusions, as generated using the IRASA. Correct recalls exhibit increased theta, decreased low-frequency activity (LFA), and increased high-frequency activity (HFA) oscillations in comparison to intrusions. (*B*) Averaged hippocampal LFA oscillations of each retrieval type in the uncategorized word list. (*C*) Averaged hippocampal LFA oscillations of each retrieval type in the categorized word list. (C, correct recall; PLI, prior-list intrusion; ELI, extralist intrusion; S-PLI, semantic prior-list intrusion; nS-PLI, nonsemantic prior-list intrusion; S-ELI, semantic extralist intrusion; D, deliberation). Error bars represent ±1 SE of the mean.

We first wanted to determine whether the broadband or the oscillatory components accounted for the differences observed between correct recalls and intrusions. We did not observe any reliable differences between the uncategorized and categorized experiments in either the broadband parameters (all $P's > 0.681$) or oscillatory activity (all $P's > 0.302$). We therefore report here the results aggregated across the two experiments (*SI Appendix*, Fig. S3 shows the results for each experiment separately). We found that both the slope ($z = 1.10$, $P = n.s.$) and intercept ($z = -0.60$, $P = n.s.$) of the broadband activity did not differ between correct recalls and intrusions. In contrast, oscillatory activity reliably distinguished between correct recalls and intrusions, with higher theta ($z = 4.40$, $P < 0.001$), lower LFA ($z = -5.17$, $P < 0.001$), and higher HFA ($z = 4.97$, $P < 0.001$) for correct recalls relative to intrusions (Fig. 4*A*). This pattern of results is similar to the one obtained from the mixed neural signal, suggesting that the initially observed effects were due to oscillatory activity.

Next, we sought to test whether oscillatory LFA differentiated between intrusions sharing high or low contextual similarity with the target context. In the uncategorized experiment, we found that oscillatory LFA was higher for PLIs ($z = 2.37$, $P = 0.04$) and ELIs ($z = 3.08$, $P = 0.003$) in comparison to correct recalls (Fig. 4*B*). Oscillatory LFA activity also differentiated between recall types in the categorized experiment, with S-PLIs showing the strongest LFA reduction, which was greater in comparison to nS-PLIs ($z = -3.24$, $P = 0.003$) and in comparison to S-ELIs ($z = -3.23$, $P = 0.003$) (Fig. 4*C*). Note that unlike the spectral analyses conducted on the mixed signal, whereby we standardize the data over 2,400 ms and then extract the averaged power across the time window of interest, the IRASA algorithm extracts the oscillatory power across the full iEEG time series chosen for the analysis. As such, the oscillatory activity obtained does not capture the relative decrease in LFA over time. However, the relative oscillatory activity for correct vs. other false recall types reveals the same patterns of decreased LFA as a function of contextual similarity between the current context and retrieved item. This similar pattern of results suggests that oscillatory LFA, rather than a tilt in the broadband power spectrum, underlies contextual similarity between the retrieved item and target context.

Temporal Specificity of Hippocampal Biomarkers of Recall Veridicality. Evaluating spectral features in the 500 ms preceding vocalization revealed that while both low-theta, LFA and HFA differentiate correct from false recalls, LFA specifically reflects the similarity between the retrieved context and the current context. If hippocampal LFA is indeed the driving force behind context-dependent retrieval, we should find this signal specifically in the moments preceding vocalization. This prediction aligns with prior studies indicating the presence of context reinstatement in the moments prior to retrieval (58, 59). Alternatively, if the hippocampal activity differentiating correct from false recalls lacks temporal specificity, it may reflect an ongoing mental state which predisposes individuals to commit false recalls, such as inattentiveness or fatigue (60). Here, we investigated this question by looking at HFA, LFA, and low-theta in the two seconds surrounding vocalization onset (a time window that avoided contamination from adjacent vocalizations). Comparing free recall of uncategorized and categorized lists failed to reveal any reliable differences in temporal specificity for either HFA ($t(298) = 0.29$, $P = n.s.$), LFA ($t(298) = 0.38$, $P = n.s.$) or low-theta ($t(298) = -0.96$, $P = n.s.$). We therefore conducted the time specificity analysis on the aggregated data across the two experiments (*SI Appendix*, Fig. S2 shows the results of the uncategorized and categorized free-recall experiments, separately). Cluster permutation testing determined time windows

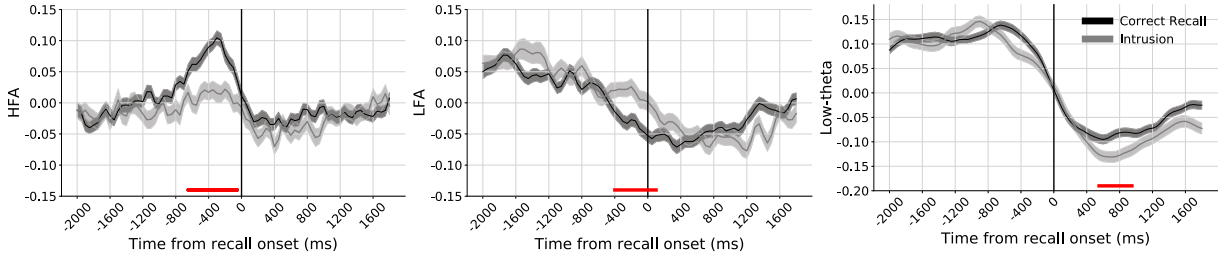

Fig. 5. Temporal specificity of hippocampal biomarkers of false memories. We measured HFA (*Left*), LFA (*Middle*), and low-theta (*Right*) at the 2 s surrounding item vocalization. Red marks on the x-axis represent time windows of significant difference between correct recalls (black) and intrusions (gray) (cluster permutation test, $P < 0.05$). shaded areas represent ±1 SE of the mean.

of significant differences between correct and false recalls (See *Temporal Specificity Analysis* under the *Materials and Methods*).

As Fig. 5 illustrates, the increased LFA reflective of intrusions relative to correct recalls was time-specific, appearing 450 ms prior to and disappearing 150 ms following vocalization. Decreased hippocampal HFA during intrusions relative to correct recalls emerged at 650 ms prior to vocalizations and dissipated at the beginning of vocalization. These time windows overlap with the time in which accessing the lexical representation of a word from the mental lexicon was estimated to occur during word production (61). The emergence of a recall veridicality biomarker prior to retrieval, as reflected in the LFA and HFA signal, strengthens the hypothesis that this hippocampal activity supports the retrieval process. In contrast, low-theta differences between correct recalls and intrusions emerged only following vocalization, between 500 and 1,850 ms, suggesting that it does not drive retrieval. To directly compare time specificity between the three frequency ranges, we predicted the difference between correct recalls and intrusions based on frequency (HFA/LFA/low-theta) and time (pre/post vocalization). We found a significant interaction effect ($F(2, 228) = 6.32, P = 0.002$). Post hoc analysis showed that, specifically for LFA, the difference between correct recalls and intrusions was higher pre vs. post vocalization (see *Temporal Specificity Analysis* for more details).

Multivariate Prediction of False Memories. Our analyses implicate two hippocampal processes that may be involved in generating false memories: the absence of HFA indicative of context-mediated retrieval itself as in prior work (17), and low-frequency synchronization indicative of the mismatch in contextual similarity of the memory to the target list. These findings emerge from a large group of participants with recordings from the hippocampus and indicate that on average, these signatures reflect properties of the ongoing retrieval process. However, they do not provide insight into how discriminable false memories are at the level of individual recalls. Further, they do not consider interactions between frequency bands or brain regions when relating hippocampal function to false memory. We thus applied multivariate classification techniques as in our prior work (62, 63) to ask whether neural states could reliably predict the veridicality of memories. Specifically, we asked whether whole-brain neural signals preceding overt recall can predict recall type and investigated which neural features were the most influential in these predictions.

We first focused on classifying recalls on uncategorized lists to avoid the influence of semantic similarity introduced in the categorized experiment (Fig. 6*A*). Classifiers trained on whole-brain neural patterns (spectral power ranging from 2 to 100 Hz, see *Materials and Methods*) identified brain states that significantly predicted correct recalls (AUC = 0.55 ± 0.01, $t(111) = 5.52, P < 0.0001$), PLIs (AUC = 0.53 ± 0.02, $t(59) = 2.25, P = 0.03$) , and ELIs (AUC = 0.54 ± 0.02, $t(55) = 3.13, P = 0.003$), as determined by permutation testing.

We also examined whether neural signals could discriminate correct and false memories on the categorized free-recall task (Fig. 6*B*), including false memories with varying degrees of semantic similarity to the target list. We trained classifiers to distinguish between all types of retrieval events: correct recalls, S-PLIs, nS-PLIs, and S-ELIs. Whole-brain classifiers predicted correct recalls (AUC = 0.59 ± 0.04, $t(24) = 2.62, P = 0.02$), S-PLIs (AUC = 0.57 ± 0.04, $t(20) = 2.22, P = 0.04$), and nS-PLIs (AUC = 0.62 ± 0.03, $t(12) = 3.49, P = 0.004$) at significantly above chance levels. S-ELIs could not be discriminated (AUC = 0.54 ± 0.02, $t(16) = 1.29, P = 0.21$) from other recalls at above chance levels. These results demonstrate that moment-to-moment changes in brain state are sufficient to identify individual, upcoming false memories.

We constructed forward models (64) to identify which brain regions and frequencies informed the predictions of each classifier. Fig. 6*C*. depicts neural sources that meaningfully covaried with classification of different retrieval types, across both experimental tasks. Effects in the hippocampus were comparable to our univariate analyses. Increases in HFA predicted correct

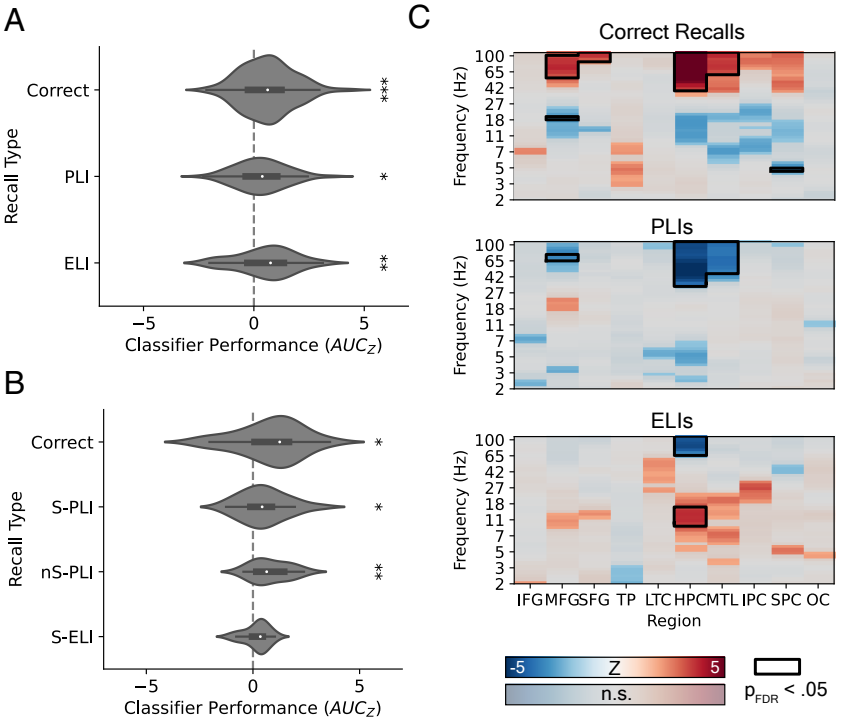

Fig. 6. Multivariate prediction of false memories. Logistic regression models distinguish correct from false memories based on local field potentials recorded from hippocampus and neocortical recording sites. (*A*) Classifier performance in the uncategorized free-recall task. Significant prediction of correct recalls, prior-list intrusions (PLI), and extralist intrusions (ELI) are denoted with asterisks. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. (*B*) Classifier performance in the categorized free-recall task. Significant prediction of correct recalls, semantic PLIs (S-PLI), nonsemantic PLIs (nS-PLI), and semantic ELIs (S-ELI) are denoted as in (*A*). (*C*) Forward-model estimates of feature importance broken down by anatomical regions of interest. Significant ($P < .05$) increases and decreases in power important for classification of correct recalls (*Top*), PLIs (*Middle*), and ELIs (*Bottom*) appear in red and blue, respectively. Features that survive FDR correction are outlined in black. Nonsignificant (n.s.) features are masked in gray. 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.

recalls, with reduced HFA indicating an upcoming intrusion. Increased hippocampal LFA was specifically predictive of ELIs, but not PLIs. These effects were not specific to the hippocampus, as we observed similar HFA effects within the MTL and prefrontal cortex (both *Middle* and superior frontal gyri).

Separately examining informative features on each task revealed similar patterns, irrespective of the semantic composition of lists (*SI Appendix*, Fig. S6). After correction for multiple comparisons, we did not find evidence for regions where neural activity specifically predicted the semantic attributes of false memories. However, with a more liberal (uncorrected) statistical threshold, decreased LFA was predictive of S-PLIs, as in our univariate analyses.

We found that brain signals preceding overt recall can reliably predict the veridicality of the to-be-recalled item and can sometimes even predict the contextual source of falsely recalled information (i.e., whether it is a PLI, ELI, S-PLI, or nS-PLI). While extending the analyses to widespread brain regions limits the statistical power of this analysis (due to multiple comparison corrections), it provides a broader look at the neural states predictive of contextual misattribution. Although the free-recall task leads to a significantly lower number of false recalls relative to paradigms designed to elicit false recalls (e.g., the DRM paradigm), single-subject classification was significant in 20.53% of participants in the uncategorized ($P < 0.001$, binomial test) and in 32% of participants in the categorized ($P < 0.001$, binomial test) experiment. These results suggest that predicting the commission of a false recall on a single-subject level is a plausible endeavor that could be optimized using tasks designed for that purpose.

Discussion

Work in both humans (49, 65–67) and in nonhuman primates (68, 69) implicates the hippocampus in context-dependent memory retrieval. Failure of contextually mediated retrieval can lead to erroneous recall of items that do not belong to a target context. Here, we asked whether hippocampal signals preceding item vocalization predict the veridicality of the to-be-recalled item, and whether these hippocampal signals differentiate between false recalls as a function of their contextual similarity with the target context. To answer these questions, we analyzed hippocampal depth electrode recordings captured while human subjects performed two variants of a free recall task: uncategorized and categorized free recall. These analyses revealed striking electrophysiological correlates of context-dependent memory retrieval, distinguishing the imminent retrieval of correct items from false recalls. We found that correct recalls exhibited increased low-theta (2 to 5 Hz), decreased LFA (6 to 18 Hz), and increased HFA (44 to 100 Hz) relative to false recalls in the moments leading up to vocalization. The contextual similarity of the false recall to the correct context did not influence the degree of low-theta or HFA increases. In contrast, false recalls that shared greater contextual similarity with the correct context also exhibited a greater LFA reduction, similar to the LFA reduction characteristic of correct recalls. This LFA reduction was present both when the false memory shared a similar contextual source (being encoded at a prior phase of the experiment) and when it shared semantic attributes with the recently encoded list. These findings suggest that the degree of hippocampal LFA reduction reflects the degree of correspondence between the retrieved item's context and the target context. The difference between correct and false recalls in hippocampal LFA emerged specifically at the moments preceding memory recall and faded rapidly afterward, supporting the involvement of hippocampal LFA in the retrieval of items using their associated context.

The observed LFA reduction prior to contextually mediated retrieval aligns with recognition memory studies showing LFA reduction especially during associative memory reinstatement (33, 70). It has been proposed that LFA desynchronization in the neocortex aids memory by increasing information coding of material-specific information, leading to a gamma synchronization in the hippocampus (32). Our findings show that LFA desynchronization is apparent not only in the neocortex but also in the hippocampus, where it represents not the content of the memory (i.e., memory for items) but the similarity between the retrieved and target context.

The association of items with their encoded context is demonstrated not only from the greater decrease in the hippocampal LFA for items sharing greater similarity with the target context but also from participants' behavioral performance in the task. Participants' false memories tended to share at least one type of contextual similarity with the desired context (either source or semantic, see *SI Appendix*, Table S1), implying that such contextual similarity led to their erroneous retrieval. It has long been shown that competitive interference between memories frequently results from their associations to similar contexts (71–73). Computational models of memory posit that during learning the hippocampus associates features representing each item with a dynamic representation of spatio-temporal context. During memory search, the current context cues item retrieval (45, 46), explaining why items learned under similar contexts may be falsely recalled. Indeed, exposing participants to an item's encoding context not only boosts correct item recognition, but also leads to false recognition of similar items that were never actually learned (74, 75). While the influence of contextual interference on memory is well established, the neural processes giving rise to such contextual interference remain elusive. The present findings offer a possible mechanism for such interference; as greater contextual similarity with the target context is reflected by greater LFA reduction, it may lead to a reduced signal-to-noise ratio in discriminating between correct and contextually similar false recalls. Alternatively, LFA reduction for both correct and contextually similar false recalls may reflect enhanced fidelity of these retrieved items in comparison to contextually dissimilar false recalls (31), rendering them more likely to be retrieved. LFA reduction could also signal enhanced confidence in the retrieved response. We investigated this option by using output position as a proxy of response confidence (40). These analyses demonstrated that the difference between correct recalls and intrusions persists even when controlling for output position, rendering this option unlikely.

Contrary to the greater LFA reduction with increased contextual similarity, HFA increases for correct relative to false recalls remained a strong predictor of recall veridicality regardless of the false recall type. Increased HFA during correct relative to false recalls, irrespective of the degree of contextual similarity between them, follows a large corpus of iEEG studies showing a widespread increase in HFA during diverse memory related processes (17, 18, 20, 23, 41, 42, 76), as well as studies suggesting that HFA reflects a domain-general marker of brain activation (28, 77).

The present study also uncovered a low-theta increase for correct relative to false recalls. This finding supports the recent suggestion that averaging theta across the traditional 3 to 8 Hz range can mask a memory-related increase in slow theta,

e.g., 3 Hz, (37). The functional role that low-theta plays in memory retrieval is still under debate (36), with some studies suggesting a distinct role of low (~3 Hz) and high (~8 Hz) theta for memory in humans (78, 79). Our findings strengthen the notion that low, rather than high, theta is specifically involved in associative retrieval (80). Our temporal specificity analysis showed that increased low-theta for correct relative to false recalls was especially evident postvocalization. This observed increase in low-theta may reflect patients' postvocalization assessment of the accuracy of their retrieved response (81, 82), or the associative strength between the current and subsequent retrievals (49).

In studying the spectral correlates of intrusions in recall of categorized and unrelated word lists we first adopted standard spectral decomposition methods. This allowed us to relate our findings to many earlier studies that have compared spectral power under different encoding and retrieval contrasts. Recent methodological advances, however, provide separate indices of narrow-band oscillatory activity and broadband (nonoscillatory) modulations of the EEG. Specifically, the FOOOF method applies an iterative procedure to fitting the $1/f^{\alpha}$ power spectrum resulting from autocorrelated EEG activity. This method then separates oscillatory components that exceed this background spectrum, similar to the earlier BOSC technique (83, 84). The IRASA similarly attempts to separate broadband and narrowband components. Rudoler et al. recently used this approach to demonstrate distinct modulations of hippocampal low-theta and high-theta activity during memory encoding and retrieval (37). In this paper, we chose to use IRASA to remain consistent with this prior work and avoid a possible distortion in broadband estimation due to low-frequency oscillations in the data (53). We found that the spectral patterns characteristic of false recall result from oscillatory activity rather than a tilt in the broadband power spectrum. Although the possibility of HFA and LFA dependency can not be ruled out, the findings that they reflect oscillatory activity and that LFA alone covaries with contextual similarity support the proposal that HFA and LFA reflect two distinct memory processes rather than a unitary phenomenon. Prior findings also support this view, as these oscillations were found to emerge at different times, in different neural regions (85) and to be deferentially influenced by the type of learned material (26).

Multivariate classification of memory retrieval established that neural signals preceding vocalization can reliably distinguish correct from false recalls, and demonstrated that even a task that elicits a paucity of false recalls can lead to significant classification of recall veridicality on a single-subject level. Tasks that induce false recalls, such as the DRM procedure, could be used to optimize such classification performance. Predicting the commission of false memories on a single-subject level is particularly important for intervention aimed at reducing false recalls when those induce significant distress or functional impairment. Individuals suffering from stress-related psychopathology, such as posttraumatic stress disorder, often experience memory intrusions of their traumatic experiences under contexts that are safe and dissimilar to the traumatic incident (86–88). Targeted interventions that disrupt retrieval of intrusive memories could spawn novel therapies for such clinical conditions (89, 90). Whether the false memory biomarkers shown in this study are causally involved in the generation of false recall, or are a correlate of false memories, is still an open question. Future studies could test this causal vs. correlative relationship by modulating those biomarkers, for example by delivering intracranial stimulation during brain states predictive of false recall manifestation (91–94). Animal models have already demonstrated the causal role of the hippocampus in context-dependent retrieval, as artificial activation of context-specific cells in the hippocampal dentate gyrus led to falsely recalling the memory encoded in the activated context (95).

Our work demonstrates that neural signals preceding overt recall can predict the veridicality of the to-be-recalled information, as well as the type of contextual misattribution in the case of false memories. The results point at LFA synchronization in the hippocampus as a specific marker of the mismatch between the current and retrieved context, therefore providing a neural signature for our ability to distinguish between similar memories that were formed on different occasions.

Materials and Methods

Intracranial Recordings. We analyzed data from subdural grids and strips (intercontact spacing 10.0 mm) or depth electrodes (intercontact spacing 2.2 to 10.0 mm) in patients undergoing surgical treatment for intractable epilepsy. All experimental protocols were approved by the Institutional Review Board of one of eight participating hospitals. Hospitals included Thomas Jefferson University Hospital (Philadelphia, PA), University of Texas Southwestern Medical Center (Dallas, TX), Emory University Hospital (Atlanta, GA), Dartmouth-Hitchcock Medical Center (Lebanon, NH), Hospital of the University of Pennsylvania (Philadelphia, PA), Mayo Clinic (Rochester, MN), NIH (Bethesda, MD), and Columbia University Hospital (New York, NY). Informed consent was obtained from all study participants. Electrode localization was confirmed via careful examination of high-resolution magnetic resonance images by qualified members of the clinical team.

We recorded brain activity while participants completed one of two experimental paradigms; 1. Uncategorized free-recall task, or 2. Categorized free-recall task. In the uncategorized free recall, a list of nouns (12 or 15 words per list) were displayed on a screen for 1,600 ms, sequentially. Words in each list were drawn from a 300-word pool and were comprised according to an algorithm that generated unique lists with a low semantic relation between them [mean pairwise Latent Semantic Analysis similarity (96) within list was ~0.2] (Fig. 1*A*). In the categorized free recall, 12 words were displayed on a screen for 1,600 ms, sequentially. Items were drawn from 25 distinct semantic categories. Each list included two same-category pairs drawn from three randomly chosen semantic categories (Fig. 3*A*). The categorized word pool was generated using Amazon Mechanical Turk to crowdsource typical exemplars for each semantic category (97). In both tasks, a 10-s countdown preceded the encoding phase of each list. Following encoding, patients completed a 20-s math distractor task consisting of a series of arithmetic problems of the form A+B+C=? (A, B, C were random integers from 1 to 9). Finally, during the recall phase (30 s), patients were required to recall as many words as possible from the most recent list, in any order. Participants' recalls were recorded via a microphone and later annotated offline using the Penn TotalRecall (http://memory.psych.upenn.edu/TotalRecall) software. Each vocalization onset time was determined as well as recall identity (a correct or false recalls). On average, participants ran in 2.3 sessions and studied 20.5 lists per session in the uncategorized free-recall experiment and in 2.5 sessions and 18 lists per session in the categorized free-recall experiment.

Intracranial EEG Data Preprocessing and Spectral Decomposition. To minimize confounds resulting from volume conduction, we analyzed the iEEG using bipolar referencing (98, 99), in which the difference in voltage between pairs of immediately adjacent electrodes is computed (22). The signal from each of these resulting bipolar signals was sampled at a minimum of 500 Hz (range: 500 to 1,600 Hz). A fourth-order 2-Hz stop-band Butterworth notch filter was applied to remove electrical line noise at 60 Hz.

We applied the Morlet wavelet transform (wave number 4) to compute spectral power as a function of time for all iEEG signals ranging from 2,500 ms and up to 100 ms preceding vocalization for the preretrieval analysis, or from 2,500 ms preceding and up to 2,500 ms following vocalization for the temporal specificity analysis (*Intracranial EEG Data Statistical Analyses*).

Frequencies were sampled logarithmically between 2 and 100 Hz, yielding a total of 46 frequencies. We included a mirrored buffer period of 1,500 ms on both sides of the data to minimize edge effects and to prevent iEEG activity measured during vocalization from bleeding into the time window of interest (100). After log transforming the power values, the data were downsampled by taking a moving average across 100-ms time windows and sliding the window every 50 ms (resulting in 47 time intervals for the preretrieval analysis, and in 99 time intervals for the temporal specificity analysis). Power values were standardized within each session, and separately for each electrode and frequency, by subtracting the mean and dividing by the SD across all retrieval events and time points. We excluded from our analysis repetitions of previously recalled items (yielding an average of 0.01 excluded recalls per list in both the uncategorized and categorized experiments). To avoid contamination from prior vocalizations, retrievals that occurred within less than 3,000 ms from the preceding recall were excluded from the analyses (yielding an average of 2.8 and 3.4 excluded recalls per list in the uncategorized and categorized experiments, respectively). The number of excluded trials did not differ significantly between the two experiments (all $P's > 0.1$). Furthermore, we excluded participants who had less than five correct recalls and five intrusions per session. Overall, using our inclusion criteria, we included 6,389 recalls across patients in the uncategorized experiment (28.5% of all total recalls) and 3,061 recalls in the categorized experiment (27.44% of all total recalls). We then divided the data into three classes of retrieval events: correctly recalled items, intrusions (items recalled that were not from the preceding list), and deliberation periods. Deliberation periods were 500-ms intervals of silence from 2,000 to 1,500 ms preceding vocalization, during which participants were attempting to recall items but made no overt vocalizations (17). For intrusions and correct recalls, we collapsed power across the 500-ms interval preceding vocalization (from −600 to −100 ms) to not include signals associated with speech production. This time window was chosen to remain consistent with prior work (17, 63, 101) and as it was shown to feature the most prominent network-wide power changes during retrieval (39). For patients who had at least five prior (PLI) and five extralist (ELI) intrusions (*Spectral Correlates of Intrusions Reflect Their Contextual Similarity*), or at least five semantic (S-I) and five nonsemantic (nS-I) intrusions (*Intrusions Semantic Categorization Procedure*), power for these different intrusion types was computed using the same method.

In the uncategorized free-recall experiment, 197 patients met these inclusion criteria. Of these, 101 patients (256 sessions) had depth electrodes in the hippocampal formation (CA fields, dentate gyrus, and the subiculum). Out of the 101 patients, 65 (167 sessions) had at least five PLIs and five ELIs. In the categorized free-recall experiment, 152 participants met the inclusion criteria. Of these, 54 patients (104 sessions) had hippocampal coverage. 27 patients (58 sessions) had at least five PLIs and five ELIs and 34 had at least five S-I and five nS-I. Electrode placement was determined solely based on clinical needs.

After applying our trial inclusion criteria, each participant in the uncategorized free recall had, on average, 33 correct recalls, 24 intrusions, and 57 deliberations. In the categorized free recall, each participant had, on average, 38 correct recalls, 19 intrusions, and 57 deliberations across sessions.

Intracranial EEG Data Statistical Analyses.

Preretrieval hippocampal biomarker analysis. Modulation of HFA, LFA, or low-theta as a function of retrieval type was analyzed on a trial-by-trial basis using a linear mixed-effects model (*SI Appendix*, Eq. **S2**). Linear mixed effects models were run using the MixedLM function in the package statsmodels in Python (102), and always included a random intercept for each session, nested in participant. In the uncategorized experiment, we added retrieval type of interest (either correct recall/intrusion/deliberation, or correct recall/PLIs/ELIs for the source similarity model) as a fixed effect. The main effect of retrieval type on each frequency range was evaluated using the likelihood ratio test between the full model and an intercept-only model. In the categorized experiment, a similar model was run with the retrieval types of interest being semantic PLIs/nonsemantic PLIs/ semantic ELIs. Since intrusions often arrive later during the retrieval phase relative to correct recalls (*SI Appendix*, Fig. S1), we next added each retrieval's output position as a secondary fixed effect to the above models to control for this possible confounding variable on the results (*SI Appendix*, Eq. **S3**). Significance of retrieval type beyond output position was evaluated using the likelihood ratio test between a reduced model containing only output position and a full model, containing both output position and retrieval type. All reported *P*-values were FDR corrected to account for the three frequency bands tested.

Temporal Specificity Analysis. To determine the temporal specificity of hippocampal biomarkers, we extracted the power signal for HFA, LFA, or low-theta at each time point from 2 s prior to 2 s following vocalization. We then tested whether there were any reliable differences in time specificity between the uncategorized and categorized free-recall experiments. Following previous studies (42), we computed the maximum *t*-statistic of the comparison between correct recalls and intrusions for each participant across trials. Independent sample *t* tests were then used to compare the distribution of maximum time-points between the uncategorized and categorized free-recall experiments for each frequency of interest (HFA, LFA, or low-theta). Since no differences in time specificity were found between the two experiments (*Results* and *SI Appendix*, Fig. S2), data were collapsed across the two experiments. To determine the temporal specificity of the difference between correct recalls and intrusions, data were permuted between conditions (correct recalls/intrusions) 1,000 times and the maximum cluster size, calculated as the sum of *t*-values, was extracted from each permutation. Cluster size of the observed data was then compared to the permuted distribution. Clusters exceeding the 5% threshold of the permuted distribution (two-sided) were considered significant (103). To determine whether there was a significant difference between frequencies in time specificity, we predicted the delta between correct recalls and intrusions as a function of time (averaged activity during pre/post retrieval) and frequency (HFA/LFA/low-theta) (*SI Appendix*, Eq. **S5**). A post hoc test (Tukey's test at α=0.05) was administered for pairwise comparisons.

Irregular-Resampling Auto-Spectral Analysis. To separate the oscillatory components of the neural power spectrum from broadband activity, we applied the IRASA to the 500 ms preceding vocalization (from −600 to −100 ms) of every retrieval event. This time window was chosen to match the time window used in the Morlet wavelets preretrieval analysis. We used a relatively conservative set of resampling factors ranging from 1.1 to 2.0, linearly spaced by 0.05, as recommended in the original methods paper (57). After separating the oscillatory and $1/f$ components, we quantified the intercept and slope from the linear regression parameters of the IRASA-decomposed fractal spectrum. The low-theta, LFA, and HFA oscillations were the averaged power across 2 to 5 Hz, 6 to 18 Hz, and 44 to 100 Hz, respectively, from the IRASA-decomposed oscillation spectrum.

We used either the broadband parameters (intercept and slope) or oscillatory components (low-theta, LFA, HFA) as dependent variables in a linear mixed-effects model. Retrieval type of interest (either correct recall/intrusion/deliberation, correct recall/PLIs/ELIs for the source similarity model, or semantic PLIs/nonsemantic PLIs/ Semantic ELIs for the semantic similarity model) was inserted as a fixed effect and sessions nested in participants were added as a random intercept effect. Differences between the categorized and uncategorized experiments were assessed by inserting experiment as an additional independent variable into the model and evaluating the event type X experiment interaction effect in each of the models (*SI Appendix*, Eq. **S4**).

Multivariate classification. We trained ridge regression models to discriminate between correct and false recalls using brain signals at the moments preceding vocalization. In the uncategorized free-recall experiment, classifiers were trained to discriminate between correct recalls, PLIs, and ELIs. In the categorized free-recall experiment, classifiers were trained to discriminate between correct recalls, semantic PLIs, nonsemantic PLIs, and semantic ELIs (see *Spectral Correlates of Intrusions Reflect Their Semantic Similarity* for details about these intrusion types). We used spectral power estimated in 46 intervals from 2 to 100 Hz averaged from 600 to 100 ms before vocalization onset as input features. We used L2 regularization, selecting the regularization strength based on prior work identifying successful retrieval states (97) to avoid overfitting (i.e., $C = 0.0007$). To avoid potential issues with class imbalance, the loss function was weighted proportionally to the number of observations in each class. We fit either three (in the uncategorized free-recall experiment) or four (in the categorized free-recall experiment) models for each participant, trained to distinguish each of the above retrieval types from all other retrieval events. To ensure sufficient training data and generalization of findings, we evaluated prediction accuracy in held-out sessions, using leave-one-session-out cross-validation. We excluded sessions without at least two observations per condition in each training fold.

After excluding these sessions, we analyzed data from participants with at least three sessions. 112 participants met these inclusion criteria in the uncategorized experiment, and 25 met these criteria in the categorized experiment. The area under the curve (AUC) (104, 105) measured predictive accuracy in a matter insensitive to the class distributions. Permutation testing ($N = 1{,}000$ shuffles of condition labels) determined classifier significance, allowing standardized AUC measures (AUC_Z) based on the mean and SD of surrogate distributions. We performed group inference through t tests of these standardized measures. To determine whether the number of participants exhibiting significant decoding accuracy was above the chance level, we used a one-way binomial test contrasting the number of participants with significant classification ($P < 0.05$) versus 5% chance (expected under the permutation null).

To assess the importance of different frequency bands and brain regions to classification performance, we constructed forward models for each participant based on trained classifiers (64). As in prior work from our group (62), we estimated model-based activation (**A**) as the product of the covariance matrix of input features ($\mathbf{\Sigma_X}$) with classifier weights (**W**) normalized by the variance of classifier predictions ($\hat{\sigma}_y^2$). We fitted linear mixed-effects models for each feature to estimate average effects across tasks, with each retrieval type as a fixed effect and subject as a random intercept and slope. False discovery rate ($q = 0.05$) corrected for multiple comparisons across frequencies and regions (106).

Intrusions Semantic Categorization Procedure. In the categorized free-recall experiment (Fig. 3*A*), we divided intrusions into those that belong to at least one of the three semantic categories presented during list encoding ("semantic intrusions") and those that do not relate to any of the encoded categories ("nonsemantic intrusions"). To achieve this, we manually coded each ELI conducted by participants as associated with either: 1. one (or more) of the 25 semantic categories from which words in the categorized free recall were drawn, 2. "None" – if the word did not belong to any of the 25 semantic categories. The semantic category associated with each word was selected by two independent raters. Interrater reliability was 92%. Only words for which agreement was achieved between raters were used in the analyses.

Intrusions belonging to one or more of the three semantic categories present during encoding of the preceding list were considered semantic intrusions (S-I), while intrusions not belonging to any of the three encoded semantic categories were considered nonsemantic intrusions (nS-I). The same procedure was applied for PLIs, though no manual categorization was needed for these intrusions as they were part of the existing word pool.

To verify that the categorized structure successfully induced semantic intrusions, we computed the mean semantic similarity of each intrusion to the recently encoded list. Semantic similarity values were obtained using word2vec, a pretrained word embedding model (107). Each word in the word2vec space is represented by 300-dimensional vectors. We computed the cosine-theta semantic similarity vector distance between a target intrusions and the 12 items encoded in the recent list and then averaged these values to obtain a single semantic similarity measure. We used a linear mixed effects model to predict semantic similarity based on experiment type (categorized/uncategorized; *SI Appendix*, Fig. S1*A*) or based on intrusion type (PLI/ELI; *SI Appendix*, Fig. S1*B*). A random intercept of session, nested within subject, was used in these models.

Data, Materials, and Software Availability. Data were collected as part of the Defense Advanced Research Projects Agency, Restoring Active Memory initiative. All processed data, along with analysis code, may be freely obtained from the senior author's website: https://memory.psych.upenn.edu/Data (108).

ACKNOWLEDGMENTS. We are grateful to the patients for their participation and thank hospital staff and researchers who were involved in data acquisition. This work was supported by the NIH grant U01-NS113198.

1. M. W. Howard *et al*., A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *J. Neurosci.* **34**, 4692–4707 (2014).
2. C. Ranganath, Binding items and contexts: The cognitive neuroscience of episodic memory. *Curr. Dir. Psychol. Sci.* **19**, 131–137 (2010).
3. M. E. Hasselmo, H. Eichenbaum, Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Networks* **18**, 1172–1190 (2005).
4. K. J. Mitchell, M. S. Zaragoza, Repeated exposure to suggestion and false memory: The role of contextual variability. *J. Mem. Lang.* **35**, 246–260 (1996).
5. M. Chang, C. Brainerd, Semantic and phonological false memory: A review of theory and data. *J. Mem. Lang.* **119**, 104210 (2021).
6. H. L. Roediger, K. B. McDermott, Creating false memories: Remembering words not presented in lists. *J. Exp. Psychol.: Learn. Mem. Cogn.* **21**, 803–814 (1995).
7. D. R. Kimball, T. A. Smith, M. J. Kahana, The fSAM model of false recall. *Psychol. Rev.* **114**, 954–993 (2007).
8. D. E. Warren, S. H. Jones, M. C. Duff, D. Tranel, False recall is reduced by damage to the ventromedial prefrontal cortex: Implications for understanding the neural correlates of schematic memory. *J. Neurosci.* **34**, 7677–7682 (2014).
9. A. S. Atkins, P. A. Reuter-Lorenz, Neural mechanisms of semantic interference and false recognition in short-term memory. *NeuroImage* **56**, 1726–1734 (2011).
10. E. Díeza, C. J. Gómez-Ariza, A. M. Díez-Álamo, M. A. Alonso, A. Fernandez, The processing of semantic relatedness in the brain: Evidence from associative and categorical false recognition effects following transcranial direct current stimulation of the left anterior temporal lobe. *Cortex* **93**, 133–145 (2017).
11. M. A. Friehs, C. Greene, B. Pastötter, Transcranial direct current stimulation over the left anterior temporal lobe during memory retrieval differentially affects true and false recognition in the DRM task. *Eur. J. Neurosci.* **54**, 4609–4620 (2021).
12. R. Cabeza, S. M. Rao, A. D. Wagner, A. R. Mayer, D. L. Schacter, Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4805–4810 (2001).
13. S. Slotnick, D. Schacter, A sensory signature that distinguishes true from false memories. *Nat. Neurosci.* **7**, 664–672 (2004).
14. S. M. Stark, M. A. Yassa, C. E. L. Stark, Individual differences in spatial pattern separation performance associated with healthy aging in humans. *Learn. Mem.* **17**, 284–288 (2010).
15. J. M. Karanian, S. D. Slotnick, False memory for context activates the parahippocampal cortex. *Cognit. Neurosci.* **5**, 186–192 (2014).
16. H. Kim, R. Cabeza, Trusting our memories: Dissociating the neural correlates of confidence in veridical versus illusory memories. *J. Neurosci.* **27**, 12190–12197 (2007).
17. N. M. Long *et al*., Contextually mediated spontaneous retrieval is specific to the hippocampus. *Curr. Biol.* **27**, 1074–1079 (2017).
18. J. F. Burke *et al*., Theta and high-frequency activity mark spontaneous recall of episodic memories. *J. Neurosci.* **34**, 11355–11365 (2014).
19. N. M. Long, J. F. Burke, M. J. Kahana, Subsequent memory effect in intracranial and scalp EEG. *NeuroImage* **84**, 488–494 (2014).
20. N. M. Long, M. J. Kahana, Successful memory formation is driven by contextual encoding in the core memory network. *NeuroImage* **119**, 332–337 (2015).
21. E. A. Solomon *et al*., Widespread theta synchrony and high-frequency desynchronization underlies enhanced cognition. *Nat. Commun.* **8**, 1704 (2017).
22. J. F. Burke *et al*., Synchronous and asynchronous theta and gamma activity during episodic memory formation. *J. Neurosci.* **33**, 292–304 (2013).
23. B. J. Griffiths *et al*., Directional coupling of slow and fast hippocampal gamma with neocortical alpha/beta oscillations in human episodic memory. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21834–21842 (2019).
24. J. R. Manning, J. Jacobs, I. Fried, M. J. Kahana, Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J. Neurosci.* **29**, 13613–13620 (2009).
25. R. Mukamel *et al*., Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* **309**, 951–954 (2005).
26. M. C. Fellner *et al*., Spectral fingerprints or spectral tilt? Evidence for distinct oscillatory signatures of memory formation. *PLoS Biol.* **17**, e3000403 (2019).
27. P. Fries, Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* **32**, 209–224 (2009).
28. J. F. Burke, A. G. Ramayya, M. J. Kahana, Human intracranial high-frequency activity during memory processing: Neural oscillations or stochastic volatility? *Curr. Opin. Neurobiol.* **31**, 104–110 (2015).
29. B. Merker, Cortical gamma oscillations: The functional key is activation, not cognition. *Neurosci. Biobehav. Rev.* **37**, 401–417 (2013).
30. S. Hanslmayr, T. Staudigl, M. Fellner, Oscillatory power decreases and long-term memory: The information via desynchronization hypothesis. *Front. Hum. Neurosci.* **6**, 74 (2012).
31. B. J. Griffiths *et al*., Alpha/beta power decreases track the fidelity of stimulus-specific information. *eLife* **8**, e49562 (2019).
32. S. Hanslmayr, B. P. Staresina, H. Bowman, Oscillations and episodic memory: Addressing the synchronization/desynchronization conundrum. *Trends Neurosci.* **39**, 16–25 (2016).
33. M. C. Martín-Buro, M. Wimber, R. N. Henson, B. P. Staresina, Alpha rhythms reveal when and where item and associative memories are retrieved. *J. Neurosci.* **40**, 2510–2518 (2020).
34. A. E. Karlsson, C. C. Wehrspaun, M. C. Sander, Item recognition and lure discrimination in younger and older adults are supported by alpha/beta desynchronization. *Neuropsychologia* **148**, 107658 (2020).
35. B. S. Katerman, Y. Li, J. K. Pazdera, C. Keane, M. J. Kahana, EEG biomarkers of free recall. *NeuroImage* **246**, 118748 (2022).
36. N. A. Herweg, E. A. Solomon, M. J. Kahana, Theta oscillations in human memory. *Trends Cognit. Sci.* **24**, 208–227 (2020).
37. J. H. Rudoler, N. A. Herweg, M. J. Kahana, Hippocampal theta and episodic memory. *J. Neurosci.* **43**, 613–620 (2023).
38. S. M. Polyn, M. J. Kahana, Memory search and the neural representation of context. *Trends Cognit. Sci.* **12**, 24–30 (2008).
39. E. A. Solomon *et al*., Dynamic theta networks in the human medial temporal lobe support episodic encoding and retrieval. *Curr. Biol.* **29**, 1100–1111 (2019).

40. J. Jou, Recall latencies, confidence, and output positions of true and false memories: Implications for recall and metamemory theories. *Lang. J. Mem.* **58**, 1049–1064 (2008).
41. J. A. Greenberg, J. F. Burke, R. Haque, M. J. Kahana, K. A. Zaghloul, Decreases in theta and increases in high frequency activity underlie associative memory encoding. *NeuroImage* **114**, 257–263 (2015).
42. J. F. Burke *et al*., Human intracranial high-frequency activity maps episodic memory formation in space and time. *NeuroImage* **85**, 834–843 (2014).
43. D. Zeithamova, A. R. Preston, Temporal proximity promotes integration of overlapping events. *J. Cognit. Neurosci.* **29**, 1311–1323 (2017).
44. C. Ranganath, L. T. Hsieh, The hippocampus: A special place for time. *Ann. N. Y. Acad. Sci.* **17**, 65–70 (2016).
45. M. W. Howard, M. J. Kahana, A distributed representation of temporal context. *J. Math. Psychol.* **46**, 269–299 (2002).
46. S. M. Polyn, K. A. Norman, M. J. Kahana, A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* **116**, 129–156 (2009).
47. M. M. El-Kalliny *et al*., Changing temporal context in human temporal lobe promotes memory of distinct episodes. *Nat. Commun.* **10**, 203 (2019).
48. M. A. Yassa, C. E. L. Stark, Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–525 (2011).
49. E. A. Solomon, B. C. Lega, M. R. Sperling, M. J. Kahana, Hippocampal theta codes for distances in semantic and temporal spaces. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24343–24352 (2019).
50. J. Poppenk, H. R. Evensmoen, M. Moscovitch, L. Nadel, Long-axis specialization of the human hippocampus. *Trends Cognit. Sci.* **17**, 230–240 (2013).
51. J. H. Coane *et al*., Manipulations of list type in the DRM paradigm: A review of how structural and conceptual similarity affect false memory. *Front. Psychol.* **12**, 668550 (2021).
52. K. J. Mitchell, *The Cognitive Neuroscience of Source Monitoring*, J. Dunlosky, U. K. Tauber, Sarah, Eds. (Oxford University, 2015).
53. M. Gerster *et al*., Separating neural oscillations from aperiodic 1/f activity: Challenges and recommendations. *Neuroinformatics* **20**, 991–1012 (2022).
54. B. J. He, Scale-free brain activity: Past, present, and future. *Trends Cognit. Sci.* **18**, 480–487 (2014).
55. B. Voytek, R. T. Knight, Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. *Biol. Psychiatry* **77**, 1089–1097 (2015).
56. K. J. Miller *et al*., Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *NeuroImage* **85**, 711–720 (2014).
57. H. Wen, Z. Liu, Separating fractal and oscillatory components in the power spectrum of neurophysiological signal. *Brain Topogr.* **29**, 13–26 (2016).
58. R. B. Yaffe *et al*., Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18727–8732 (2014).
59. S. J. Gershman, A. C. Schapiro, A. Hupbach, K. A. Norman, Neural context reinstatement predicts memory misattribution. *J. Neurosci.* **33**, 8590–8595 (2013).
60. S. A. Dewhurst, C. Barry, E. R. Swannell, S. J. Holmes, G. L. Bathurst, The effect of divided attention on false memory depends on how memory is tested. *Mem. Cognit.* **34**, 660–667 (2007).
61. P. Indefrey, W. Levelt, The spatial and temporal signatures of word production components. *Cognition* **92**, 101–104 (2004).
62. Y. Ezzyat *et al*., Direct brain stimulation modulates encoding states and memory performance in humans. *Curr. Biol.* **27**, 1251–1258 (2017).
63. J. E. Kragel *et al*., Similar patterns of neural activity predict memory function during encoding and retrieval. *NeuroImage* **155**, 60–71 (2017).
64. S. Haufe *et al*., On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* **87**, 96–110 (2014).
65. R. A. Diana, A. P. Yonelinas, C. Ranganath, Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends Cognit. Sci.* **11**, 379–386 (2007).
66. J. F. Miller *et al*., Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science* **342**, 1111–1114 (2013).
67. L. Davachi, Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* **16**, 693–700 (2006).
68. J. Bachevalier, S. Nemanic, M. C. Alvarado, The influence of context on recognition memory in monkeys: Effects of hippocampal, parahippocampal and perirhinal lesions. *Behav. Brain Res.* **285**, 89–98 (2015).
69. J. J. Sakon, Y. N. Sylvia Wirth, W. A. Suzuki, Context-dependent incremental timing cells in the primate hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18351–18356 (2014).
70. B. P. Staresina *et al*., Hippocampal pattern completion is linked to gamma power increases and alpha power decreases during recollection. *eLife* **5**, e17397 (2016).
71. J. R. Anderson, G. H. Bower, Interference in memory for multiple contexts. *Mem. Cognit.* **2**, 509–514 (1974).
72. F. M. Zaromb *et al*., Temporal associations and prior-list intrusions in free recall. *J. Exp. Psychol.: Learn. Mem. Cognit.* **32**, 792–804 (2006).
73. C. M. MacLeod, *Oxford Handbook of Human Memory*, M. J. Kahana, A. D. Wagner, Eds. (Oxford University Press, Oxford, UK, ed. 2, 2022).
74. M. K. Doss, J. K. Picart, D. A. Gallo, The dark side of context: Context reinstatement can distort memory. *Psychol. Sci.* **29**, 914–925 (2017).
75. M. Racsmány, D. Bencze, P. Pajkossy, Á. Szőllősi, M. Marián, Irrelevant background context decreases mnemonic discrimination and increases false memory. *Sci. Rep.* **11**, 6204 (2021).
76. N. M. Long, M. J. Kahana, Modulation of task demands suggests that semantic processing interferes with the formation of episodic associations. *J. Exp. Psychol.: Learn. Mem. Cognit.* **43**, 167–176 (2017).
77. J. P. Lachaux, N. Axmacher, F. Mormann, E. Halgren, N. E. Crone, High-frequency neural activity and human cognition: Past, present, and possible future of intracranial EEG research. *Progr. Neurobiol.* **98**, 279–301 (2012).
78. A. Goyal *et al*., Functionally distinct high and low theta oscillations in the human hippocampus. *Nat. Commun.* **11**, 2469 (2020).
79. B. Lega, J. Jacobs, M. Kahana, Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus* **22**, 748–761 (2012).
80. S. Kota, M. D. Rugg, B. C. Lega, Hippocampal theta oscillations support successful associative memory formation. *J. Neurosci.* **40**, 9507–9518 (2020).
81. S. C. Wynn, E. Nyhus, Brain activity patterns underlying memory confidence. *Eur. J. Neurosci.* **55**, 1774–1797 (2022).
82. S. C. Wynn, S. M. Daselaar, R. P. Kessels, D. J. Schutter, The electrophysiology of subjectively perceived memory confidence in relation to recollection and familiarity. *Brain Cognit.* **130**, 20–27 (2019).
83. M. J. Kahana, R. Sekuler, J. B. Caplan, M. Kirschen, J. R. Madsen, Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature* **399**, 781–784 (1999).
84. J. B. Caplan *et al*., Human theta oscillations related to sensorimotor integration and spatial learning. *J. Neurosci.* **23**, 4726–4736 (2003).
85. V. S. Marks *et al*., Independent dynamics of low, intermediate, and high frequency spectral intracranial EEG activities during human memory formation. *NeuroImage* **245**, 118637 (2021).
86. D. T. Acheson, J. E. Gresack, V. B. Risbrough, Hippocampal dysfunction effects on context memory: Possible etiology for posttraumatic stress disorder. *Neuropharmacology* **62**, 674–685 (2012).
87. C. R. Brewin, J. D. Gregory, M. Lipton, N. Burgess, Intrusive images in psychological disorders: Characteristics, neural mechanisms, and treatment implications. *Psychol. Rev.* **117**, 210–232 (2010).
88. R. T. Cohen, M. J. Kahana, A memory based theory of emotional disorders. *Psychol. Rev.* **129**, 742–776 (2022).
89. N. Herz *et al*., Neuromodulation of visual cortex reduces the intensity of intrusive memories. *Cerebral Cortex* **32**, 408–417 (2022).
90. L. Iyaduraia *et al*., Intrusive memories of trauma: A target for research bridging cognitive science and its clinical application. *Clin. Psychol. Rev.* **69**, 67–82 (2019).
91. J. Jacobs *et al*., Direct electrical stimulation of the human entorhinal region and hippocampus impairs memory. *Neuron* **92**, 983–990 (2016).
92. M. E. Lacruz *et al*., Single pulse electrical stimulation of the hippocampus is sufficient to impair human episodic memory. *Neuroscience* **170**, 623–632 (2010).
93. S. G. Coleshill *et al*., Material-specific recognition memory deficits elicited by unilateral hippocampal electrical stimulation. *J. Neurosci.* **24**, 1612–1616 (2004).
94. M. B. Merkow *et al*., Stimulation of the human medial temporal lobe between learning and recall selectively enhances forgetting. *Brain Stimul.* **10**, 645–650 (2017).
95. S. Ramirez *et al*., Creating a false memory in the hippocampus. *Science* **341**, 387–391 (2013).
96. T. K. Landauer, P. W. Foltz, D. Laham, An introduction to latent semantic analysis. *Discourse Process.* **25**, 259–284 (1998).
97. C. T. Weidemann *et al*., Neural activity reveals interactions between episodic and semantic memory systems during retrieval. *J. Exp. Psychol.: General* **148**, 1–12 (2019).
98. P. L. Nunez, R. Srinivasan, *Electric Fields of the Brain* (Oxford University Press, New York, NY, 2006).
99. C. K. Kovach *et al*., Manifestation of ocular-muscle EMG contamination in human intracranial recordings. *NeuroImage* **54**, 213–233 (2011).
100. M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice* (MIT Press, Cambridge, MA, 2014).
101. P. B. Sederberg *et al*., Gamma oscillations distinguish true from false memories. *Psychol. Sci.* **18**, 927–932 (2007).
102. S. Seabold, J. Perktold, "Statsmodels: Econometric and statistical modeling with Python" in *Proceedings of the 9th Python in Science Conference* (2010), vol. 57.
103. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
104. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning* (Springer-Verlag, New York, NY, 2001).
105. K. A. Spackman, "Signal detection theory: Valuable tools for evaluating inductive learning" in *Proceedings of the Sixth International Workshop on Machine Learning* (1989), pp. 160–163.
106. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
107. T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space. arXiv [Preprint] (2013). http://arxiv.org/abs/1301.3781 (Accessed 25 February 2023).
108. N. Herz, B. R. Bukala, J. E. Kragel, M. J. Kahana, Hippocampal activity predicts contextual misattribution in false memories. Cognitive Electrophysiology Data Portal. https://memory.psych.upenn.edu/Publications. Deposited 20 April 2023.