

Hippocampal mechanisms of false recall

Noa Herz (✉ herz.noa@gmail.com)

University of Pennsylvania

Bernard Bukala

University of Pennsylvania <https://orcid.org/0000-0002-9582-949X>

James Kragel

University of Chicago

Michael Kahana

University of Pennsylvania

Article

Keywords: false memory, context, hippocampus, free-recall

Posted Date: August 24th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1946629/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Hippocampal mechanisms of false recall

Noa Herz¹, Bernard R. Bukala¹, James E. Kragel²
and Michael J. Kahana¹

¹Department of Psychology, University of Pennsylvania,
Philadelphia, 19104, PA, USA.

²Department of Neurology, University of Chicago, Street,
Chicago, 60637, IL, USA.

Contributing authors: noaherz@sas.upenn.edu;
bernard.bukala@nhs.net; jkrangel@uchicago.edu;
kahana@psych.upenn.edu;

Abstract

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 contextually-mediated retrieval, the neural process leading to false recalls is not yet understood. Using direct electrical recordings from the human hippocampus, we investigate the neural mechanisms underlying the false recall phenomenon. In two large datasets, we characterize key physiological differences between correct and false recalls, emerging immediately prior to vocalization. By differentiating between false recalls that share high or low contextual similarity with the target context, we identify the neural process underlying retrieval of item-context associations. Applying multivariate decoding methods, we were able to reliably predict whether the to-be-recalled item would be a veridical or false memory. Our findings provide a mechanistic insight into the process of retrieving context-bound memories, and open new avenues for interventions aimed at reducing false recalls when those lead to functional impairment.

Keywords: false memory, context, hippocampus, free-recall

1 Introduction

When remembering an event, we may retrieve contextual features representing the time and place of its occurrence. Our ability to encode and retrieve the temporal and situational context of our past experiences is a hallmark of episodic memory. Numerous studies have shown that the hippocampus plays a crucial role in both the encoding and retrieval of episodic memories, and that hippocampal reactivation underlies contextual reinstatement [1–4]. Although episodic memory retrieval often reflects veridical recall, false recall can happen when we retrieve items outside of their associated context. Consider, for example, thinking about your last birthday while recalling a birthday cake that was actually served on a different occasion. If the hippocampus stores the associations of items with their encoded contexts, hippocampal activity should discriminate between correct and false recalls. In line with this proposal, multiple lines of evidence suggest that while distinct neural pathways store context and item information, the hippocampus specifically represents the relation between the two [2, 5–9]. Computational models have also suggested the distinct representations of items and context, with the hippocampus forming the associations between them [3, 10]. Supporting the causal role of the hippocampus in context-dependent retrieval, artificial activation of context-specific cells in the hippocampal dentate gyrus led to falsely recalling the memory encoded in the activated context [11]. Here, we tested the prediction that hippocampal neural activity in the moments preceding memory retrieval reliably differentiates correct from false memories. To do so we analyzed direct electrical recordings from the human hippocampus of neurosurgical epileptic patients as they studied and subsequently recalled lists of items. We further tested the prediction that false memories sharing greater contextual similarity with the correct information will show less discriminable hippocampal activity from correct recalls.

Previous electrophysiological studies demonstrating hippocampal contributions to episodic memory recall have compared activity between correct recall and matched silent periods during memory search [12, 13]. These events, however, do not uniquely isolate the correlates of contextual retrieval as they also differ by the mere retrieval process, as well as by the motor activity associated with item vocalization. The comparison between correct and false retrievals allows a more controlled and nuanced investigation of the role the hippocampus plays in computing the association between retrieved items and their encoded context. Although not all neuroimaging studies have found hippocampal correlates of successful episodic retrieval ([14–16], but see [17]), recordings from hippocampal depth electrodes may provide a more direct readout of hippocampal physiology. Indeed, analyzing hippocampal depth electrode recordings, Long et al. found elevated high-frequency activity (HFA) (44–100 Hz) for correct relative to false recalls just prior to vocalization [18]. Similar HFA increases also emerged during encoding of subsequently remembered relative to forgotten information [13, 19], possibly reflecting successful item-context binding that enables later retrieval of the encoded information [20]. In these studies,

74 decreased low-frequency activity (LFA) typically accompanied increased HFA
75 [13, 21, 22]. The twin findings of increased HFA and decreased LFA may rep-
76 resent a ‘tilt’ in the broadband power spectrum, and different studies have
77 suggested that this neural pattern serves as a proxy for neuronal firing rate
78 [23, 24]. Whereas high-theta and alpha oscillations decrease prior to successful
79 recall, low-frequency theta oscillations sometimes increase [25, 26], an effect
80 that can be masked when aggregating across the full 2-8 Hz theta band [27].

81 Relying on the idea that the hippocampus supports retrieval using the
82 associative strength between items and their encoded context, here we hypoth-
83 esized that hippocampal activation will reflect the degree of similarity between
84 the target context and the contextual information of the retrieved item. While
85 correct responses should manifest the highest degree of contextual overlap,
86 false recalls should also vary in their degree of contextual similarity to the tar-
87 get context. For example, false recall of items encoded in a prior list in the
88 experiment share a more similar contextual information to the target context
89 relative to false recall of items never presented in the experimental session. We
90 predicted that false recalls sharing a high, compared to low, degree of contex-
91 tual similarity to the target context will manifest a higher degree of overlap in
92 their underlying hippocampal features to those reflective of correct recalls. To
93 test our predictions we used broad spectral features, including HFA (44-100
94 Hz), LFA (6-18 Hz) and low-theta (2-5 Hz), as well as multivariate classifica-
95 tion methods, to characterize the hippocampal activity distinguishing correct
96 from false recalls varying in their contextual similarity to the target context.
97 We show that hippocampal activity can reliably differentiate correct from false
98 retrievals, and that this activity emerges specifically in the moments (< 1 sec)
99 preceding memory retrieval and fades rapidly afterwards. We further show that
100 hippocampal LFA maps the degree of similarity between the falsely-recalled
101 item’s context and the target context, with greater LFA reduction signaling
102 greater overlap between the target context and the context of the retrieved
103 item.

104 2 Results

105 We report five major sets of analyses across two large studies of human hip-
106 pocampal electrophysiology. Analyzing data from free recall of unrelated word
107 lists, we first ask whether human hippocampal activity at the moments pre-
108 ceding memory retrieval could reliably differentiate correct from false recalls.
109 We then test how these observed biomarkers of recall veridicality change as
110 a function of the temporal similarity between the false recall and the target
111 context. Using data from free recall of semantically organized word lists, we
112 then test whether similar hippocampal biomarkers coding for temporal simi-
113 larity also code for the semantic similarity between the false recall and the
114 target context. Next, we ask whether the observed hippocampal biomarkers
115 of recall veridicality drive the retrieval process by testing for their emergence
116 specifically at the time prior to item vocalization. Finally, using multivariate

4 *Hippocampal mechanisms of false recall*

117 classification methods, we predict the type of the to-be-recalled information
118 and test whether such prediction is possible not only at the group level, but
119 also at a single-subject level.

120 **2.1 Hippocampal Activity Distinguishes Veridical Recall** 121 **from False Memories**

122 We first investigated whether hippocampal activity at the moments preceding
123 memory retrieval could reliably differentiate correct from false recalls. In line
124 with prior studies, we found that correct recalls exhibited a tilt in the power
125 spectrum, with increased HFA and decreased LFA relative to deliberation periods.
126 For false recalls, however, we found a reduced spectral-tilt strength relative
127 to correct recalls (Figure 1.D). To assess the statistical significance of these
128 effects, we predicted hippocampal power (either HFA, LFA or low-theta, sep-
129 arately) as a function of retrieval type using a linear mixed-effects model. We
130 found a main effect of retrieval type on HFA ($\chi^2_{(2)} = 317.323, p < .001$), with
131 decreased HFA for intrusions relative to correct recalls ($z = -6.460, p < .001$)
132 but increased relative to deliberations ($z = 8.599, p < .001$). In addition, we
133 found a main effect of retrieval type on LFA ($\chi^2_{(2)} = 58.741, p < .001$), with
134 increased LFA for intrusions relative to correct recalls ($z = 2.703, p = .010$)
135 but decreased relative to deliberations ($z = -3.746, p < .001$). In the low-theta
136 range, we found decreased low-theta power for intrusions relative to correct
137 recalls ($z = -2.426, p = .015$) and deliberations ($z = -6.362, p < .001$)
138 ($\chi^2_{(2)} = 45.210, p < .001$). All of these three effects remained significant after
139 including output position (i.e. the position of each recalled event throughout
140 the retrieval period of a given list) in the model (all p 's $< .05$), suggesting that
141 output position of the recalled event does not account for these hippocampal
142 biomarkers. These results demonstrate that correct recalls show a similar
143 spectral pattern to the one previously found during successful encoding, with
144 increased HFA and decreased LFA relative to deliberations [19, 28, 29]. These
145 same biomarkers are attenuated for false recalls, and reliably differentiate
146 between these two retrieval types prior to their commission.

147 **2.2 Spectral Correlates of Intrusions Reflect Their** 148 **Temporal Similarity**

149 Prominent theories [30–32], as well as single-unit observations [33, 34], sug-
150 gest the role of the hippocampus in the association of items with an intrinsic
151 and gradually drifting representation of time. If the hippocampus codes the
152 association of items with their temporal context, hippocampal activity during
153 retrieval of items encoded in greater temporal proximity to one another should
154 exhibit greater neural similarity. Therefore, the degree of separation between
155 correct and false recalls may be dependent on the similarity between the tem-
156 poral context in which these items were encoded. The free-recall paradigm
157 enables differentiation of intrusions based on the similarity of their associ-
158 ated temporal context to the recently encoded list. Prior-list intrusions (PLIs)

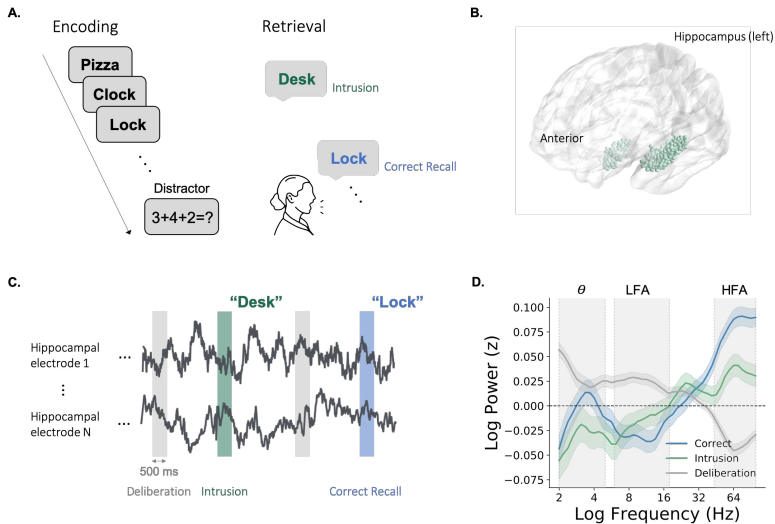


Fig. 1 Hippocampal biomarkers of false memory in uncategorized word lists. **A.** Uncategorized free-recall design. During encoding, semantically unrelated words were presented on the screen, sequentially. Participants were then required to recall as many words as they could remember from the recent list, in any order. During retrieval, participants either correctly recalled one of the recently presented words (correct recall) or falsely recalled a word not recently presented (intrusion). **B.** Multi-patient localization of hippocampal electrodes in the uncategorized free-recall task. **C.** Spectral power was computed across hippocampal electrodes during the -2500 to -100 ms preceding vocalization. Then, mean power across the 500 ms preceding vocalization (either correct recalls or intrusions) was extracted. Deliberation periods were 500 ms of ‘silence’ extracted from 2000 ms - 1500 ms preceding each vocalization. **D.** Correct recalls (blue) are characterized by increased high-frequency activity (HFA) and decreased low frequency activity (LFA) relative to deliberations in the 500 ms preceding recall in the hippocampus. This spectral-tilt effect is lower for intrusions (green). Gray background marks the frequency ranges used in the analyses (frequencies were log transformed to enable better inspection of activity at the low-frequency ranges).

159 reflect cases where subjects incorrectly recall an item that was not presented
 160 on the target list, but was presented in one of the prior lists of the experi-
 161 ment. Extra-list intrusions (ELIs), on the other hand, are intrusion of items
 162 never presented in the experiment. Since PLIs were encoded in a prior list,
 163 they share greater temporal similarity with the current list’s context relative
 164 to ELIs. If the hippocampus stores information about the temporal context
 165 in which items were encoded, we should expect PLIs to share a greater spec-
 166 tral similarity with correct recalls relative to ELIs. This prediction is in line
 167 with findings showing that events encoded in greater temporal proximity have
 168 higher chances of becoming linked into an integrated representation [35].

169 To test this hypothesis, we used a linear mixed-effects model predicting hip-
 170 pocampal power (HFA, LFA or low-theta separately) as a function of retrieval
 171 type while differentiating between intrusion types (PLIs/ELIs/correct recalls).
 172 When predicting HFA, we found a general HFA reduction for both intrusion
 173 types relative to correct recalls (PLIs vs. correct recalls: $z = -6.355$, $p < .001$;

6 *Hippocampal mechanisms of false recall*

174 ELIs vs. correct recalls: $z = -4.139$, $p < .001$), without any difference between
 175 PLIs and ELIs ($z = -1.651$, $p = 0.1$).

176 For LFA, however, both correct recalls and contextually similar intrusions
 177 (i.e. PLIs) exhibit reduced LFA relative to contextual dissimilar intrusions
 178 (ELIs) (ELIs vs. correct recalls: $z = 3.486$, $p < .001$; ELIs vs. PLIs: $z = 2.185$,
 179 $p = 0.028$). LFA does not distinguish between correct recalls and contextually
 180 similar intrusions (PLIs vs. correct recalls: $z = 0.842$, $p = 0.400$) (Figure 2.B).
 181 In the low-theta range, contextually similar and dissimilar intrusions exhibited
 182 similar low-theta reduction (PLIs vs. ELIs: $z = 1.200$, $p = 0.2$).

183 The results demonstrate that hippocampal LFA is gradually reduced as a
 184 function of intrusions' temporal similarity to the correct context, with PLIs
 185 showing greater LFA reduction relative to ELIs (Figure 2). These findings sug-
 186 gest that hippocampal LFA codes the association of items with the temporal
 187 context in which they were acquired, therefore allowing the distinction between
 188 items that were encoded under different temporal contexts. By distinguish-
 189 ing between memories that were encoded in temporally distinct episodes, the
 190 LFA reflects the role the hippocampus play in temporal organization of mem-
 191 ories (e.g. differentiating between our last birthday party and an event that
 192 occurred a week beforehand).

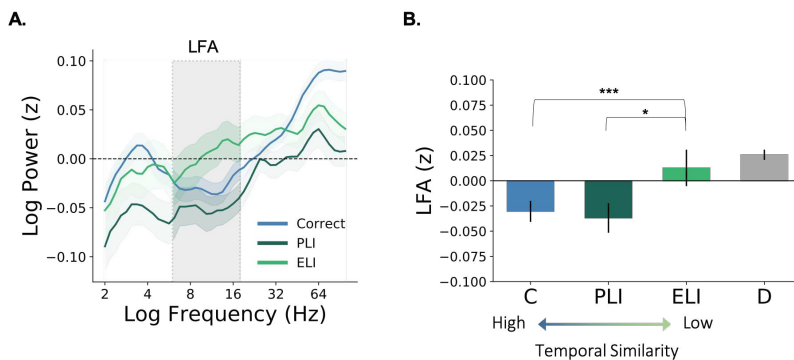


Fig. 2 Hippocampal LFA decreases as a function of intrusions' temporal similarity to the recently encoded list. **A.** Spectral power of correct recalls, PLIs and ELIs. PLIs exhibit similar LFA reduction to the one characterizing correct recalls. ELIs, on the other hand, does not show similar LFA reduction. Gray background marks the LFA range used in the analysis. **B.** Mean hippocampal LFA for each retrieval type. LFA decreases as a function of intrusions' temporal similarity to the recently encoded list (C; correct recalls, PLI; prior-list intrusion, ELI; extra-list intrusion, D; deliberations). Error bars represent ± 1 standard error of the mean. * $p < .05$, *** $p < .001$ in a linear mixed-effects model, FDR corrected.

193 2.3 Spectral Correlates of Intrusions Reflect Their 194 Semantic Similarity

195 To the extent that hippocampal decreases in LFA mark the successful rein-
 196 statement of temporal context, one might expect to find a similar LFA decrease

197 for the reinstatement of semantic context. We used an independent dataset in
198 which participants studied a list of items categorized into three semantic cat-
199 egories to test this prediction (see Figure 3.A. and ‘*Intracranial Recordings*’
200 section for more details on the experimental design).

201 We first sought to test whether the same hippocampal biomarkers differ-
202 entiating correct from false recall emerge in this new dataset of semantically
203 organized lists. Similar to the results obtained in the uncategorized experi-
204 ment (Figure 1.D), we again found that correct recalls exhibited increased
205 HFA and decreased LFA relative to deliberation periods. Relative to correct
206 recalls, false recalls had reduced spectral-tilt strength (Figure 3.C). When
207 predicting hippocampal power as a function of retrieval type in the catego-
208 rized free-recall experiment, we found a main effect of retrieval type on HFA
209 ($\chi^2_{(2)} = 140.729$, $p < .001$), with decreased HFA for intrusions relative to cor-
210 rect recalls ($z = -5.793$, $p < .001$) but increased relative to deliberations
211 ($z = 3.188$, $p = 0.001$). In addition, we found a main effect of retrieval type on
212 LFA ($\chi^2_{(2)} = 31.703$, $p < .001$), with decreased LFA for intrusions ($z = -3.141$,
213 $p = .001$) and correct recalls ($z = -5.433$, $p < .001$) relative to deliberations.
214 In the low-theta range, a similar main effect of retrieval type to the one found
215 for the LFA emerged, with decreased low-theta for correct recalls ($z = -2.423$,
216 $p = 0.015$) and intrusions ($z = -3.599$, $p < .001$) relative to deliberations
217 ($\chi^2_{(2)} = 14.755$, $p < .001$). All of these three main effects remained significant
218 after including output position in the model (all p 's $< .05$).

219 These results demonstrate that similar hippocampal biomarkers of recall
220 veridicality to those found in an uncategorized list emerge in an independent
221 dataset, where the encoded list contains semantically organized information.
222 Correct recalls exhibit increased HFA and decreased LFA relative to delib-
223 erations, effects that are attenuated for false recalls. The presence of similar
224 hippocampal biomarkers in this new dataset support the generalizability of
225 these biomarkers across participants and study conditions.

226 In the uncategorized experiment, we found that hippocampal LFA decrease
227 as a function of the temporal similarity between the intrusion and the target
228 context. Here we asked - does hippocampal LFA also decrease as a function
229 of the semantic similarity between the intrusion and the target context? We
230 tested this question by contrasting two subclasses of intrusions observed in
231 the categorized free recall experiment. Owing to the categorical nature of the
232 study lists in this experiment (e.g., flowers, insects, fruits), subjects would often
233 incorrectly recall non-studied items belonging to one of the studied categories
234 (e.g., a flower that was not presented on the target list). This allowed us to
235 compare intrusions that are semantically related to the encoded information to
236 those that lack such semantic relatedness. We hypothesized that the degree to
237 which intrusions reflect retrieval of semantic context will determine the mag-
238 nitude of the observed hippocampal LFA decrease. We therefore categorized
239 each intrusion committed by participants as either S-I (e.g. an intrusion that
240 was semantically related to at least one of the three semantic categories pre-
241 sented during encoding) or nS-I (an intrusion that was not related to any of the

three encoded categories) (see ‘*Intrusions Semantic Categorization Procedure*’ under the *Methods* section for more details on the categorization procedure).

When assessing the contribution of both temporal and semantic similarity factors to the intruded items in this experiment, 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 non-related intrusions (chi-square test of independence: $\chi^2_{(1)} = 112.317, p < .001$) (see Supplementary Table 1 for exact intrusions counts). This suggests that intrusions tend to share at least one source of contextual similarity with the encoded information (either temporal or semantic). Due to the rarity of intrusions that are neither temporal nor semantically related to the encoded information (i.e. non-semantic ELIs), we investigated whether HFA, LFA or low-theta change as a function of three intrusion types differing in their contextual similarity levels; semantic PLIs (S-PLI), non-semantic PLIs (nS-PLI) and semantic ELIs (S-ELI).

As expected, we found that hippocampal LFA gradually changed as a function of the contextual similarity between the intrusion and the recently encoded list ($\chi^2_{(2)} = 14.383, p < .001$). Specifically, S-PLIs showed the strongest LFA reduction, which was greater in comparison to nS-PLIs ($z = -2.513, p = 0.035$) and in comparison to S-ELIs ($z = -3.715, p < .001$). These effects remained significant after including output position in the model ($\chi^2_{(2)} = 15.057, p < .001$, S-PLI vs. nS-PLI: $z = -2.687, p = 0.021$, S-PLI vs. S-ELI: $z = -3.768, p < .001$). Figure 3 shows the mean LFA for each retrieval type as a function of either the temporal (3.D, top-left), semantic (3.D, bottom-left), or combined semantic and temporal similarity of the intrusion to the recently encoded list (Figure 3.D, right). In contrast to LFA, neither HFA ($\chi^2_{(2)} = 1.074, p = 0.584$) nor low-theta ($\chi^2_{(2)} = 2.0166, p = 0.338$) changed as a function of intrusions’ temporal and semantic similarity to the recently encoded list. These findings show that hippocampal LFA is modulated by both the semantic and the temporal similarity of the committed intrusion to the correctly-recalled context and suggest that hippocampal LFA is involved in associating items with both their temporal and semantic context. The findings further support our conclusion that hippocampal LFA underlies retrieval of items bounded to a given context, thus allowing us to distinguish between semantically similar experiences that happened at different points in time (e.g. differentiating between the cake served at our last birthday from the one served at our birthday two years ago).

2.4 Temporal Specificity of Hippocampal Biomarkers of Recall Veridicality

So far, we have tested whether broad spectral features (low-theta, LFA, HFA) differentiate correct from false retrievals in the 500 ms preceding vocalization. We found that while both HFA and LFA differentiate correct from false

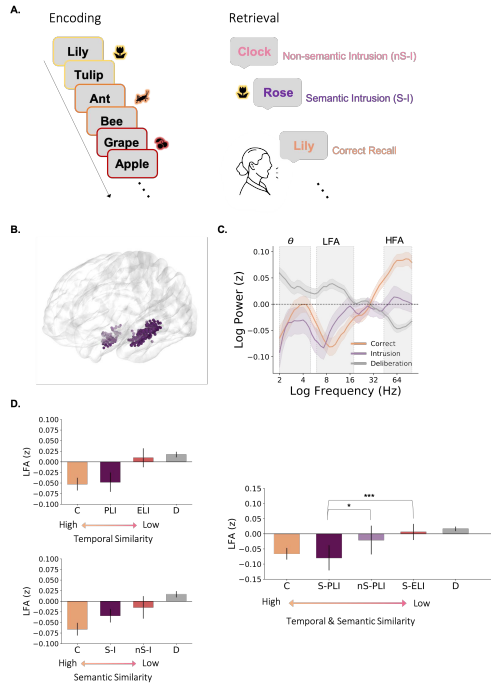


Fig. 3 Hippocampal biomarkers of false recall in categorized word lists. **A.** Categorized free-recall design. During encoding, 12 words drawn from three semantic categories (here: flowers, insects and fruits) were presented in pairs of two. During retrieval, participants either correctly recalled one of the presented words (e.g. ‘Lily’), committed an intrusion of a word semantically related to one of the encoded categories (e.g. ‘Rose’), or committed an intrusion non-semantically related to any of the encoded categories (e.g. ‘Clock’). **B.** Multi-patient localization of hippocampal electrodes in the categorized free-recall task. **C.** Correct recalls (orange) are characterized by increased high-frequency activity (HFA) and decreased low frequency activity (LFA) relative to deliberations in the 500 ms preceding recall in the hippocampus. This spectral-tilt effect is lower for intrusions (purple). Gray background marks the frequency ranges used in the analyses (frequencies were log transformed to enable better inspection of activity at the low-frequency ranges). **D.** Mean hippocampal LFA for each retrieval type as a function of intrusions’ temporal similarity to the recently encoded list (top left); intrusions’ semantic similarity to the recently encoded list (bottom left), or mean hippocampal LFA for each retrieval type as a function of the combined contribution of temporal and semantic factors (right). (C: correct recall, PLI; prior-list intrusion, ELI; extra-list intrusion, S-I: semantic intrusion, nS-I: non-semantic intrusion, S-PLI: semantic prior-list intrusion, nS-PLI: non-semantic prior-list intrusion, S-ELI: semantic extra-list intrusion, D: deliberation). Error bars represent ± 1 standard error of the mean. * $p < .05$, *** $p < .001$ in a linear mixed-effects model, FDR corrected.

284 recalls, LFA specifically codes the similarity between the retrieved item context
 285 and the current retrieval context. If the hippocampal LFA is indeed
 286 the driving force of retrieval of items bound to their encoded context, we
 287 should expect to find this hippocampal biomarker specifically at the moments
 288 preceding item retrieval. Alternatively, if the hippocampal activity differentiating
 289 correct from false recalls is not time specific, extending beyond the

retrieval event in time, it may reflect an ongoing memory state which predispose individuals for committing false recalls, such as a state of inattentiveness or fatigue. Here, we investigated this question by looking at HFA, LFA and low-theta at each time point beginning at two seconds preceding and up to two seconds following vocalization (a time window surrounding vocalization that was not contaminated by adjacent vocalizations). When comparing the uncategorized and categorized free-recall experiments, we did not find any reliable differences in temporal specificity between the two for neither HFA ($t(298) = 0.299, p = .765$), LFA ($t(298) = 0.385, p = .700$) or low-theta ($t(298) = -0.965, p = .334$). We therefore conducted the time specificity analysis on the aggregated data across the two experiments (Supplementary Figure 2 shows the results of the uncategorized and categorized free-recall experiments, separately).

As Figure 4 illustrates, we found that increased LFA of intrusions relative to correct recalls was time-specific, appearing at 450 ms prior to and disappeared 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. The emergence of a recall veridicality biomarker specifically prior to retrieval, as reflected in the LFA and HFA signal, support the hypothesis that this hippocampal activity drives the retrieval process. In contrast, low-theta differences between correct recalls and intrusions emerged only following vocalization, between 500 - 1850 ms. This post-vocalization effect suggests that hippocampal low-theta reflects a post-processing signal, rather than a signal that drives retrieval of items bound to a given context.

Overall, rather than an ongoing memory state that promotes the generation of false recalls, these findings support the role of the hippocampus in the retrieval of items bound to their encoded context and point to the time window preceding retrieval as a potential target for interventions aimed at reducing false recalls.

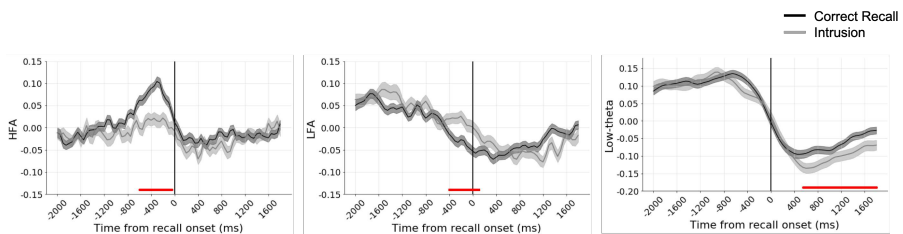


Fig. 4 Temporal specificity in the hippocampus. HFA (left), LFA (middle) and low-theta (right) measured at each time point from two seconds prior to two seconds following vocalization. Red marks on the x-axis represent time windows of significant difference between correct recalls (black) and intrusions (grey) (cluster permutation test, $p < .05$). shaded areas represent ± 1 standard error of the mean.

2.5 Predicting False Memories from Hippocampal and Neocortical Field Potentials

Next, we used multivariate prediction models to determine whether neural signals preceding retrieval can predict the type of the to-be-recalled event (Figure 5). We first focused on distinguishing correct from false recalls, irrespective of semantic or temporal attributes. Classifiers trained on patterns of hippocampal activity (spectral power ranging from 2 to 100 Hz, see *Methods*) identified retrieval events as either false or correct recalls in the uncategorized experiment ($AUC = 0.52 \pm 0.01$, $t(111) = 2.23$, $p = 0.03$) and on a trend level in the categorized experiment ($AUC = 0.59 \pm 0.04$, $t(24) = 2.06$, $p = 0.05$) as determined by permutation testing. Including additional signals from all available neocortical sites (excluding seizure onsets zones, epileptogenic sites, and other atypical recording sites) improved prediction of correct from false recalls in the uncategorized ($AUC = 0.55 \pm 0.01$, $t(111) = 2.45$, $p = 0.02$) but not the categorized ($AUC = 0.59 \pm 0.04$, $t(24) = 0.98$, $p = 0.33$) experiment.

We then asked whether we could predict whether a false recall came from the current session (PLI) or outside of the experiment (ELI). Similarly to the univariate analysis, we used the uncategorized experiment to differentiate PLIs and ELIs without the potential confound of semantic similarity introduced in the categorized experiment. Classifiers trained on hippocampal signals could discriminate ELIs on a trend level ($AUC = 0.53 \pm 0.02$, $t(55) = 1.78$, $p = 0.08$), and incorporating additional signals from the cortex did not reliably improve prediction ($AUC = 0.54 \pm 0.02$, $t(55) = 1.47$, $p = 0.14$). PLIs, on the other hand, could be reliably discriminated from other retrieval types using hippocampal signals alone ($AUC = 0.53 \pm 0.02$, $t(59) = 2.49$, $p = 0.02$), comparable to classifiers incorporating neocortical signals ($AUC = 0.53 \pm 0.02$, $t(59) = 0.03$, $p = 0.97$; see Figure 5.A).

We also examined whether hippocampal signals predicted the semantic relatedness of false memories to the current list, focusing on the categorized free-recall task. We trained classifiers to distinguish between each of these intrusions types from all other retrieval events: S-PLIs, nS-PLIs, and S-ELIs. Hippocampal classifiers predicted both S-PLIs ($AUC = 0.60 \pm 0.04$, $t(20) = 2.50$, $p = 0.02$) and nS-PLIs ($AUC = 0.59 \pm 0.04$, $t(13) = 2.19$, $p = 0.049$) at significantly above chance levels. Classifiers incorporating additional neocortical sites performed at comparable levels for both semantic ($AUC = 0.57 \pm 0.04$, $t(20) = -0.48$, $p = 0.63$) and non-semantic ($AUC = 0.62 \pm 0.04$, $t(13) = 0.74$, $p = 0.47$) PLIs. S-ELIs could not be reliably discriminated based on exclusively hippocampal or additional neocortical signals (all p 's > 0.21) (Fig. 5.B).

Overall, the findings suggest that brain signals preceding vocalization can reliably predict the veridicality of the to-be-recalled item, and can sometime even predict the type of the falsely-recalled information (e.g. whether it is a PLI, S-PLI or nS-PLI). While the classification of correct vs. false memories in the categorized free-recall experiment was marginal ($p = 0.050$), it is worth noting that the sample size in this experiment was significantly smaller relative

365 to the uncategorized free-recall. Interestingly, although hippocampal signals
 366 provided far fewer neural sources compared to aggregation across all neocortical
 367 sites (5% of all available electrodes), they were sufficient to identify false
 368 recalls at above chance levels in 15% of participants across experiments. The
 369 number of participants exhibiting significant classification using hippocampal
 370 signals alone was significant in both the uncategorized (14.28%, $p < 0.001$,
 371 binomial test) and categorized (20%, $p < 0.001$, binomial test) experiment.
 372 Using all available recording sites, classification was significant in 20.53% of
 373 participants in the uncategorized ($p < 0.001$, binomial test) and in 32% of
 374 participants in the categorized ($p < 0.001$, binomial test) experiment. These
 375 results indicate that biomarkers of recall veridicality can reliably predict the
 376 commission of a false recall even on a single-subject level. By targeting brain
 377 activity predictive of the imminent commission of false recall, future inter-
 378 ventions could interfere with the retrieval process to reduce false recalls when
 379 those lead to functional impairment.

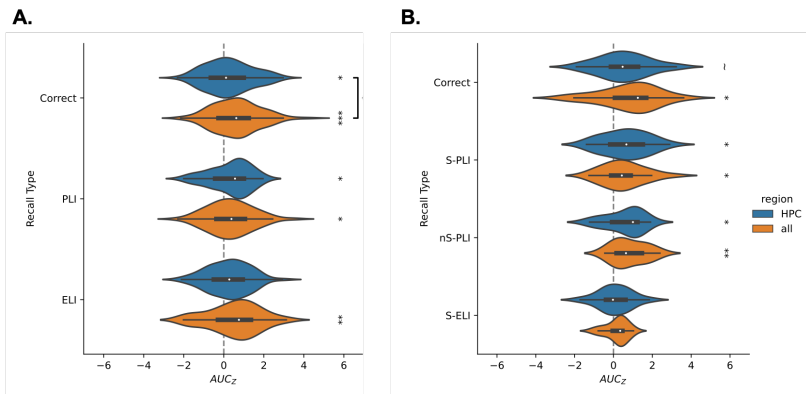


Fig. 5 Multivariate prediction of false memories. Logistic regression models reliably distinguish correct from false memories based on local field potentials recorded from the hippocampus (HPC, blue) and all available recording sites (all, orange). **(A)** Classifier performance in the uncategorized free-recall task. Significant prediction of correct recalls, prior-list intrusions (PLI) and extra-list intrusions (ELI) are denoted with asterisks. **(B)** Classifier performance in the categorized free-recall task. Significant prediction of correct recalls, semantic PLIs (S-PLI), non-semantic PLIs (nS-PLI) and semantic ELIs (S-ELI) are denoted with asterisks. $\sim p = 0.05$, $* p < .05$, $** p < .01$, $*** p < .001$.

380 3 Methods

381 3.1 Intracranial Recordings

382 We analyzed data from subdural grids and strips (intercontact spacing 10.0
 383 mm) or depth electrodes (intercontact spacing 2.2-10.0 mm) in patients under-
 384 going surgical treatment for intractable epilepsy. Data were recorded at eight

385 hospitals, including Thomas Jefferson University Hospital (Philadelphia, PA),
386 University of Texas Southwestern Medical Center (Dallas, TX), Emory Univer-
387 sity Hospital (Atlanta, GA), Dartmouth-Hitchcock Medical Center (Lebanon,
388 NH), Hospital of the University of Pennsylvania (Philadelphia, PA), Mayo
389 Clinic (Rochester, MN), National Institutes of Health (Bethesda, MD), and
390 Columbia University Hospital (New York, NY). All experimental protocols
391 were approved at participating hospitals or by the Institutional Review Board
392 at the University of Pennsylvania via a reliance agreement prior to data
393 collection. Informed consent was obtained from all study participants.

394 We recorded brain activity while participants completed one of two
395 experimental paradigms; 1. Uncategorized free-recall task, or 2. Categorized
396 free-recall task. In the uncategorized free-recall, a list of nouns (12 or 15
397 words per list) were displayed on a screen for 1,600 ms, sequentially. Words
398 in each list were drawn from a 300 word pool and were comprised accord-
399 ing to an algorithm that generated unique lists with a low semantic relation
400 between them (mean pairwise Latent Semantic Analysis similarity [36] within
401 list was 0.2) (Figure 1.A). In the categorized free-recall, 12 words were dis-
402 played on a screen for 1,600 ms, sequentially. Items were drawn from 25
403 distinct semantic categories. Each list included two same-category pairs drawn
404 from three randomly chosen semantic categories (Figure 3.A). The catego-
405 rized word pool was generated using Amazon Mechanical Turk to crowdsource
406 typical exemplars for each semantic category [37]. In both tasks, a ten sec-
407 ond countdown preceded the encoding phase of each list. Following encoding,
408 patients completed a 20 sec math distractor task consisting of a series of arith-
409 metic problems of the form $A+B+C=?$ (A,B,C were random integers from
410 1-9). Finally, during the recall phase (30 sec), patients were required to recall
411 as many words as possible from the most recent list, in any order. Patients
412 vocalizations were recorded and later annotated offline using Penn TotalRecall
413 (<http://memory.psych.upenn.edu/TotalRecall>) to determine correct or false
414 recalls. On average, participants participated in 2.3 sessions and studied 20.5
415 lists per session in the uncategorized free-recall experiment and in 2.5 sessions
416 and 18 lists per session in the categorized free-recall experiment.

417 **3.2 Intracranial EEG Data Preprocessing and Spectral** 418 **Decomposition**

419 To minimize confounds resulting from volume conduction, we analyzed the
420 iEEG using bipolar referencing [38, 39], in which the difference in voltage
421 between pairs of immediately adjacent electrodes is computed [21]. The signal
422 from each of these resulting bipolar signals was amplified and sampled at a
423 minimum of 500 Hz (range: 500 - 1600 Hz). A fourth order 2-Hz stop-band
424 Butterworth notch filter was applied to remove electrical line noise at either
425 50 Hz (for data collected in Germany) or 60 Hz (USA).

426 We applied the Morlet wavelet transform (wave number 4) to compute
427 spectral power as a function of time for all iEEG signals ranging from 2500 ms
428 and up to 100 ms preceding vocalization for the pre-retrieval analysis, or from

2500 ms preceding and up to 2500 ms following vocalization for the temporal specificity analysis (see *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 1500 ms on both sides of the data to minimized edge effects and to prevent iEEG activity measured during vocalization from bleeding into the time window of interest [40]. After log transforming, 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 pre-retrieval 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 standard deviation power. Mean and standard deviation were calculated across all retrieval events and time points. We excluded from our analysis repetitions of previously recalled items. To avoid contamination from prior vocalizations, retrievals that were within less than 3,000 ms from the preceding recall were excluded from the analyses. Furthermore, we excluded participants who had less than five correct recalls and five intrusions per session. We then divided the data to 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 [18]. 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. For patients who had at least five prior (PLI) and five extra-list (ELI) intrusions (see *"Spectral Correlates of Intrusions Reflect Their Temporal Similarity"*), or at least five semantic (S-I) and five non-semantic (nS-I) intrusions (see *'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). 65 patients (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. Electrode localization was confirmed via careful examination of high-resolution magnetic resonance images by qualified members of the clinical team.

After applying our trial inclusions 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.

3.3 Intracranial EEG Data Statistical Analyses

3.3.1 Pre-retrieval 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, with retrieval type of interest (either correct recall/intrusion/deliberation, correct recall/PLIs/ELIs for the temporal similarity model, or semantic PLIs/non-semantic PLIs/ Semantic ELIs for the semantic similarity model) as a fixed effect and sessions nested in participants as a random intercept effect. Main effect of retrieval type on each frequency range was evaluated using likelihood ratio test between the full model and an intercept-only model. Since intrusions often arrive later during the retrieval phase relative to correct recalls (see Supplementary Figure 1), we next added each retrieval's output position as a secondary fixed effect to the model to control for this possible confounding variable on the results. 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. Linear mixed effects models were run using the MixedLM function in the package statsmodels in Python [41]. All reported p-values were FDR corrected to account for the three frequency bands tested.

3.3.2 Temporal Specificity Analysis.

To determine the temporal specificity of hippocampal biomarkers, we extracted the power signal for either HFA, LFA or low-theta at each time point from two seconds prior to two seconds 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 [29], 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 (see *Results* and Supplementary Figure 2), 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) 1000 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 [42].

3.3.3 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, non-semantic 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 [37] to avoid overfitting (i.e., $C = 0.0007$). 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) [43, 44] measured predictive accuracy. Permutation testing ($N = 1000$ shuffles of condition labels) determined classifier significance, allowing standardized AUC measures (AUC_Z) based on the mean and standard deviation 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 chance level, we used a one-way binomial test contrasting the number of participants with significant classification ($p < .05$) versus 5% chance.

3.4 Intrusions Semantic Categorization Procedure

In the categorized free-recall experiment (Figure 3.A), we divided intrusions to 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 ('non-semantic intrusions'). For this purpose, 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. Inter-rater reliability was 92%. Only words for which agreement was achieved between raters were used in the analysis.

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

559 **3.5 Data and Code Sharing**

560 All raw and pre-processed data, along with our analysis code, may be freely
561 obtained from the senior author’s website: <http://memory.psych.upenn.edu>.

562 **4 Discussion**

563 Work in both humans [5, 45?–47] and in non-human primates [48, 49] impli-
564 cates the hippocampus in context-dependent memory retrieval. Failure of
565 contextually mediated retrieval can lead to erroneous recall of items that do
566 not belong to a target list. Here, we asked whether hippocampal signals pre-
567 ceding item vocalization predict the veridicality of the to-be-recalled item, and
568 whether these hippocampal signals differentiate between false recalls as a func-
569 tion of their contextual similarity with the correct information. To answer these
570 questions, we analyzed hippocampal depth electrode recordings captured while
571 human subjects performed two variants of a free recall task: uncategorized and
572 categorized free recall. These analyses revealed striking electrophysiological
573 correlates of context-dependent memory retrieval, distinguishing the imminent
574 retrieval of correct items from false recalls without relying on patients’ real-
575 ization that such memory errors were made. We found that correct recalls
576 exhibited decreased LFA (6-18 Hz) and increased HFA (44-100 Hz) relative to
577 false recalls in the moments leading up to vocalization. The contextual simi-
578 larity of the false recall to the correct context did not influence the degree of
579 these HFA increases. In contrast, false recalls that shared a higher contextual
580 similarity with the correct context also exhibited a greater LFA reduction,
581 similar to the LFA reduction characteristic of correct recalls. These findings
582 suggest that the degree of hippocampal LFA reduction reflects the degree of
583 correspondence between the retrieved item and its associated context. This
584 gradual LFA reduction goes in line with previous memory recognition studies
585 showing LFA reduction especially during associative memory reinstatement
586 [4, 50]. The difference between correct and false recalls in hippocampal LFA
587 in our study emerged specifically at the moments preceding memory retrieval
588 and faded rapidly afterwards, supporting the role of hippocampal LFA in driv-
589 ing the retrieval of items using their associated context. The hippocampal
590 LFA therefore reflect our ability to retrieve memories bound to a given con-
591 text, enabling the distinction between similar memories that happened under
592 different circumstances.

593 Competitive interference between memories frequently results from their
594 associations to similar contexts [51–53]. Computational models of memory

595 posit that during learning the hippocampus associates features representing
596 each item with a dynamic representation of spatio-temporal context. During
597 memory search, the current context cues item retrieval [31, 32], explaining why
598 items learned under similar contexts may be falsely-recalled. Indeed, exposing
599 participants to an item’s encoding context not only boosts correct item recognition,
600 but also leads to false recognition of similar items that were never actually
601 learned [54?]. While the influence of contextual interference on memory is well
602 established, the neural processes giving rise to such contextual interference
603 remain elusive. In our study, false recalls tended to have at least one source of
604 contextual similarity with the desired context (either semantic or temporal, see
605 Supplementary Table 1), implying that such contextual similarity led to their
606 erroneous retrieval. The hippocampal LFA (6-18 Hz) coded for item-context
607 associative strength, with greater LFA reduction signaling greater similarity
608 between the retrieved item and the target context. False recalls sharing greater
609 contextual similarity with the target context manifested a similar LFA reduction
610 to the one characterizing correct recalls. Such LFA reduction may lead to
611 a reduced signal-to-noise ratio in discriminating between correct and contex-
612 tually similar false recalls, yielding higher chances of falsely retrieving those
613 items. Alternatively, LFA reduction for both correct and contextually similar
614 false recalls may reflect enhanced fidelity of these retrieved item in compari-
615 son to contextually dissimilar false recalls [55], rendering them more likely to
616 be retrieved.

617 Contrary to the greater LFA reduction with increased contextual similarity,
618 HFA increases for correct relative to false recalls remained a strong predictor
619 of recall veridicality regardless of the false recall type. HFA increases reflective
620 of correct relative to false recalls, independently of the degree of contextual
621 similarity, follows a large corpus of iEEG studies showing a widespread increase
622 in HFA during diverse memory related processes [13, 18, 20, 22, 28, 29, 56], as
623 well as studies suggesting that HFA reflects a domain-general marker of brain
624 activation [57, 58].

625 The present study also uncovered a low-theta increase for correct relative
626 to false recalls. This finding supports the recent suggestion that averaging the
627 theta signal across the traditional 2-8 Hz range can mask a low-theta increase
628 associated with successful memory retrieval [27]. Our temporal specificity anal-
629 ysis showed that increased low-theta for correct relative to false recalls was
630 especially evident post-vocalization, suggesting that this biomarker reflects a
631 post-processing signal rather than a signal that solely drives memory retrieval.
632 This post-vocalization increase in low-theta may reflect successful reinstatement
633 of the retrieved context [59, 60], or patients’ post-vocalization assessment
634 of the accuracy of their retrieved response [61, 62]. The functional role that
635 low-theta plays in memory retrieval extends an ongoing debate about the rela-
636 tion between theta oscillations and successful memory [26] and the putative
637 distinct roles of low (~ 3 Hz) and high (~ 8 Hz) theta for memory in humans
638 [63, 64].

639 Our multivariate analysis allowed us to establish that neural signals pre-
640 ceding vocalization can reliably distinguish correct from false recalls at the
641 single-subject level. These findings pave the way for interventions aimed at
642 reducing false recalls when those induce significant distress or functional
643 impairment. Individuals suffering from stress-related psychopathology, such as
644 post-traumatic stress disorder, often experience memory intrusions of their
645 traumatic experiences under contexts that are safe and dissimilar to the trau-
646 matic incident [65–67]. Targeted interventions for interfering with the retrieval
647 process of intrusive memories could prove fruitful for such clinical conditions
648 [68, 69]. Along these lines, several studies have shown that intracranial stimu-
649 lation of the hippocampus interferes with memory retrieval [70–73]. Utilizing
650 such methods using closed-loop stimulation, delivered during brain-states pre-
651 dictive of false recalls, is a promising direction for future research aimed at
652 alleviating memory intrusions in such mental health conditions.

653 Our work demonstrates that hippocampal neural signals can predict false
654 memories prior to their commission. The results extend prior studies indicat-
655 ing a widespread HFA increase reflective of successful memory retrieval [18, 25]
656 to LFA, which more specifically codes for the retrieval of items associated with
657 a given context. The findings provide a better understanding of the compu-
658 tational process implemented by the hippocampus during episodic memory
659 retrieval, unraveling the neural signals representing item-context associative
660 strength. Based on these findings, future interventions could be developed for
661 preventing the reinstatement of memories under inappropriate contexts when
662 those become detrimental to mental health.

663 **Supplementary information.** Accompanying supplementary information
664 is included with this manuscript.

665 **Acknowledgments.** We are grateful to the patients for their participation
666 and thank hospital staff and researchers who were involved in data acquisition.
667 This work was supported by the National Institutes of Health grant U01-
668 NS113198.

669 References

- 670 [1] Howard, M.W., MacDonald, C.J., Tiganj, Z., Shankar, K.H., Du, Q.,
671 Hasselmo, M.E., Eichenbaum, H.: A unified mathematical framework for
672 coding time, space, and sequences in the hippocampal region. *The Journal*
673 *of Neuroscience* **34**(13), 4692–4707 (2014). [https://doi.org/10.1523/
674 JNEUROSCI.5808-12.2014](https://doi.org/10.1523/JNEUROSCI.5808-12.2014)
- 675 [2] Ranganath, C.: Binding items and contexts: The cognitive neuroscience
676 of episodic memory. *Current Directions in Psychological Science* (2010).
677 <https://doi.org/10.1177/0963721410368805>

- 678 [3] Hasselmo, M.E., Eichenbaum, H.: Hippocampal mechanisms for the
679 context-dependent retrieval of episodes. *Neural Networks* **18**(9), 1172–
680 1190 (2005). <https://doi.org/10.1016/j.neunet.2005.08.007>
- 681 [4] Staresina, B.P., Michelmann, S., Bonnefond, M., Jensen, O., Axmacher,
682 N., Fell, J.: Hippocampal pattern completion is linked to gamma power
683 increases and alpha power decreases during recollection. *eLife* (2016).
684 <https://doi.org/10.7554/eLife.17397>
- 685 [5] Davachi, L.: Item, context and relational episodic encoding in humans.
686 *Current Opinion in Neurobiology* **16**(6), 693–700 (2006). <https://doi.org/10.1016/j.comb.2006.10.012>
687
- 688 [6] Davachi, L., Wagner, A.D.: Hippocampal contributions to episodic encod-
689 ing: Insights from relational and item-based learning. *Journal of Neu-*
690 *rophysiology* **88**, 982–990 (2002). [https://doi.org/10.1152/jn.2002.88.2.](https://doi.org/10.1152/jn.2002.88.2.982)
691 [982](https://doi.org/10.1152/jn.2002.88.2.982)
- 692 [7] Yonelinas, A.P., Ranganath, C., Ekstrom, A.D., Wiltgen, B.J.: A con-
693 textual binding theory of episodic memory: systems consolidation recon-
694 sidered. *Nature Reviews Neuroscience* (2019). [https://doi.org/10.1038/](https://doi.org/10.1038/s41583-019-0150-4)
695 [s41583-019-0150-4](https://doi.org/10.1038/s41583-019-0150-4)
- 696 [8] Slotnick, S.D.: Does the hippocampus mediate objective binding or sub-
697 jective remembering? *NeuroImage* **49**, 1769–1776 (2010). <https://doi.org/10.1016/j.neuroimage.2009.09.039>
698
- 699 [9] Sadeh, T., Maril, A., Bitan, T., Goshen-Gottstein, Y.: Putting humpty
700 together and pulling him apart: Accessing and unbinding the hippocam-
701 pal item-context engram. *NeuroImage* **60**(1), 808–817 (2012). [https://doi.](https://doi.org/10.1016/j.neuroimage.2011.12.004)
702 [org/10.1016/j.neuroimage.2011.12.004](https://doi.org/10.1016/j.neuroimage.2011.12.004)
- 703 [10] Polyn, S.M., Kahana, M.J.: Memory search and the neural representation
704 of context. *Trends in Cognitive Sciences* **12**(1), 24–30 (2008). [https://doi.](https://doi.org/10.1016/j.tics.2007.10.010)
705 [org/10.1016/j.tics.2007.10.010](https://doi.org/10.1016/j.tics.2007.10.010)
- 706 [11] Ramirez, S., Liu, X., Lin, P.-A., Suh, J., Pignatelli, M., Redondo, R.L.,
707 Ryan, T.J., Tonegawa, S.: Creating a false memory in the hippocam-
708 pus. *Science* **341**(6144), 387–391 (2013). [https://doi.org/10.1126/science.](https://doi.org/10.1126/science.1239073)
709 [1239073](https://doi.org/10.1126/science.1239073)
- 710 [12] Solomon, E.A., Kragel, J.E., Sperling, M.R., Sharan, A.D., Worrell,
711 G.A., Kucewicz, M.T., Inman, C.S., Lega, B.C., Davis, K.A., Stein,
712 J.M., Jobst, B.C., Zaghoul, K.A., Sheth, S.A., Rizzuto, D.S., Kahana,
713 M.J.: Widespread theta synchrony and high-frequency desynchronization
714 underlies enhanced cognition. *Nature Communications* **8**(1), 1704 (2017).
715 <https://doi.org/10.1038/s41467-017-01763-2>

- 716 [13] Burke, J.F., Sharan, A.D., Sperling, M.R., Ramayya, A.G., Evans, J.J.,
717 Healey, M.K., Beck, E.N., Davis, K.A., Lucas, T.H., Kahana, M.J.: Theta
718 and high-frequency activity mark spontaneous recall of episodic mem-
719 ories. *Journal of Neuroscience* **34**(34), 11355–11365 (2014). [https://doi.
720 org/10.1523/JNEUROSCI.2654-13.2014](https://doi.org/10.1523/JNEUROSCI.2654-13.2014)
- 721 [14] Cabeza, R., Rao, S.M., Wagner, A.D., Mayer, A.R., Schacter, D.L.: Can
722 medial temporal lobe regions distinguish true from false? An event-related
723 functional MRI study of veridical and illusory recognition memory. *Pro-
724 ceedings of the National Academy of Sciences, USA* **98**(8), 4805–4810
725 (2001). <https://doi.org/10.1073/pnas.081082698>
- 726 [15] Slotnick, S.D., Schacter, D.L.: A sensory signature that distinguishes true
727 from false memories. *Nature Neuroscience* **7**(6), 664–672 (2004). [https:
728 //doi.org/10.1038/nm1252](https://doi.org/10.1038/nm1252)
- 729 [16] Stark, C.E.L., Yoko, O., Loftus, E.F.: Imaging the reconstruction of true
730 and false memories using sensory reactivation and the misinformation
731 paradigms. *Learning and Memory* **17**(10), 485–488 (2010). [https://doi.
732 org/10.1101/lm.1845710](https://doi.org/10.1101/lm.1845710)
- 733 [17] Karanian, J.M., Slotnick, S.D.: False memory for context activates the
734 parahippocampal cortex. *Cognitive Neuroscience* **5**(3-4), 186–192 (2014).
735 <https://doi.org/10.1080/17588928.2014.938035>
- 736 [18] Long, N.M., Sperling, M.R., Worrell, G.A., Davis, K.A., Gross, R.E.,
737 Lega, B.C., Jobst, B.C., Sheth, S.A., Zaghoul, K., Stein, J.M., Kahana,
738 M.J.: Contextually mediated spontaneous retrieval is specific to the hip-
739 pocampus. *Current Biology* **27**(7), 1074–1079 (2017). [https://doi.org/10.
740 1016/j.cub.2017.02.054](https://doi.org/10.1016/j.cub.2017.02.054)
- 741 [19] Long, N.M., Burke, J.F., Kahana, M.J.: Subsequent memory effect in
742 intracranial and scalp EEG. *NeuroImage* **84**, 488–494 (2014). [https://doi.
743 org/10.1016/j.neuroimage.2013.08.052](https://doi.org/10.1016/j.neuroimage.2013.08.052)
- 744 [20] Long, N.M., Kahana, M.J.: Successful memory formation is driven by
745 contextual encoding in the core memory network. *NeuroImage* **119**, 332–
746 337 (2015). <https://doi.org/10.1016/j.neuroimage.2015.06.073>
- 747 [21] Burke, J.F., Zaghoul, K.A., Jacobs, J., Williams, R.B., Sperling, M.R.,
748 Sharan, A.D., Kahana, M.J.: Synchronous and asynchronous theta and
749 gamma activity during episodic memory formation. *Journal of Neu-
750 roscience* **33**(1), 292–304 (2013). [https://doi.org/10.1523/JNEUROSCI.
751 2057-12.2013](https://doi.org/10.1523/JNEUROSCI.2057-12.2013)
- 752 [22] Griffiths, B.J., Parish, G., Roux, F., Michelmann, S., van der Plas, M.,
753 Kolibius, L.D., Chelvarajah, R., Rollings, D.T., Sawlani, V., Hamer, H.,

- 754 Gollwitzer, S., Kreiselmeyer, G., Staresina, B., Wimber, M., Hanslmayr,
755 S.: Directional coupling of slow and fast hippocampal gamma with neocortical
756 alpha/beta oscillations in human episodic memory. *PNAS* **116**(43),
757 21834–21842 (2019). <https://doi.org/10.1073/pnas.1914180116>
- 758 [23] Manning, J.R., Jacobs, J., Fried, I., Kahana, M.J.: Broadband shifts
759 in local field potential power spectra are correlated with single-neuron
760 spiking in humans. *Journal of Neuroscience* **29**(43), 13613–13620 (2009).
761 <https://doi.org/10.1523/JNEUROSCI.2041-09.2009>
- 762 [24] Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., Malach, R.:
763 Coupling between neuronal firing, field potentials, and fMRI in human
764 auditory cortex. *Science* **309**(5736), 951–954 (2005). <https://doi.org/10.1126/science.111091>
- 766 [25] Katerman, B.S., Li, Y., Pazdera, J.K., Keane, C., Kahana, M.J.: EEG
767 biomarkers of free recall. *NeuroImage* **246**, 118748 (2022). <https://doi.org/10.1016/j.neuroimage.2021.118748>
- 769 [26] Herweg, N.A., Solomon, E.A., Kahana, M.J.: Theta oscillations in human
770 memory. *Trends in Cognitive Science* **24**(3), 208–227 (2020). <https://doi.org/10.1016/j.tics.2019.12.006>
- 772 [27] Rudoler, J.H., Herweg, N.A., Kahana, M.J.: Hippocampal theta and
773 episodic memory. *bioRxiv* (Submitted). <https://doi.org/10.1101/2022.03.13.484014>
- 775 [28] Greenberg, J.A., Burke, J.F., Haque, R., Kahana, M.J., Zaghoul, K.A.:
776 Decreases in theta and increases in high frequency activity underlie
777 associative memory encoding. *NeuroImage* **114**, 257–263 (2015). <https://doi.org/10.1016/j.neuroimage.2015.03.077>
- 779 [29] Burke, J.F., Long, N.M., Zaghoul, K.A., Sharan, A.D., Sperling, M.R.,
780 Kahana, M.J.: Human intracranial high-frequency activity maps episodic
781 memory formation in space and time. *NeuroImage* **85**, 834–843 (2014).
782 <https://doi.org/10.1016/j.neuroimage.2013.06.067>
- 783 [30] Ranganath, C., Hsieh, L.-T.: The hippocampus: a special place for time.
784 *Annals of the New York Academy of Sciences* (2016). <https://doi.org/10.1111/nyas.13043>
- 786 [31] Howard, M.W., Kahana, M.J.: A distributed representation of tempo-
787 ral context. *Journal of Mathematical Psychology* **46**(3), 269–299 (2002).
788 <https://doi.org/10.1006/jmps.2001.1388>
- 789 [32] Polyn, S.M., Norman, K.A., Kahana, M.J.: A context maintenance and
790 retrieval model of organizational processes in free recall. *Psychological*

- 791 Review **116**(1), 129–156 (2009). <https://doi.org/10.1037/a0014420>
- 792 [33] Umbach, G., Kantak, P., Jacobs, J., Kahana, M.J., Pfeiffer, B.E., Sperling,
793 M., Lega, B.: Time cells in the human hippocampus and entorhinal cortex
794 support episodic memory. *PNAS* **117**(45), 28463–28474 (2020). [https://](https://doi.org/10.1073/pnas.2013250117)
795 doi.org/10.1073/pnas.2013250117
- 796 [34] Eichenbaum, H.: Time cells in the hippocampus: a new dimension for
797 mapping memories. *Nature Reviews Neuroscience* **15**(11), 732–744 (2014).
798 <https://doi.org/10.1038/nrn3827>
- 799 [35] Zeithamova, D., Preston, A.R.: Temporal proximity promotes integration
800 of overlapping events. *Journal of Cognitive Neuroscience* **29**(8), 1311–1323
801 (2017). https://doi.org/10.1162/jocn_a.01116
- 802 [36] Landauer, T.K., Foltz, P.W., Laham, D.: An introduction to latent seman-
803 tic analysis. *Discourse Processes* **25**, 259–284 (1998). [https://doi.org/10.](https://doi.org/10.1080/01638539809545028)
804 [1080/01638539809545028](https://doi.org/10.1080/01638539809545028)
- 805 [37] Weidemann, C.T., Kragel, J.E., Lega, B.C., Worrell, G.A., Sperling, M.R.,
806 Sharan, A.D., Jobst, B.C., Khadjevand, F., Davis, K.A., Wanda, P.A.,
807 Kadel, A., Rizzuto, D.S., Kahana, M.J.: Neural activity reveals interac-
808 tions between episodic and semantic memory systems during retrieval.
809 *Journal of Experimental Psychology: General* **148**(1), 1–12 (2019). [https:](https://doi.org/10.1037/xge0000480)
810 [//doi.org/10.1037/xge0000480](https://doi.org/10.1037/xge0000480)
- 811 [38] Nunez, P.L., Srinivasan, R.: *Electric Fields of the Brain*. Oxford University
812 Press, New York (2006)
- 813 [39] Kovach, C.K., Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M.A.,
814 Adolphs, R.: Manifestation of ocular-muscle EMG contamination in
815 human intracranial recordings. *NeuroImage* **54**, 213–233 (2011). [https:](https://doi.org/10.1016/j.neuroimage.2010.08.002)
816 [//doi.org/10.1016/j.neuroimage.2010.08.002](https://doi.org/10.1016/j.neuroimage.2010.08.002)
- 817 [40] Cohen, M.X.: *Analyzing Neural Time Series Data: Theory and Practice*.
818 MIT press, Cambridge, Massachusetts (2014)
- 819 [41] Seabold, S., Perktold, J.: *Statsmodels: Econometric and statistical mod-*
820 *eling with python*. *Proceedings of the 9th Python in Science Conference*
821 **57**(61) (2010). <https://doi.org/10-25080>
- 822 [42] Maris, E., Oostenveld, R.: Nonparametric statistical testing of EEG- and
823 MEG-data. *Journal of Neuroscience Methods* **164**, 177–190 (2007). [https:](https://doi.org/10.1016/j.jneumeth.2007.03.024)
824 [//doi.org/10.1016/j.jneumeth.2007.03.024](https://doi.org/10.1016/j.jneumeth.2007.03.024)
- 825 [43] Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical*
826 *Learning*. Springer, New York (2001)

- 827 [44] Spackman, K.A.: Signal detection theory: Valuable tools for evaluat-
828 ing inductive learning. Proceedings of the Sixth International Work-
829 shop on Machine Learning, 160–163 (1989). [https://doi.org/10.1016/
830 B978-1-55860-036-2.50047-3](https://doi.org/10.1016/B978-1-55860-036-2.50047-3)
- 831 [45] Diana, R.A., Yonelinas, A.P., Ranganath, C.: Imaging recollection and
832 familiarity in the medial temporal lobe: a three-component model. Trends
833 in Cognitive Sciences **11**(9) (2007). [https://doi.org/10.1016/j.tics.2007.
834 08.001](https://doi.org/10.1016/j.tics.2007.08.001)
- 835 [46] Miller, J.F., Neufang, M., Solway, A., Brandt, A., Trippel, M., Mader, I.,
836 Hefft, S., Merkow, M., Polyn, S.M., Jacobs, J., Kahana, M.J., Schulze-
837 Bonhage, A.: Neural activity in human hippocampal formation reveals
838 the spatial context of retrieved memories. Science **342**(6162), 1111–1114
839 (2013). <https://doi.org/10.1126/science.1244056>
- 840 [47] Solomon, E.A., Lega, B.C., Sperling, M.R., Kahana, M.J.: Hippocampal
841 theta codes for distances in semantic and temporal spaces. Proceedings
842 of the National Academy of Sciences **116**(48), 24343–24352 (2019). [https:
843 //doi.org/10.1073/pnas.1906729116](https://doi.org/10.1073/pnas.1906729116)
- 844 [48] Bachevalier, J., Nemanic, S., Alvarado, M.C.: The influence of context on
845 recognition memory in monkeys: Effects of hippocampal, parahippocam-
846 pal and perirhinal lesions. Behavioural Brain Research **285**, 89–98 (2015).
847 <https://doi.org/10.1016/j.bbr.2014.07.010>
- 848 [49] Sakon, J.J., Sylvia Wirth, Y.N., Suzuki, W.A.: Context-dependent incre-
849 mental timing cells in the primate hippocampus. Proceedings of the
850 National Academy of Sciences **111**(51), 18351–18356 (2014). [https://doi.
851 org/10.1073/pnas.1417827111](https://doi.org/10.1073/pnas.1417827111)
- 852 [50] Martín-Buro, M.C., Wimber, M., Henson, R.N., Staresina, B.P.: Alpha
853 rhythms reveal when and where item and associative memories are
854 retrieved. The Journal of Neuroscience **40**(12), 2510–2518 (2020). [https:
855 //doi.org/10.1523/JNEUROSCI.1982-19.2020](https://doi.org/10.1523/JNEUROSCI.1982-19.2020)
- 856 [51] Anderson, J.R., Bower, G.H.: Interference in memory for multiple con-
857 texts. Memory and Cognition **2**(3), 509–514 (1974). [https://doi.org/10.
858 3758/BF03196913](https://doi.org/10.3758/BF03196913)
- 859 [52] Zaromb, F.M., Howard, M.W., Dolan, E.D., Sirotin, Y.B., Tully, M.,
860 Wingfield, A., Kahana, M.J.: Temporal associations and prior-list intru-
861 sions in free recall. Journal of Experimental Psychology: Learning,
862 Memory, and Cognition **32**(4), 792–804 (2006). [https://doi.org/10.1037/
863 0278-7393.32.4.792](https://doi.org/10.1037/0278-7393.32.4.792)
- 864 [53] MacLeod, C.M.: Interference Theory: History and Current Status. In:

- 865 Kahana, M.J., Wagner, A.D. (eds.) Oxford Handbook of Human Memory,
866 2nd edn. Oxford University Press, Oxford, U. K. (In Press)
- 867 [54] Doss, M.K., Picart, J.K., Gallo, D.A.: The dark side of context: Context
868 reinstatement can distort memory. *Psychological Science* **29**(6), 914–925
869 (2017). <https://doi.org/10.1177/0956797617749534>
- 870 [55] Griffiths, B.J., Mayhew, S.D., Mullinger, K.J., Jorge, J., Charest, I., Wim-
871 ber, M., Hanslmayr, S.: Alpha/beta power decreases track the fidelity of
872 stimulus-specific information. *eLife* **8**(e49562.) (2019). <https://doi.org/10.7554/eLife.49562>
873
- 874 [56] Long, N.M., Kahana, M.J.: Modulation of task demands suggests that
875 semantic processing interferes with the formation of episodic associations.
876 *Journal of Experimental Psychology: Learning, Memory, and Cognition*
877 **43**(2), 167–176 (2017). <https://doi.org/10.1037/xlm0000300>
- 878 [57] Burke, J.F., Ramayya, A.G., Kahana, M.J.: Human intracranial high-
879 frequency activity during memory processing: Neural oscillations or
880 stochastic volatility? *Current Opinion in Neurobiology* **31**, 104–110
881 (2015). <https://doi.org/10.1016/j.conb.2014.09.003>
- 882 [58] Lachaux, J.P., Axmacher, N., Mormann, F., Halgren, E., Crone, N.E.:
883 High-frequency neural activity and human cognition: Past, present, and
884 possible future of intracranial EEG research. *Progress in Neurobiology*
885 **98**(3), 279–301 (2012). <https://doi.org/10.1016/j.pneurobio.2012.06.008>
- 886 [59] Kota, S., Rugg, M.D., Lega, B.C.: Hippocampal theta oscillations support
887 successful associative memory formation. *The Journal of Neuroscience*
888 **40**(49) (2020). <https://doi.org/10.1523/JNEUROSCI.0767-20.2020>
- 889 [60] Staudigl, T., Hanslmayr, S.: Theta oscillations at encoding mediate the
890 context-dependent nature of human episodic memory. *Current Biology*
891 **23**(12), 1101–1106 (2013). <https://doi.org/10.1016/j.cub.2013.04.074>
- 892 [61] Wynn, S.C., Nyhus, E.: Brain activity patterns underlying memory confi-
893 dence. *European Journal of Neuroscience* **55**(7), 1774–1797 (2022). <https://doi.org/10.1111/ejn.15649>
894
- 895 [62] Wynn, S.C., Daselaar, S.M., Kessels, R.P.C., Schutter, D.J.L.G.: The
896 electrophysiology of subjectively perceived memory confidence in relation
897 to recollection and familiarity. *Brain and Cognition* **130**, 20–27 (2019).
898 <https://doi.org/10.1016/j.bandc.2018.07.003>
- 899 [63] Goyal, A., Miller, J., Qasim, S.E., Watrous, A.J., Zhang, H., Joel M. Stein,
900 C.S. Inman, Gross, R.E., Willie, J.T., Lega, B., Lin, J.-J., Sharan, A., Wu,
901 C., Sperling, M.R., Sheth, S.A., McKhann, G.M., Smith, E.H., Catherine,

- 902 S., Jacobs, J.: Functionally distinct high and low theta oscillations in the
903 human hippocampus. *Nature Communications* **11**(2469) (2020). <https://doi.org/10.1038/s41467-020-15670-6>
904
- 905 [64] Lega, B., Jacobs, J., Kahana, M.J.: Human hippocampal theta oscillations
906 and the formation of episodic memories. *Hippocampus* **22**(4), 748–761
907 (2012). <https://doi.org/10.1002/hipo.20937>
- 908 [65] Acheson, D.T., Gresack, J.E., Risbrough, V.B.: Hippocampal dysfunction
909 effects on context memory: Possible etiology for posttraumatic stress
910 disorder. *Neuropharmacology* **62**(2), 674–685 (2012). <https://doi.org/10.1016/j.neuropharm.2011.04.029>
911
- 912 [66] Brewin, C.R., Gregory, J.D., Lipton, M., Burgess, N.: Intrusive images
913 in psychological disorders: Characteristics, neural mechanisms, and treat-
914 ment implications **117**(1), 210–232 (2010). <https://doi.org/10.1037/a0018113>
915
- 916 [67] Cohen, R.T., Kahana, M.J.: A memory based theory of emotional disor-
917 ders. *Psychological Review* (In Press). <https://doi.org/10.1101/817486>
- 918 [68] Herz, N., Bar-Haim, Y., Tavor, I., Tik, N., Sharon, H., Holmes, E.A.,
919 Censor, N.: Neuromodulation of visual cortex reduces the intensity of
920 intrusive memories. *Cerebral Cortex* **32**(2), 408–417 (2022). <https://doi.org/10.1093/cercor/bhab217>
921
- 922 [69] Iyadurai, L., Visser, R.M., Lau-Zhu, A., Porcheret, K., Horsche, A.,
923 Holmes, E.A., James, E.L.: Intrusive memories of trauma: A target for
924 research bridging cognitive science and its clinical application. *Clinical*
925 *Psychology Review* **69**, 67–82 (2019). <https://doi.org/10.1016/j.cpr.2018.08.005>
926
- 927 [70] Jacobs, J., Miller, J., Lee, S.A., Coffey, T., Watrous, A.J., Sperling, M.R.,
928 Sharan, A., Worrell, G., Berry, B., Lega, B., Jobst, B., Davis, K., Gross,
929 R.E., Sheth, S.A., Ezzyat, Y., Das, S.R., Stein, J., Gorniak, R., Kahana,
930 M.J., Rizzuto, D.S.: Direct electrical stimulation of the human entorhinal
931 region and hippocampus impairs memory. *Neuron* **92**(5), 983–990 (2016).
932 <https://doi.org/10.1016/j.neuron.2016.10.062>
- 933 [71] Lacruz, M.E., Valentín, A., Seoane, J.J.G., Morris, R.G., Selway, R.P.,
934 Alarcón, G.: Single pulse electrical stimulation of the hippocampus is
935 sufficient to impair human episodic memory. *Neuroscience* **170**(2), 623–32
936 (2010). <https://doi.org/10.1016/j.neuroscience.2010.06.042>
- 937 [72] Coleshill, S.G., Binnie, C.D., Morris, R.G., Alarcon, G., van Emde Boas,
938 W., Velis, D.N., Simmons, A., Polkey, C.E., van Veen, C.W.M., van

939 Rijen, P.C.: Material-specific recognition memory deficits elicited by uni-
940 lateral hippocampal electrical stimulation. *Journal of Neuroscience* **24**(7),
941 1612–16 (2004). <https://doi.org/10.1523/JNEUROSCI.4352-03.2004>

942 [73] Merkow, M.B., Burke, J.F., Ramayya, A.G., Sharan, A.D., Sperling,
943 M.R., Kahana, M.J.: Stimulation of the human medial temporal lobe
944 between learning and recall selectively enhances forgetting. *Brain Stimu-*
945 *lation* **10**(3), 645–650 (2017). <https://doi.org/10.1016/j.brs.2016.12.011>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementary.pdf](#)